

A hybrid approach based on a robotic system and playware technology to adapt and predict behaviours from children with Autism Spectrum Disorder

UMinho | 2023 Vinicius Silva



Universidade do Minho
Escola de Engenharia

Vinícius Corrêa Alves da Silva

A hybrid approach based on a robotic system and playware technology to adapt and predict behaviours from children with Autism Spectrum Disorder

Fevereiro de 2023



Universidade do Minho
Escola de Engenharia

Vinícius Corrêa Alves da Silva

A hybrid approach based on a robotic system and playware technology to adapt and predict behaviours from children with Autism Spectrum Disorder

Tese de Doutoramento
Programa Doutoral em Engenharia Eletrónica e de Computadores

Trabalho efetuado sob a orientação da
Professora Doutora Filomena Maria da Rocha Menezes de Oliveira Soares
Professor Doutor João Miguel Clemente de Sena Esteves

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição

CC BY

https://creativecommons.org/licenses/by/4.0/

DECLARAÇÃO

Nome: Vinícius Corrêa Alves da Silva

Endereço eletrónico: _____ Telefone: 917672420 (alternativo)

Bilhete de Identidade/Cartão do Cidadão: 14374652

Título da tese:

Orientador(es): Professora Doutora Filomena Maria da Rocha Menezes de Oliveira Soares e Professor Doutor João Miguel Clemente de Sena Esteves

Ano de conclusão: _____

Designação do Ramo de Conhecimento do Doutoramento: Engenharia Eletrónica e de Computadores

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA TESE/TRABALHO.

Universidade do Minho, ____/__/_____(date)

Assinatura: _____

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledge the Code of Ethical Conduct of the University of Minho.

University of Minho, _____

Full name: _____

Signature:

_____

ACKNOWLEDGEMENTS

I am using this opportunity to express my gratitude to everyone who supported me throughout the development and writing of the present work. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work.
In particular, I would like to give my thanks to:

- My supervisor Dr. Filomena Soares for her support, guidance, and freedom to pursue my own ideas. She consistently allowed this dissertation to be my own work but steered me in the right direction whenever she thought I needed it. All advice she gave were invaluable and I thank her sincerely.
- My co-supervisor Dr. João Sena Esteves for his assistance, advices, supervision, and excellent feedback.
- My colleague Sandra Queirós for her knowledge on behaviour analysis. I am gratefully indebted for her very valuable comments on this thesis.
- Dr. Ana Paula Pereira for her insight, enthusiasm, and for sharing her vast knowledge on Autism Spectrum disorder and especial education.
- Dr. Celina Leão for her insight, enthusiasm, and for sharing her vast knowledge in statistical analysis.
- The professionals (therapists, psychologists, and teachers), parents, and the children who participated in the studies presented in this dissertation.
- My family, especially my mother for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this dissertation. This accomplishment would not have been possible without them.

FUNDS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia under the SFRH/BD/SFRH/BD/133314/2017 scholarship.

AGRADECIMENTOS

Estou a usar esta oportunidade para expressar a minha gratidão a todos que me apoiaram ao longo do desenvolvimento e escrita do presente trabalho. Eu sou grato pela orientação, pelas críticas construtivas indispensáveis e conselhos durante a dissertação.

Em particular, gostaria de agradecer:

- A minha supervisora Dra. Filomena Soares, pelo seu apoio incondicional, orientação e liberdade para perseguir as minhas ideias. Ela constantemente permitiu que esta dissertação fosse o meu próprio trabalho, mas me guiou na direção certa sempre que fosse necessário. Todos os conselhos que me deu forma inestimáveis e agradeço-lhe sinceramente.
- Ao meu coorientador Dr. João Sena Esteves, pela sua ajuda, conselhos, supervisão e um excelente *feedback*.
- A minha colega Sandra Queirós pelo seu valioso conhecimento em análise comportamental a partir de vídeos. Estou eternamente agradecido pelas suas valiosas observações em relação ao trabalho desenvolvido.
- A Dra. Ana Paula Pereira pela sua visão e entusiasmo, e por compartilhar o seu vasto conhecimento sobre Perturbações do Espectro do Autismo (PEA).
- A Dra. Celina pela sua visão e entusiasmo, e por compartilhar o seu vasto conhecimento em análise estatística.
- Aos profissionais (terapeutas, psicólogos e professores), pais e as crianças que participaram dos estudos apresentados nesta dissertação.
- A minha família, especialmente a minha mãe por fornecer apoio incondicional e incentivo contínuo ao longo dos meus anos de estudo e também no processo de pesquisa e escrita da dissertação. Esta conquista não teria sido possível sem eles.

FINANCIAMENTO

Este trabalho foi financiado pela FCT – Fundação para a Ciência e Tecnologia no âmbito da bolsa SFRH/BD/SFRH/BD/133314/2017.

ABSTRACT

Social interaction is a fundamental aspect of human experience and plays a crucial role in shaping an individual's emotions, thoughts, and relationships with the others. It encompasses a wide range of behaviours, including speaking, listening, recognizing facial expressions, body language, and other nonverbal cues. Individuals with Autism Spectrum Disorder (ASD) can face challenges in understanding and participating in social interactions. ASD is characterized by a triad of impairments in interaction, communication, and reduced interest. Different technological tools have been employed in the supporting processes with children with ASD, for example social robots, providing the opportunities to practice social interaction and social-emotional skills in a non-threatening and supportive environment.

