This is a pre-print of a contribution published in: WAINA 2020 conference proceedings (L. Barolli et al. (Eds.) published by Springer Nature Switzerland AG. The final authenticated version is available online at: https://doi.org/10.1007/978-3-030-44038-1_117

Please use the following for reference:
Landowska A., Robins B.: Robot Eye Perspective in Perceiving Facial Expressions in Interaction with Children with Autism, in: WAINA 2020, AISC 1150, pp. 1287–1297, Springer Nature, 2020.

Robot Eye Perspective in Perceiving Facial Expressions in Interaction with Children with Autism

Agnieszka Landowska[1], Ben Robins[2]
[1] Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gdansk, Poland
nailie@pg.edu.pl
[2] School of Computer Science, University of Hertfordshire, Hatfield, UK
b.robins@herts.ac.uk

Abstract. The paper concerns automatic facial expression analysis applied in a study of natural "in the wild" interaction between children with autism and a social robot. The paper reports a study that analyzed the recordings captured via a camera located in the eye of a robot. Children with autism exhibit a diverse level of deficits, including ones in social interaction and emotional expression. The aim of the study was to explore the possibility of applying automatic emotion recognition in analyzing human-robot interaction. The study revealed some challenges, that might be classified as activity-based, child condition-based and setup-based ones. Despite those, the facial expressions in children with autism were on average more positive than in a control group of typically developing children. Children with autism seemed to enjoy the interaction with the robot more. The paper might be interesting for researchers and practitioners who plan to combine social robots and emotion recognition in children with autism.

1 Introduction

The population that is likely to benefit from the impact of the technology are individuals with ASD. Autism Spectrum Disorder (ASD) is a lifelong, neurodevelopmental disorder that can occur to different degrees and in a variety of forms [1]. Children with autism suffer from multiple deficits, and limited social and emotional skills are among those, that influence their ability to involve in interaction and communication. The deficits in social interaction and emotional expression are the motivation to introduce social robots in intervention in children with autism [2-4].
There are promising results in the use of robots in supporting the social and emotional development of children with autism [5]. It is not clear, why children with autism are eager to interact with human-like looking robots and not with humans. Regardless of the reason, social robots proved to be a way to get through the social obstacles of a child and make him/her involved in the interaction. Once the interaction happens, there is a unique opportunity to engage a child in gradually building and practicing social and emotional skills.
In the paper, we report a study of applying emotion recognition technologies in analyzing robot-supported intervention in children with autism. We applied automatic

facial expression analysis to videos from interaction sessions with children on the autism spectrum and the typically developing ones. There were two research questions of the study: *can we obtain valuable information on the child-robot interaction process using the automatic emotion analysis tools?* and *what are the challenges of applying the automatic facial expression analysis to the videos captured from the robot eye perspective?* The study fits into the in-the-wild approach. The paper is organized as follows. Section II summarizes related works, section III presents the design of the study, while section IV - the results. Section V summarizes lessons learned and validity threats of the study that are followed with an overview of challenges and forthcoming research in section VI.

2 Related Work

Robot-assisted therapy in autism has been a growing area of research in recent years [2-4]. Many robots have been used to encourage social interaction and collaborative play amongst children with ASD, e.g. mobile robot IROMEC [5] creature/cartoon-like robots such as Probo [6] and Keepon [7], artificial pets like the teddy bear Huggable and the baby seal robot Paro [8][9] and more. Humanoid robots, e.g., Robota, Nao, Kaspar, Milo have been used with children with autism to help mediate interactions with peers and adults [10][11][12][13]. Studies examined the use of these robots as mediators focusing on communication and social interaction, e.g. self-initiated interactions, collaboration, verbal communication, turn-taking, imitation, joint attention, etc. The ASD practitioners expressed the robot Kaspar to be of added value to ASD objectives in domains such as communication, interpersonal interaction, social relations, and emotional wellbeing. [14][15][16] Kaspar robot was not used in conjunction with emotion recognition technologies before.