The present thesis focus on the use of a humanoid robot to promote social interaction and communication in children with ASD cooperating in a triadic interaction with an experimenter, a robot, and an Object with Playware Technology (OPT), as object of shared attention, which will actively adapt its behaviour to the child's actions in a supervised manner. In order to fulfil this goal, in an initial step OPT devices were developed based on the expertise of professionals and they were successfully tested in pilot studies. Then, a supervised behavioural framework, consisting of three layers – Perception, Reasoning, and Actuation – was developed. This framework uses machine learning techniques to identify nonverbal actions of the child in real-time and propose possible adaptation system prompts to the experimenter, which will choose if they will be executed. The system was tested in a pilot study with game scenarios targeting emotion recognition skills. Finally, a study was conducted in order to assess if the proposed approach can promote social interaction while teaching children with ASD emotion recognition skills. Verbal and non-verbal communicative behaviours, as well as the children's attention patterns, and the children's performance in the game scenarios, are used as measures to evaluate the social interaction in the triadic configuration. Along the sessions, both the children's performance and their attention levels increased. They also tended to verbalize more in the sessions where more system adaptation prompts where executed.

The findings suggest that a hybrid supervised autonomous approach may enhance the social interaction of children and promote the development of emotion recognition skills in children with ASD.

Keywords: Human-Robot Interaction, Social Robots, Machine Learning, Objects with Playware Technology (OPT), Autism Spectrum Disorder (ASD), Social Interaction, Socio-Emotional Skills Development.

RESUMO

A interação social é um aspeto fundamental da experiência humana e desempenha um papel crucial na formação das emoções, pensamentos e relações de um individuo com os outros. Abrange uma ampla gama de comportamentos, incluindo falar, ouvir, reconhecer expressões faciais, linguagem corporal e outros indicadores não verbais. Indivíduos com Perturbação do Espetro do Autismo (PEA) podem enfrentar desafios na compreensão e participação nas interações sociais. A PEA é caracterizada por uma tríade de comprometimentos na interação, comunicação e interesses reduzidos. Diferentes ferramentas tecnológicas têm vindo a serem usadas no apoio a crianças com PEA, por exemplo robôs sociais, fornecendo aos indivíduos com PEA oportunidades de praticar a interação social e competências socio-emocionais num ambiente controlado e não ameaçador.

A presente tese foca-se no uso de um robô humanoide para promover a interação social e a comunicação em crianças com PEA, cooperando numa interação triádica com um investigador, um robô e um objeto de atenção compartilhada com tecnologia *playware* (OPT). O robô adapta ativamente o seu comportamento às ações da criança de forma autónoma, mas supervisionada. Para cumprir este objetivo, numa etapa inicial desenvolveram-se dispositivos OPT com base na opinião de profissionais e testaram-se com sucesso esses dispositivos em estudos piloto. De seguida, desenvolveu-se uma arquitetura comportamental supervisionada, composta por três camadas – Perceção, Raciocínio e Ação. A arquitetura proposta usa técnicas de *machine learning* para identificar em tempo real ações não verbais da criança e propor ao experimentador possíveis formas de adaptação do sistema, podendo o experimentador decidir se serão executadas. O sistema foi testado num estudo piloto com duas atividades que visam promover competências de reconhecimento de emoções. Por fim, conduziu-se um estudo para avaliar se a abordagem proposta pode promover a interação social e ensinar competências de reconhecimento de emoções em crianças com PEA. Para se avaliar a interação social na configuração triádica usaram-se como medidas várias métricas baseadas nos comportamentos de comunicação verbais e não verbais, padrões de atenção e de desempenho das crianças. Ao longo das sessões, aumentaram o desempenho e a atenção das crianças, tendo havido mais verbalização nas sessões onde ocorreram mais eventos de adaptação do sistema. Os resultados sugerem que a abordagem proposta pode promover a interação social e competêcnias de reconhecimento de emoções em crianças com PEA.

PALAVRAS CHAVE: Interação Humano-Robô, Robôs Sociais, *Machine Learning*, Objetos com Tecnologia *Playware* (OPT), Perturbação do Espectro do Autismo (PEA), Interação Social, Desenvolvimento de Competências Socio-Emocionais.

CONTENTS

Acknowledgements vi
Agradecimentos vii
Abstract viii
Resumo ix
List of Figures xiii
List of Tables xx
Abbreviations xxii
1 Introduction 1
1.1 Problem Statement, Motivations, and Scope 3
1.2 Objectives 5
1.3 Contribution to Knowledge 6
1.4 Results of the Developed Scientific Activity 7
1.5 Thesis Structure 9
2 Literature Review 14
2.1 Autism Spectrum Disorder 17
2.1.1 Social Interaction and Deficits in Emotional Process 19
2.1.2 Supporting Intervention Methods for Children with ASD 21
2.2 Affective Computing 23
2.2.1 Affect Sensing – Facial Tracking, Head Pose Estimation 24
2.2.2 Human Action Recognition and Detection of Typical and Stereotypical Behaviours 27
2.3 Objects based on Playware Technology (OPT) 31
2.4 Human-Robot Interaction – Social Assistive Robotics and ASD 34
2.4.1 Facial Expressive Robots 40
2.4.2 Social Assistive Robots and Autonomy 46
2.5 Summary and Conclusions 52
3 System Design: Architecture, Ethics, and Methodology 68
3.1 Introduction 69
3.2 Experimental Set-Up 70

3.2.1 Depth Sensing Device 71
3.2.2 This sensor was manly chosen due to its smaller size The Robot 72
3.2.3 The Activities 73
3.2.4 Room Setup 74
3.3 Ethical concerns 75
3.4 Methodological Considerations 76
4 Development of Object with Playware Technology 81
4.1 Introduction 82
4.2 Development of the Object with Playware Technology – PlayCube 82
4.2.1 Results 83
4.2.2 Summary and Conclusions 86
4.3 Development of the Object with Playware Technology – PlayBrick 86
4.3.1 Methods 87
4.3.2 Results 92
4.3.3 Summary and Conclusions 100
5 Behavioural Control Architecture 105
5.1 Perception Layer – Hardware, OPT, Head Pose estimation, and body pose estimation 107
5.1.1 Perception Layer – Detection of non-verbal behaviours for action recognition 109
5.1.2 Summary and Conclusions 125
5.2 The Reasoning Layer 126
5.2.1 Structure of the Interview 126
5.2.2 Results 128
5.2.3 Summary and Conclusions 137
5.3 The Actuation Layer 137
5.3.1 Summary and conclusions 140
5.4 Pilot Study 140
5.4.1 Results 142
5.4.2 Discussion 146

6 Exploring the use of a supervised architecture for promotion of socio-emotional skills in children with ASD 150
6.1 Research Questions 152
6.2 Methods 153
6.2.1 Participants 155
6.2.2 Procedures 156
6.3 Results 157
6.3.1 Analysis of Eye Gaze 158
6.3.2 Verbal communication 162
6.3.3 Performance in the game scenarios 163
6.3.4 Pre- and post- test analysis 168
6.4 Discussion of the results 169
6.5 Summary of hypothesis and conclusions 173
7 Conclusions and Future Work 178
7.1 Conclusions 179
7.2 Challenges and Rewards of Multidisciplinary Research 184
7.3 Operational Constraints 185
7.4 Future Work 186
References 188
Appendixes 209

LIST OF FIGURES

Figure 2.1: OpenFace facial behaviour analysis pipeline, including: landmark detection, head pose and eye gaze estimation, facial action unit recognition. The outputs of all of these are highlighted in green, (Baltrusaitis et al., 2018). 25
Figure 2.2: Head pose and eye gaze estimation on video sequences. Eye gaze estimation is highlighted in green and heap pose highlighted with the blue 3D box (Baltrusaitis et al., 2018). 25
Figure 2.3: Output of OpenPose, detecting body, foot, hand, and facial landmarks, retrieved from (Cao et al., 2021). 26
Figure 2.4: Nuitrack main features including facial tracking, head pose estimation, and body key-points estimation, (Nuitrack, 2019). 27
Figure 2.5: Interactive tiles as a playware object (Lund et al., 2009). 31
Figure 2.6: Children playing the activity with the developed tiles (Yannakakis et al., 2007). 32
Figure 2.7: Children with ASD playing with modular tiles (Lund et al., 2009). 32
Figure 2.8: Social Playware: Cyber-physical system (Suzuki, 2014). 33
Figure 2.9: The developed approach consisting on a serious game (running on a tablet) and the OPT developed (Azevedo et al., 2018). 34
Figure 2.10: Social robots used in ASD research (Pennisi et al., 2016). 35
Figure 2.11: The robotic platform used, LABO (Dautenhahn & Werry, 2004). 36
Figure 2.12: On the left: Several prototypes of Robota. On the right: Robota mimicking the user by lifting up its arm (Dautenhahn & Billard, 2002). 37
Figure 2.13: The KASPAR robot (Dautenhahn et al., 2009b). 38
Figure 2.14: The Keepon robot with the details to the internal and external structures (Kozima et al., 2008). 38
Figure 2.15: The IROMEC robot in the horizontal and vertical configurations with the main screen providing emotional feedback, (Ferrari et al., 2009) 39
Figure 2.16: Humanoid robots used in ASD research – 1) FACE (Mazzei et al., 2012); 2) EDDIE (Sosnowski et al., 2006); 3) Kismet (Breazeal, 2000); 4) SAYA (Hashimoto et al., 2011); 5) Zeno R50 (S. Costa, 2014); 6) Zeno R25 (Robots4Autism, 2018); 7) QT (LuxAI, 2023). 41
Figure 2.17: On the left: the room apparatus. On the right: the humanoid robot NAO used in the experiments with individuals with ASD (E. Bekele et al., 2014). 47

Figure 2.18: Experimental setup configuration with the fixed array of sensory devices (Esteban et al., 2017). 48
Figure 2.19: The different robot control paradigms: (a) Wizard of Oz, (b) full autonomy, and (c) supervised autonomy, (Esteban et al., 2017). 49
Figure 2.20: Diagram illustrating the room setup configuration with the robot used (ASTRO), (Melo et al., 2019). 50
Figure 2.21: The trial involved children with ASD in a school setting playing dyadic and triadic games with the semi-autonomous Kaspar robot, (Zaraki et al., 2018). 51
Figure 2.22: An example of a typical interaction between the robot (Jibo on the right), the child, and the caregiver, (Scassellati et al., 2018). 51
Figure 3.1: The proposed system. Starting from the left bottom: PlayBrick, Intel RealSense D435, , computer Intel NUC, and the humanoid robot ZECA. 70
Figure 3.2: On the left: Humanoid robot ZECA. On the right: set of facial expressions displayed by ZECA - a) anger, b) fear, c) happiness, d) surprise, and e) sadness. 72
Figure 3.3: Activities flow chart concerning the Recognize and Storytelling game scenarios. 74
Figure 3.4: The proposed experimental design with a triadic configuration, where the child is approximately 25cm from therapist and 85cm from the robot. 75
Figure 4.1: PlayCube prototype. On the left: 3D design. On the middle and right: real images of the prototype. 82
Figure 4.2: Set of emojis used in the OPT device. Starting from the left: sadness, happiness, anger, scared, and surprised. 83
Figure 4.3: Four-step study consisting of a focus group session, think aloud sessions, an online questionnaire, and a pilot study. 87
Figure 4.4: A sample of two sample images for the scenario representing the emotion anger. In scenario A (left), the background and the characters used were based on real-world environments. Scenario B (right) uses a more cartoon alike appearance. 89
Figure 4.5: The physical appearance of the main character of the stories. On the right, a photo of the robot is shown. On the left, a visual cartoon representation of the same robot is shown. 94
Figure 4.6: The interface that is prompted to the child on the OPT where he/she has to select the correct answer by pressing the arrows in order to scroll through the different emojis reflecting the emotion presented in the storytelling scenario. 95
Figure 4.7: Interaction diagram for the storytelling game scenario. 97