There are several works on facial expressions in children with autism [14-18]. Trevisan et al. provided a systematic review and meta-analysis in 2018 that revealed that multiple studies suggest that participants with ASD display facial expressions less frequently and shortly, and they are less likely to share facial expressions with others. Their facial expressions are also judged to be lower in quality and less accurate. However, participants with ASD do not express emotions less intensely, nor is their reaction time of expression onset slower [17]. The findings were observed in several studies that analyzed spontaneous expressions and used diverse play activities: free, semi-structured or structured [18][19][20][21]. None of the studies reported were based on interaction with a social robot [17]. Some studies reported emotional expressions in autism analyzed with automatic recognition software [22][24]. In 2018 a study was published, that directly dealt with automatic emotion recognition applied in child-robot interaction in autism [23] and it reported the feasibility of robot perception of affect and engagement in children with autism. Among the constraints, the authors reported that the video came from a fixed (in position) background camera and advised on using “active vision” view from the robot’s (moving) perspective to enable a more naturalistic interaction setting and viewpoint [23]. The study presented in this paper takes the latter suggestion as a goal and provides a preliminary case study of facial expressions perceived via a camera located in the eye of a robot.

3 Method of the Study

The concept of the study was to analyze videos taken during the interaction sessions between Kaspar robot and children with autism (ASD) as well as with typically developing children (as a control group, denoted TD). The videos were captured via a camera that is located in the eye of Kaspar robot. The study is based on typical interaction videos and the procedure of interaction was not modified in any way for capturing the facial expressions of a child. The emotions were not evoked in any way - natural interaction expressions were observed only. The latter condition might be considered as an "in-the-wild" approach, however, the validity of such a decision is further discussed as well.

3.1 Videos of Child-Robot Interaction

In this study, existing recordings were analyzed, that captured the face of a child interacting with Kaspar robot. The video clips are taken randomly from a previous study UH conducted in a special education school for children with autism. The study was a part of or research within the EU Horizon 2020 project BabyRobot. In the study, there were different play scenarios for children with autism to playfully explore elements that are important in developing Visual Perspective Taking (VPT) skills. The sessions consisted of a series of play activities. The tasks in these play scenarios were designed to progressively move from very simple games that children play with the robot to more complex opportunities for interactions. The "Wizard of Oz" protocol was used here as the robot is only semi-autonomous and requires a human operator. It is important to note here, not only that interactive scenarios with low functioning children with autism often feature free or less-structured interactions, but also to note that children with autism are an extremely heterogeneous population and although they share the same core difficulties, each child displays these in an individual way. Therefore at times, the task order and scenarios were also adapted during the sessions to the needs or abilities of a specific child.

In the study reported within this paper, the recorded sessions from the BabyRobot project are taken as-is and analyzed in terms of facial expressions of the children. Participants included children with ASD (n=12) and typically developing children (denoted as TD, n=9). For typically developing children single session was performed that included all of the tasks in a predefined order. For children on the autism spectrum, 2 up to 6 sessions per child were recorded and analyzed. In total 49 sessions were selected for the analysis.

3.2 Analysis of Facial Expressions

Facial expressions were analyzed using off-the-shelf Noldus FaceReader 6 solution with child model for facial expressions applied. We have performed the start-to-finish context-blind analysis at first, then tagged selected videos with tasks (or interaction process if tasks were mixed and interrelated) and cut the relevant information. The recognized emotional expressions: Ekman's six basic emotions (*Happy, Angry, Scared, Surprised, Disgusted, Sad*) plus *Neutral* emotional state. The emotional expressions are reported using (0,1) scale, with intermediate values representing intensity. The basic emotions were additionally mapped into the

Valence-Arousal model of emotions with valence represented in (-1,1) scale, while arousal in (0,1) scale. Moreover, the frame count of face found/not found was used to quantify the observation of difficulty with maintaining the setup with ASD children. Data processing and statistical analysis were performed using the Knime toolkit.