Figure 4.8: Sample images of two storytelling game scenarios. On the left it represents the anger emotion. On the right it represents the surprise emotion. 97
Figure 4.9: Final design of the OPT device - PlayBrick. 98
Figure 4.10: Average matching accuracy per emotion (%) for children and adults. 98
Figure 4.11: Number of gazes (A) towards the robot and the OPT for the 4 children (A, B, C, and D) during the sessions. 99
Figure 4.12: Mean gaze time (B) towards the robot and the OPT for the 4 children (A, B, C, and D) during the sessions. 99
Figure 4.13: Answers of the children to the robot's prompts during the sessions. 99
Figure 4.14: Children's mean response time in seconds (with confidence interval) to the robot prompts during the four sessions. 100
Figure 4.15: Number of times that the children needed help per session. 100
Figure 5.1: The behavioural architecture is composed of three main layers – the Perception (blue), Reasoning (yellow), and Actuation (green) layers. 106
Figure 5.2: The PlayBrick and gestures that can be detected - forward tilt (grey arrow) and backward tilt (green arrow). 107
Figure 5.3: Facial features detection and extraction by using OpenFace with the frame rate (top left corner) and confidence of face detection (top right corner). 108
Figure 5.4: Nuitrack detection capabilities. It can track the user and its skeleton data plus it also offers facial tracking and pose estimation. 109
Figure 5.5: The experimental configuration used to extract the data for the dataset. The data was extracted by placing the Intel RealSense in three configurations: in front of the children, on his/her left, and his/her right. 110
Figure 5.6: The overall system pipeline. It starts by detecting and tracking the user joints through the Nuitrack SDK. Then, the data is served as an input to the learning model to classify the action. The joints used are highlighted in orange (left arm, shoulder, elbow, and wrist joints), yellow (right arm, shoulder, elbow, and wrist joints), light blue (left leg hip, knee, and ankle joints), dark blue (right leg hip, knee, and ankle joints), and red (head, neck, and spine joints). 111
Figure 5.7: The joints are encoded by color in the image space, where N is the number of frames. For example, the encoding process for the wrist joint is highlighted in the middle, where the x, y, and z coordinates are normalized by using Equation (1) into the range of 0–255. Additionally, on the right, the encoding process for the N skeleton frame is highlighted in more detail. The final image dimensions are

15 joints (vertical axis) × 32 frames (horizontal axis) × 3 channels (comprising the converted coordinates to RGB). 112
Figure 5.8: Representation of the pixels in image space according to the human body's physical structure. Each coloured strip represents the spatial position of a numbered joint (on the right) over the frames. 112
Figure 5.9: Some samples of skeleton-based image representations from the TSBD dataset used in this work. 113
Figure 5.10: Network architecture used in the present work. BN: Batch Normalization; MP: Max Pooling; GAP: Global Averaging Pooling; FC: Fully Connected layer. The values 64, 128, and 256 are the number of filters used in each convolutional layer. The values 32, 16, and 8 are the resulted tensor width after the MP layer. As an example, the image representation (on the left) serves as input to the model. Adapted from (Ludl et al., 2019). 114
Figure 5.11: Some samples of skeleton-based image representations from the KARD dataset. On the left side, it is shown the images for the skeleton configurations $B2$ and $B4$. On the right side, the images for the skeleton configurations $C2$ and $C4$ are displayed. 116
Figure 5.12: Plot of the training accuracy over the 20 runs. The accuracy for the first epochs (5 to 100) is high-lighted. 117
Figure 5.13: The error plot bar of the mean accuracy per class as well as the standard deviation for the model trained with the skeleton configuration $B1$. 117
Figure 5.14: Some samples of skeleton-based image representations from the TSBD dataset. On the left side, it is shown the images for the skeleton configurations $B3$ and $B5$. On the right side, the images for the skeleton configurations $C3$ and $C5$ are displayed. 118
Figure 5.15: Skeleton-based image representations for the ground-truth class 'COVER_EARS' from the TSBD dataset. The predicted class and its score are shown below each sample on the left. On the right, it is possible to see the activation maps for each sample. It can be observed that the highlighted regions vary depending on the skeleton sequence, e.g., for $B1$ the upper part of the region in the image is highlighted. On $C2$ the regions highlighted are more spread. 120
Figure 5.16: Skeleton-based image representations for the ground-truth class 'ROCKING' from the TSBD dataset. The predicted class and their score are shown below each sample on the left. On the right, it is possible to see the activation maps for each sample. It can be noticed that the highlighted areas are, in general, in the same region on the image for both $B1$ and $C2$. 120

Figure 5.17: Plot of the training accuracy over the 20 runs. The accuracy for the first epochs (5 to 100) is highlighted. 121
Figure 5.18: The error bar plot of the mean accuracy per class as well as the standard deviation for each of the 9 classes over the 20 runs. 121
Figure 5.19: The generated image representations with the spatial and temporal information that serves as input for the model (**below**), as well as frames of the actions 'CLAP' (**above**) for the seated (**on the left**) and upright (**on the right**) positions. In both scenarios, the model successfully classified the action as 'CLAP. 122
Figure 5.20: The generated image representations with the spatial and temporal information that serves as input for the model (**below**), as well as frames of the actions 'TURN' (**above**) for the seated (**on the left**) and upright (**on the right**) positions. In both scenarios, the model successfully classified the action as 'TURN'. 122
Figure 5.21: Tree code based on the questions and participants answers. 129
Figure 5.22: 50 most frequent words said in the SSIs (in Portuguese). 130
Figure 5.23: Robot adaptation procedure flow. 136
Figure 5.24: Reinforcement feedback images prompted by the OPT - on the left for a correct answer, on the right for a wrong answer. The LEDs showcase the colours green (on the bottom left) or red (on the bottom right) according to the correctness of the answer. 137
Figure 5.25: GUI main tab where it is possible to visualize the skeleton data and gestures as well as configure the session time, choose the activity and establish the communication with ZECA and OPT. 138
Figure 5.26: GUI behaviour tab displaying face, body, and session data. It is possible to visualize the session flow. 139
Figure 5.27: GUI child tab where it is possible to add the child data and visualize the child performance over the sessions. 139
Figure 5.28: Remote device used by the researcher/experimenter to confirm or not the system recommended actions as well as pause/resume the activity and repeat the robot prompt. 140
Figure 5.29: Children's performance in terms of right, wrong, and no answers along the sessions for both game scenarios – child 1 at the top; child 2 at the bottom. 142
Figure 5.30: Comparison of the average duration of the behaviours obtained from video analysis (in ■ and _N) with the average duration of the detected behaviour by the trained model (in ■) for child 1. 144

Figure 5.31: Comparison of the average duration of the behaviours obtained from video analysis (in ■ and _N) with the average duration of the detected behaviour by the trained model (in ■) for child 2. 145
Figure 6.1: Images used in pre- and post-tests sessions. At the top it shows the 5 PECS cards which were matched to the figures with the man and the woman, respectively. From left to right, the portrayed emotions are: happiness, surprise, fear, sadness, and anger. 156
Figure 6.2: Average percentage of eye gaze time along the sessions. The percentage time of children gaze towards the OPT increased along the sessions. Overall, children stayed focused on the task since the percentage of gaze to elsewhere is small in comparison to the other percentages of gaze and it decreased over the sessions (with exception of session 4). 159
Figure 6.3: Mean percentage of the joint attention time along the sessions for both game scenarios. The joint attention time in the Storytelling stayed relatively higher in comparison to the Recognize game scenario. 160
Figure 6.4: Comparison of the average duration of eye gaze towards the robot and OPT along the sessions in the Recognize game scenario. 161
Figure 6.5: Comparison of the average duration of eye gaze towards the robot and OPT along the sessions in the Storytelling game scenario 161
Figure 6.6: Mean percentage of eye gaze towards the experimenter along the sessions in both game scenarios. it stayed relatively constant in both game scenarios. 162
Figure 6.7: Mean duration of the Verbal Behaviour in both game scenarios along the sessions. The mean duration of verbal behaviour was higher in the Storytelling game scenario. 163
Figure 6.8: Children's performance in the Recognize game scenario along the sessions. The percentage of successful answers increased along the sessions. 164
Figure 6.9: Children's performance in the Storytelling game scenario along the sessions. The percentage of successful answers increased along the sessions. 164
Figure 6.10: Percentage of successful answers along all the sessions, per game scenario. The percentage of successful answers increased in both game scenarios when comparing the first and last sessions. 165
Figure 6.11: Frequency of system prompts concerning behaviour adaptation along the sessions for both game scenarios. 166
Figure 6.12: Comparison of the average duration of the detected behaviours (in ■) with the duration of the same behaviours obtained from video analysis (in ■ and _N), in seconds. 168

Figure 6.13: Number of successful (SM) and unsuccessful matches (UM) in the Performance Task for all emotions in the Pre- and Post-Test. There was no significant differences in the number of correct matches when comparing the pre- and post-test sessions. 169