4 Results

4.1 Availability of Facial Expressions from Robot Eye Perspective

There is a challenge of camera location in designing the studies of a human interacting with an object or a person. In the study of the social robot [23], authors used a fixed background position of the camera and reported it as a challenge of the study. In the study presented here, we used videos captured via robot eye camera as an alternative location. The approach seems more natural, however, still holds some drawbacks. Preliminary analysis of the videos revealed two recordings (one ASD and one TD) with face recognized less than 1% of the time and the reason for those was the robot hair, that slipped down to the front of the camera and although being thin, still covered parts of the child's face.

In the case of natural child-robot interaction, it is typical and expected that a child moves, covers his/her face (one of the play scenarios) or shows some items to the robot, sometimes influencing the camera view. However, for children with ASD, the setup was even harder to maintain. Further 6 recordings were excluded from the analysis due to the child not being interested in sitting in front of the robot (all three sessions of one child and one session per the other three children from ASD group). For the ASD group face was not detected in 28-84% of frames (45% on the average), while for TD group 28-44% of frames (37% on the average). Unavailability of the face in the recordings was the result of the procedure and was confirmed with manual tagging. The reasons were three-fold: activity-based, child condition-based, and setup-based.

Regarding the activity-based ones - some of the play scenarios were designed to include a child or a robot covering the face or showing objects.

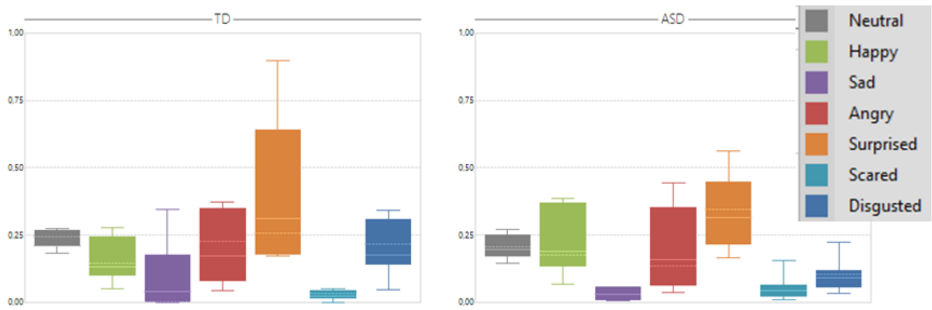
Regarding the child condition-based ones – while typically developing children followed the activities according to the planned order of raising difficulty and were able to accomplish the tasks within one session only, children with ASD exhibited refusal to interaction (at the beginning) and difficult behaviors later on. As a result, the activities were performed in an order the child wanted to follow and were sometimes mixed. Usually, it took two up to six sessions to try to perform all activities and still, some remained unaccomplished.

Regarding the set-up based ones – they resulted from the camera located in the eye of the robot. During the planned activity scenarios the robot was closing eyelids (blinking), was turning head sideways, or covered the face with hands. In experimental design, one might partially eliminate the third group of the challenges by multiplying cameras or choosing an alternative, less natural, scenarios. However, the task-based and child condition-based challenges would probably remain.

4.2 General Level of Emotions for ASD and TD Groups

Having identified the challenges, the study aimed at analyzing whether the recordings with acceptable face availability would reveal any information on the facial expression of the children. The summative level of neutral, happy, sad, angry, surprised, scared and disgusted expressions is visualized in Figure 1a. The summative level of valence and arousal is provided in Figure 1b. The summative levels of expressions differed between the ASD and the control group.