LIST OF TABLES

Table 2.1: Some approaches of the state of the art that employs skeleton-driven action classification. 30
Table 2.2: Summary of the facial recognition rates for all the referred projects. 45
Table 3.1: 3D capable sensing devices specifications. 71
Table 3.2: Principal characteristics of the Zeno R-50 robot from Robokind. 73
Table 4.1: Children's answers matching score. 84
Table 4.2: Children answers to the robot prompts where C1 to C8 is the number of the child. 84
Table 4.3: Children answers and mean response time to the robot prompts during the four sessions (SD): RT – response time (s), RA – right answer, WA – wrong answer, NA – no answer. 85
Table 4.4: Professionals' characteristics. 88
Table 4.5: Think Aloud Specialists Characteristics. 90
Table 4.6: Main conclusions of the focus group. 96
Table 5.1: Grouped joints in four limbs. (P1, P2, P4, and P5) and one trunk (P3) following the convention presented in Figures 1 and 3. 115
Table 5.2: Skeleton configurations of groups of joints used where B1 is the control group. Here the order of the joint groups is changed for each skeleton configuration. 115
Table 5.3: Skeleton configurations used where B1 is the control group. Here the sequence of the joints was randomly changed for each skeleton configuration. 115
Table 5.4: Inter-class mean accuracy comparison between the control group (B1) and C2 with the present work test dataset. 119
Table 5.5: Comparison in terms of accuracy of other works in the literature that focus on the detection of stereotypical behaviour. 125
Table 5.6: Participants characterization. 128
Table 5.7: Sessions with the performed game scenarios where the system detected the children's behaviours (marked as ✓). Session 1 and 4 only the Recognize (R) and Storytelling (S) game scenarios were performed, respectively. 143
Table 5.8: Frequency of the analysed behaviour with Noldus (■) vs. the detected behaviours (■) by the model per session and child. 145
Table 6.1: Selected behaviours coded/used during the video analysis procedure. Verbal and non-verbal behaviours as well as typical and stereotypical were selected. 154

Table 6.2: The number of system prompts concerning behaviour adaptation (AT = Add more time; OPT call = call the child's attention to interact with the OPT) along the sessions for both game scenarios (R = Recognize; S = Storytelling) as well as the number of prompts that resulted in correct answers. 166
Table 6.3 : Sessions with the performed game scenarios where the system detected the children behaviours (marked as ✓). Session 1 and 4 only the Recognize (R) and Storytelling (S) game scenarios were performed, respectively. 167
Table 6.4: Statistical comparison of the children's joint attention time. 170
Table 6.5: Statistical comparison of the children's percentage of successful answers. 172
Table 6.6: Statistical comparison of the average percentage of successful answers (SA) vs. the average percentage of unsuccessful plus unanswered prompts (UUP). 172

ABBREVIATIONS

ANN	**A**rtificial **N**eural **N**etwork
ANOVA	**AN**alysis **O**f **VA**riance
API	**A**pplication **P**rogramming **I**nterface
ASD	**A**utism **S**pectrum **D**isorder
BN	**B**atch **N**ormalisation
CE-CLM	**C**onvolutional **E**xperts **C**onstrained **L**ocal **M**odel
CNN	**C**onvolutional **N**eural **N**etwork
DoF	**D**egrees **o**f **F**reedom
FACS	**F**acial **A**ction **C**odding **S**ystem
FC	**F**ully **C**onnected
GAP	**G**lobal **A**verage **P**ooling
GUI	**G**raphical **U**ser **I**nterface
HAR	**H**uman **A**ction **R**ecognition
HCI	**H**uman **C**omputer **I**nteraction
HRI	**H**uman-**R**obot **I**nteraction
IDE	**I**ntegrated **D**evelopment **E**nvironment
JAT	**J**oint **A**ttention **T**ime
MCC	**M**atthews **C**orrelation **C**oefficient
MP	**M**ax **P**ooling
OPT	**O**bject with **P**layware **T**echnology
ReLU	**R**ectified **L**inear **U**nit
RQ	**R**esearch **Q**uestion
SDK	**S**oftware **D**evelopment **K**it
SSIs	**S**emi-**S**tructured **I**nterviews
ToM	**T**heory **o**f **M**ind
TSBD	**T**ypical and **S**tereotypical **B**ehaviours **D**ataset
UUP	**U**nsuccessful and **U**nanswered **P**rompts
ZECA	**Z**eno **E**ngaging **C**hildren with **A**utism

1 INTRODUCTION

Summary

In this chapter, an overview of the problem, motivations, and scope of the work is introduced. It starts with a brief explanation of the concept of social interaction and the difficulties faced by individuals with Autism Spectrum Disorder (ASD) in engaging in such interactions, emphasizing the significance of technological strategies in supporting individuals with ASD. The objectives and contributions of the thesis are then outlined. Finally, the outcomes of the research and the organization of the thesis are presented.

1 Introduction

1.1 Problem Statement, Motivations, and Scope

1.2 Objectives

1.3 Contribution to Knowledge

1.4 Results of the Developed Scientific Activity

1.5 Thesis Structure

The ability to social interaction forms the basis of social relations. At the core of a social interaction there are several direct social behaviours including verbal actions such as talking and non-verbal behaviours such as smiling, eye contact, body gestures, among others, that are exchanged between two or more persons mutually oriented towards each other's selves, (Weber, 2011). These behaviours are fundamental for inferring any mental state that is part of a social interaction process. Moreover, a successful interaction requires that the parties involved are able to extract and understand these social cues. However, automatic understanding of social behaviours may be a very difficult task for some individuals, especially, for children with Autism Spectrum Disorder (ASD).

Distinct technological strategies have been used in the supporting process with children with ASD, by using, for example, tangible interfaces or Objects based on Playware Technology (OPT) and assistive robots (Diehl et al., 2012; Saleh et al., 2021). The latter are designed to identify, measure, and react to social behaviours, being repeatable and objective (Diehl et al., 2012). Additionally, technological devices can offer an exceptional occasion for quantifying social behaviour. According to studies, it was observed that children with autism disorders can exhibit certain positive social behaviours when interacting with robots in contrast to what is perceived when interacting with their peers, caregivers, and therapists (Dautenhahn & Werry, 2007; Ricks & Colton, 2010a).

The robot physical appearance plays an important role in the interaction process with a person. The physical appearance of robots used for autism intervention varies greatly, from simple designs, e.g. four-wheeled mobile robots, to many levels of anthropomorphic forms, including humanoid (S. Costa, 2014), animal-like (Kim et al., 2013), and machine-like systems (Michaud et al., 2005). Consequently, it is important to discuss the appropriate features a robot must have, in order to be used in intervention processes with children with ASD. A humanoid robot can be a useful tool to develop social-emotional skills in the intervention of children with ASD, due to the engagement and positive learning outcome (S. Costa, 2014).

Robots should also be capable of interpreting social cues in order for the interaction process to be more natural and the experience adaptable to the user, since successful human-human communication relies on the ability to read affective and emotional signals. Thus, to make future progress in Human-Computer Interaction (HCI), it is thought that it is necessary for the machine to recognise Users' affect and adapt its behaviour to them, giving an appropriate response for those emotions. Indeed, affective computing tries to build a bridge between the emotionally expressive human and the emotionally lacking computer (D'Mello & Calvo, 2013). Moreover, affective computing can illuminate early emotional dynamics and

provide tools for intervention in disordered emotional functioning. It can be used to understand emotional communication in typically developing children and children with ASD (Robinson & el Kaliouby, 2009; S.Messinger, 2014)(R. Picard & Klein, 2002; Robinson & el Kaliouby, 2009).

The affective behaviour displayed by humans is multi-modal, subtle, and complex. Humans can effortlessly extract and recognize different human behavioural cues, such as facial expressions, eye gaze, hand gestures, posture, deducing successfully, in general, the others state of mind (Ambady & Rosenthal, 1992).

The present work uses a humanoid robot where the interaction process is adaptable to the child. It is based on a prompt-answer-reward, using verbal commands from the robot and visual cues from the developed OPT device. Additionally, the robot can adapt its behaviour to the child actions in a supervised manner, by analysing non-verbal cues in real-time. This process can provide a safer, more pleasant and adaptable atmosphere for the child with ASD (Zaraki et al., 2018).

1.1 Problem Statement, Motivations, and Scope

Non-verbal behavioural cues are present in every human-to-human communication. The successful interpretation of these cues is paramount for an effective interaction. In particular, facial expressions tend to play an important role during a social interaction (Nachson, 1995).

According to Baron Cohen, empathy can be measured, studied and it is often characterized by two distinct components – affective and cognitive. Cognitive empathy is the ability to imagine someone else's thoughts and feelings, knowingly called as 'theory of mind'. On the other hand, affective empathy is characterised as the drive to respond with an appropriate emotion of someone else's feelings (Baron-Cohen et al., 2009). Mental states are comprised of thoughts and emotions, where thoughts include beliefs, desires, intentions, goals, and perceptions (Wilson, 1996). Conversely, emotions as defined by Ekman, appear to be universally recognized, being fractioned into six "basic" emotions: anger, disgust, fear, happiness, sadness and surprise (Ekman et al., 1987) and numerous 'complex' emotions that are more context and cultural dependent (R. C. Solomon & Griffiths, 1999).

Individuals with ASD have difficulties with cognitive empathy, they struggle to interpret the emotional and mental states of others, being confused by other people and consequently, withdraw socially (Baron-Cohen et al., 2009). The lack of interpretation of emotional cues is not only centred around facial expressions but also from vocal intonation and body language (Baron-Cohen et al., 2001; Hobson, 1986; Yirmiya et al., 1992). This difficulty can persist throughout adulthood, especially for more complex

emotions (Baron-Cohen et al., 1997; Golan et al., 2006).

The difficulties are in part of the result of face altered processing (Dawson et al., 2004; Klin et al., 2002), which it may be due to a failure to interpret the mentalistic information conveyed by the eyes (Baron-Cohen et al., 1997). Moreover, children with ASD show reduced attention to faces and, in particular, the eyes (Swettenham et al., 1998). The empathy deficits may have a crucial influence on their social behaviour. This led researchers to question – "Can empathy be taught?" – (Baron-Cohen et al., 2009). Therefore, Baron-Cohen conducted a study, where it was used an animation series, "*The Transporters*", created to motivate children with ASD (4-7 years old) to learn about emotions and facial expressions (Baron-Cohen et al., 2009). The episodes were watched every day for four weeks and the participants were tested before and after the support sessions on emotional vocabulary and emotion recognition at three levels of generalization. The results showed that the support group improved significantly when compared to the control group on all tasks levels. Furthermore, the authors concluded the use of computer based tools can offer a special occasion for teaching of emotions, which can also include robots (Dautenhahn & Werry, 2007).These tools, should also offer some adaptation to the children's behaviours. Classical support sessions use the Wizzard of OZ (WoZ) method, where the therapist or researcher fully controls the robot. Although effective (Huijnen et al. 2016), it requires an additional operator other than the therapist that is engaging with the child in the triadic setup. Additionally, systems that employ the WoZ method usually do not record the child's behaviour (body posture, facial expressions, eye gaze, among others) which might not be suitable to use in live Human-Robot Interaction (HRI) scenarios (Zaraki et al., 2018). Therefore, it is paramount to introduce some degree of autonomy, enabling the robot to perform some autonomous behaviours whilst keeping track of the interaction data. Affective computing tries to bridge this gap, by endowing computational system of capabilities to recognize and even express emotions (R. W. Picard, 1997a, 2009), therefore allowing the introduction of some autonomy to robotic platforms, for example (E. Bekele et al., 2014; E. T. Bekele et al., 2013a; Wainer et al., 2010). Affective sensing and recognition can be core technologies in assistive applications and learning devices for individuals with ASD (el Kaliouby et al., 2006). Some approaches use wearable devices to derive affective cues from individuals with ASD (E. Bekele et al., 2014; E. T. Bekele et al., 2013b). However, these approaches can be invasive, making it dependent on the willingness of the participants to wear it (E. T. Bekele et al., 2013a, 2013b). Thus, there is a need for the development of non-invasive systems and the use of such technologies with children with ASD who may be sensible to sound or wearable artifacts (Rogers & Ozonoff, 2005). This can further enrich the interaction process and provide individualized support (E. T. Bekele et al., 2013a). The opportunities are rich for two-way collaboration in technology-

enhanced human interaction.