(a)



(b)

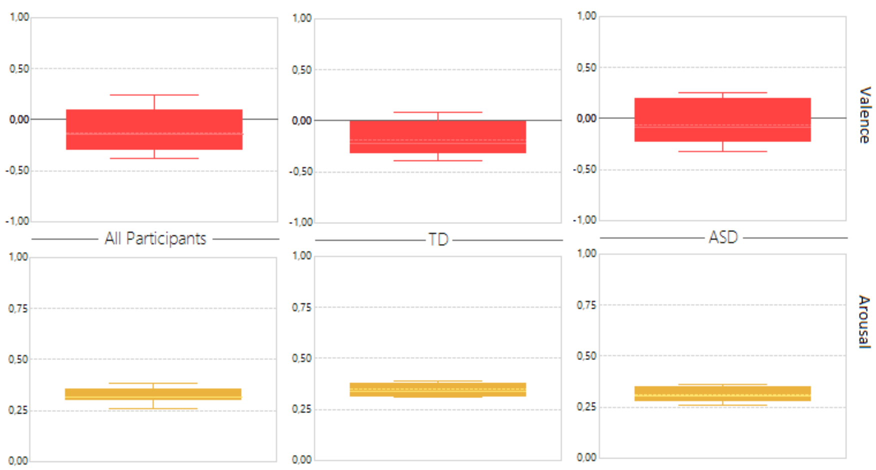


Fig. 1. Comparison of basic expressions summative level for TD and ASD participant groups. (a) Six basic emotions and neutral state in order specified within the legend. (b) Valence - arousal model of emotions.

Higher levels of *happy* and *scared* expressions are observed in children with ASD on the average, while higher levels of *angry* and *disgusted* expressions might be noticed in the TD group. Mean and standard deviation values are provided in Table 1.

Table 1. Comparison of expressions between ASD and TD groups

Expression	TD (n=8) Mean (SD)	ASD(n=11) Mean (SD)	t-value (df=17)	p-value (95%)
Neutral	0,23 (0,03)	0,20 (0,04)	1,77	0,095
Happy	0,15 (0,07)	0,12 (0,04)	-0,59	0,56
Sad	0,001 (0,00)	0,03 (0,02)	-4,16	**0,002**
Angry	0,19 (0,13)	0,16 (0,15)	0,39	0,70
Surprised	0,24 (0,08)	0,31 (0,14)	-1,35	0,19
Scared	0,03 (0,02)	0,05 (0,04)	-1,32	0,21
Disgusted	0,20 (0,09)	0,09 (0,06)	2,97	**0,01**
Valence	-0,20 (0,16)	-0,099(0,23)	-1,15	0,27
Arousal	0,34 (0,04)	0,30 (0,04)	1,72	0,11

Participant averages of emotional expressions over time must be taken with precaution, as there is no "average" emotional state of a person. The automatic emotion recognition applies an averaged model to a face and individual differences might influence the result, depending on how well the actual face fits the averaged model. Therefore a comparison between individual participants was not performed.

Independent groups t-test was performed (with variance not equal as confirmed by *Levene* test). Only for "scared" and "disgusted" expressions, the differences were found significantly different. Please note that for ASD children multiple sessions were treated as one and the recordings with excessive face occlusions were excluded from the analysis and as a result, the groups are smaller which might cause non-significant results for the other emotional expressions.

4.3 Session-by-Session Changes in ASD Group

For the ASD group, multiple sessions were held (2-6/child). Differences between sessions might be interesting from the perspective of the adoption of the robot presence and interaction. Table 2 shows session levels of *neutral* expression.

Table 2. Session-by-session percentage of neutral emotional expressions in ASD group (P-participants, while P10 up to P21 are children with ASD, S-sessions numbered according to time order, na - not applicable).

P	S1	S2	S3	S4	S5	S6
P10	17,30	15,00	14,90	5,00	na	na
P11	12,30	16,60	na	na	na	na
P12	17,00	16,00	13,40	18,20	na	na
P13	24,20	11,90	16,90	na	na	na
P14	20,70	15,20	12,10	12,90	na	na
P15	na	na	na	na	na	na
P16	27,00	18,60	19,70	18,40	na	na
P17	17,10	18,90	na	na	na	na
P18	19,30	22,90	16,40	na	na	na
P19	21,90	29,80	na	na	na	na
P20	31,60	27,30	29,00	23,20	30,50	25,70
P21	na	26,30	18,30	na	na	na

Neutral expression was chosen due to the fact of being assigned as a default state when no other emotional expression was recognized. The purpose of this analysis was to observe whether atypical facial expressions of children with autism are recognized and reported. The *neutral* emotional expressions are reported for 5 up to 27% of session times only. In those sessions of interaction with the robot Kaspar, children with autism performed natural and spontaneous facial expressions. Verification of the recognition of the actual emotional states is beyond this study.

4.4 Within-Session Analysis with Manual Tagging

The goal of manual tagging was to cut the videos according to the start and stop of activity scenarios. Context-informed analysis might provide more information on individuals as well as particular scenarios. Not all session recordings were manually tagged. Three ASD children and one TD child recordings were tagged.

One TD child session recording was tagged with tasks start and end times. Task 4 was excluded from analysis as for the task face was recognized for a small number of frames only. The results on basic emotional expressions (mean values) are provided in Table 3, while valence and arousal are visualized in Figure 3.

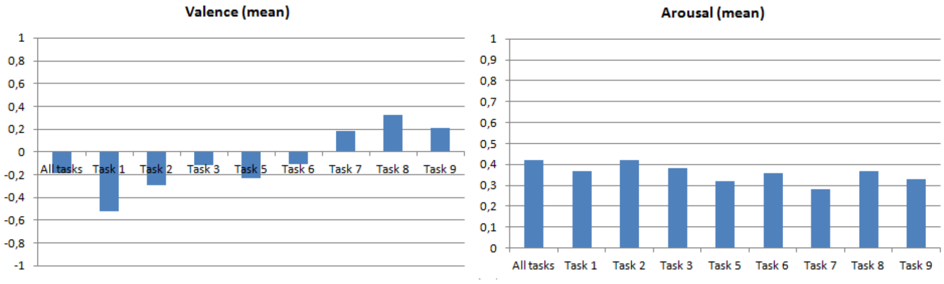


Fig. 2. Task-by-task analysis for selected TD participant represented with valence-arousal model of emotions.

Table 3. Results of emotion recognition per tasks for selected participant (as expressed with Ekman's six basic emotions model plus neutral state) Tasks were child-robot interaction scearios based on turn taking and collaborative activities.

Task ID	Neutral	Happy	Sad	Angry	Surprised	Scared	Disgusted
Task 1	0,18	0,06	0,01	0,42	0,04	0,03	0,27
Task 2	0,21	0,11	0,01	0,13	0,15	0,09	0,29
Task 3	0,27	0,14	0,01	0,09	0,17	0,08	0,12
Task 4	na	na	na	na	na	na	na
Task 5	0,20	0,09	0,00	0,07	0,32	0,04	0,27
Task 6	0,27	0,11	0,00	0,08	0,32	0,04	0,14
Task 7	0,22	0,26	0,00	0,04	0,44	0,01	0,05
Task 8	0,21	0,35	0,00	0,02	0,65	0,00	0,02
Task 9	0,23	0,32	0,00	0,10	0,52	0,01	0,01
All tasks	0,22	0,12	0,02	0,21	0,29	0,05	0,12

The observations from the task-based analysis revealed low levels of sadness and fear for all of the tasks, while the highest expression intensity was observed for surprise and anger. The observations change with time and task progress – more happiness, more surprise, decreasing anger and disgust might be observed. The valence level changes from negative during the first tasks to positive for the last tasks, while arousal level does not change much over time.

The same manual tagging procedure and analysis were to be performed for ASD children group. However, the taggers found it very difficult to explicitly tag activity start and end times (and among taggers, there was a person who actually performed the interaction sessions as a robot operator). As said before – the activities were performed in an order that followed the child's ability and mood, moreover, some activities were re-started multiple times, some remained unfinished and some were mixed. As a result for 3 children with ASD (11 sessions) tagging revealed start/stop times of any scenario-based interaction, without pointing out to a particular task. As no additional information could be obtained from this analysis regarding the differences between tasks (play activities scenarios), the analysis is not continued. Manual tags were used as verification of the analysis performed without the elimination of tags instead. The values obtained after the manual cut (according to the tags) did not differ from the values (for all metrics) obtained for all time of recordings.

5 Lessons Learned and Discussion

Despite the limitations, the analysis allows to draw some basic conclusions on the possibility of applying automatic emotion recognition to an analysis of Kaspar-child interaction:

(1) There is an observable difference in emotional expressions between TD and ASD condition groups. Two of the differences (for scared and disgusted expressions) were confirmed as statistically significant.

(2) Contextual analysis taking into account the task/activity part of the interaction would allow to interpret individual reactions of a child (for example a child expressed low levels of sadness and fear, and the highest level of surprise; with time and tasks performed more happiness, more surprise, increasing valence, decreasing anger and disgust were observed; there was no change in arousal with time; 8th activity seemed to be child's favorite). This type of analysis was not possible for children with autism. A lesson learned is that following the predefined scenarios for children with autism is difficult, if not impossible for some children. There is a trade-off between observation of natural, spontaneous expressions and capturing precise data required for verification of the research hypothesis.

(3) During 20% of time on average (9 TD children, 12 with ASD condition), the facial expressions were recognizable; the availability of the face recognition was limited by child sitting sideways, Kaspar looking sideways, children putting objects in front of Kaspar's eyes, Kaspar hair in front of the camera or simply lids closed, children walking and moving around;

(4) The availability span might be extended with simple changes during sessions eg. removal of hair from the camera view, child sitting in front of Kaspar, removing a child's hat. Even after adjusting and optimizing the procedure, some issues regarding task-related and condition-related challenges would still remain.

6 Conclusions

The study showed the potential for the use of facial expression recognition in analyzing human-robot interaction in autism interventions. The identified challenges are to be taken into account while planning further experiments. The results of the study were promising, however, the study also showed us the need for further, focused and detailed study. Among directions for future works, one might consider the adaptation of the facial expression recognition algorithms to the specificity of the expression possibilities of children with autism.

Acknowledgments. This publication was co-funded by the Erasmus Plus project of European Commission: EMBOA, Affective loop in Socially Assistive Robotics as an intervention tool for children with autism, contract no 2019-1-PL01- KA203-065096. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein. The pre-print of this publication is distributed free of charge. The research was co-funded by Santander Universidades stipend, as well as FETI Funds, Gdansk University of Technology. The study was inspired by "LUDI - Play for children with disabilities" COST action TD1309, 2014-2018. The videos were analyzed due to the courtesy of Adaptive Systems Research Group, the University of Hertfordshire, led by prof. Kerstin Dautenhahn. Special thanks to Silvia Moros and Dr. Luke Wood for helping us with manual video annotations.

References

[1] American Psychiatric Association: Diagnostic and statistical manual of mental disorders,(DSM-5®). American Psychiatric Pub. (2013)

[2] Richardson K., Coeckelbergh M., Wakunuma K., Billing E., Ziemke T., Gomez P., Vanderborght B., Belpaeme T.: Robot Enhanced Therapy for Children with Autism (DREAM): A Social Model of Autism. IEEE Technology and Society Magazine, vol. 37, no. 1, (2018) 30-39.

[3] Pennisi P., Tonacci A., Tartarisco G., Billeci L., Ruta L., Gangemi S., Pioggia G.: Autism and Social Robotics: A Systematic Review, Autism Research no 9 (2016) 165–183.

[4] Scassellati B., Admoni H., Matarić M.: Robots for Use in Autism Research, Annual Review of Biomedical Engineering, Vol. 14 (201) 275-294.

[5] Ferari E., Robins B., Dautenhahn K.: Robot as a Social Mediator - a Play Scenario Implementation with Children with Autism. 8th international Conference on Interaction Design and Children, Como, Italy, (2009)

[6] Saldien J., Goris K., Yilmazyildiz S., Verhelst W., Lefeber D.: On the design of the huggable robot Probo, Journal of Physical Agents 2(2) (2008) 3-11.

[7] Kozima H., Yasuda Y., and Nakagawa C.: Social interaction facilitated by a minimally-designed robot: Findings from longitudinal therapeutic practices for autistic children, Proceedings of 16th IEEE International Conference on Robot & Human Interactive Communication (2007)

[8] Stiehl W. D., Lee J. K., Breazeal C., Nalin M., Morandi A., Sanna A.: The Huggable: A Platform for Research in Robotic Companions for Pediatric Care, Creative Interactive Play for Disabled Children Workshop at the 8th International Conference on Interaction Design and Children, Como, Italy (2009)

[9] Marti P., Pollini A., Rullo A., and Shibata T.: Engaging with artificial pets, Proceedings of Annual Conference of the European Association of Cognitive Ergonomics, Chania, Greece (2005)

[10] Billard,A., Robins,B., Nadel,J., Dautenhahn, K.: Building Robota, a mini-humanoid robot for the rehabilitation of children with autism, RESNA Assistive Technology Journal, vol. 19, (2006) 37-49.

[11] Tapus A., Peca,A., Aly, A, Pop,C., Jisa,L., Pintea,S. Rusi,S., David, O.: Children with autism social engagement in interaction with Nao, an imitative robot – A series of single case experiments, Interaction Studies vol. 13(3) (2012) 315-347.

[12] Dautenhahn K., Nehaniv C. L., Walters M. L., Robins B., Kose-Bagci H., Mirza N. A., and Blow M.: KASPAR - A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research, Special Issue on Humanoid Robots, Applied Bionics and Biomechanics, vol. 6(3) (2009) 369-397.

[13] Milo, https://robots4autism.com/, 2019.

[14] Huijnen,C., Lexis,M., de Witte, L.: Matching Robot KASPAR to Autism Spectrum Disorder (ASD), International Journal of Social Robotics, 8 (2016) 445–455.

[15] Huijnen C., Lexis M., Jansens R., de Witte L.: Mapping Robots to Therapy and Educational Objectives for Children with Autism Spectrum Disorder. Journal of Autism Developmental Disorders, 46(6) (2016) 2100–2114.

[16] Huijnen C., Lexis M., Jansens R., de Witte L., How to Implement Robots in Interventions for Children with Autism? A Co-creation Study Involving People with Autism, Parents and Professionals. Journal of Autism Developmental Disorders, 47(10) (2017) 3079–3096

[17] Trevisan DA, Hoskyn M, Birmingham E. Facial Expression Production in Autism: A Meta-Analysis, Autism Res,. 11(12) (2018) 1586-1601.

[18] Czapinski, P., Bryson, S. E.: Reduced facial muscle movements in Autism: Evidence for dysfunction in the neuromuscular pathway? Brain and Cognition, 51(2), (2003) 177–179.

[19] Markodimitraki, M., Kypriotaki, M., Ampartzaki, M., &Manolitsis, G.. Effects of context and facial expression on imitation tasks in preschool children with autism. Early Child Development and Care, 183(9), (2013) 1276–1292.

[20] Reddy, V., Williams, E., & Vaughan, A. Sharing humour and laughter in Autism and Down's syndrome. British Journal of Psychology, 93(2), (2002) 219–242.

[21] Misailidi, P. Affective expressions during joint attention interactions with an adult: The case of autism. The Journal of the Hellenic Psychological. Society, 9, (2002) 9–21.

[22] Trevisan D. A., Bowering M., Birmingham A.: Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions, Mol Autism. (2016) 7- 46.

[23] Rudovic O., Lee J., Dai M., Schuller B., Picard R.W.: Personalized machine learning for robot perception of affect and engagement in autism therapy, Science Robotics 3(19) (2018)

[24] Capriola-Hall N., Trubanova A., Swain D. White S.: Group differences in facial emotion expression in Autism: Evidence for the utility of machine classification, Science Robotics Vol. 3, Issue 19 (2018)