Following this idea, one of the motivations for the research presented in this thesis is to develop a hybrid supervised autonomous approach with the robot and OPT being objects of shared attention and assess the use of this approach in social interaction with children with ASD. Therefore, this research aims to understand if the proposed approach may benefit the social interaction of the children with another person while encouraging the acquisition of emotion recognition skills by analysing joint attention patterns, verbal and non-verbal behaviours.

1.2 Objectives

The present work aims to conceive and implement a hybrid supervised autonomous approach composed by a robot equipped with the developed model and playware technology to interact with children with ASD. This interaction is performed while adapting the robot behaviour to the child's actions during a support session. The focus of this research involves the promotion of social interaction behaviours between a child with ASD and other person, using technology as a tool and object of shared attention. To summarize, this research pursues the following goals:

G1: to evaluate the use of objects based on playware technology in a Human-Robot Interaction (HRI) scenario;

G2: to modify the robot behaviour according to eye/head movements patterns and non-verbal actions from children with ASD;

G3: to equip a humanoid robot with a behaviour model by employing machine learning techniques which predicts the behaviour of a child with ASD based on previous non-verbal behaviours, eye/head movements and actions on objects based on playware technology;

G4: to evaluate the social interaction between the child with ASD, the robot and the experimenter, using the developed system.

From a technological point-of-view, this research targets a hybrid approach using a humanoid robot and Objects based on Playware Technology (OPT). Additionally, the system possesses an adaptive behaviour that is contingent upon the child's actions. From a support point-of-view, this research targets a social interaction between the child with ASD and the experimenter, encouraging a triadic interaction between the child, the experimenter, and the robot plus OPT as objects of shared attention.

The research questions guiding this research are:

RQ1: What are the differences in the children's behaviour when OPT is added to a HRI support session?

RQ2: What are the differences in the children's behaviour, or performance, when the robot modifies its behaviour according to the children's non-verbal actions or performance?

RQ3: In what extent does the inclusion of social awareness behaviours and adaptation in a hybrid approach based on a robotic and OPT system benefit social interaction of children with another person (experimenter)?

The following hypotheses are proposed:

H1: Children will be more engaged in the HRI with OPT.

H2: Children will be more engaged and with better performance in the HRI when the robot adapts its behaviour to the children's non-verbal actions or performance.

H3: Children will stay engaged with another person for longer when the robot predicts the children's actions based on the children's eye gaze/head movement, non-verbal and verbal actions, and performance in the activity.

1.3 Contribution to Knowledge

The study presented in this thesis shows the potential advantages of using technological tools such as robots and OPT devices in the support of children with ASD. The use of robots can help capture the attention of the children and provide an opportunity to improve their social skills, which are typically impaired. The research outlined in this thesis suggests that these tools may aid in the development of these skills, making the learning process easier. The key contributions of this thesis include:

- A new hybrid approach using robots and OPT to improve social interaction in children with ASD. The method centres on a play-based interaction between the child and the robot, with the goal of introducing a new object (OPT) to the classical HRI scenario and fostering interaction with an adult during support sessions.
- A four-step study was conducted to provide crucial considerations for designing storyboards for a HRI scenario. The study employed a mix of qualitative (focus group and think-aloud sessions) and quantitative (online questionnaire and pilot study) research methods. The focus group method was used to facilitate discussions between experts and, where possible, validate findings with existing literature. The combination of the focus group and think-aloud sessions outputs generated innovative insights concerning the design features of the OPT to be used in the

interaction process with the robot and the child, namely its format and forms of feedback, among others.

- Development of a machine learning model to detect in real time the child non-verbal actions. This model can detect both typical actions such as hand raise and stereotypical actions as for example rocking, in real-time.
- A three-layer behavioural framework, consisting of Perception, Reasoning, and Actuation, enables the modification of the robot/system behaviour based on the child's actions. The interaction flow of the session is determined by the input from the perception layer. A collection of semi-structured interviews (SSIs) with professionals, from diverse backgrounds, and a parent were conducted to gather information on the possible scenarios where the robot could intervene. Based on their feedback, an adaptation procedure was proposed, which outlines the situations in which the robot could intervene.
- A quantitative study with children with ASD was conducted in order to verify if the developed approach may benefit the social interaction of the children with another person while encouraging the acquisition of emotion recognition skills by children with ASD.

1.4 Results of the Developed Scientific Activity

The findings of the research presented in this thesis were shown and published in the following international publications:

Journal Articles:

- Silva, V., Soares, F., Leão, C. P., Esteves, J. S., & Vercelli, G. (2021). Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network. Sensors, 21(13).

International Conference Papers:

- Silva, V., Soares, F., Esteves, J. S., & Pereira, A. P. (2018). Building a hybrid approach for a game scenario using a tangible interface in human robot interaction. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-02762-9_25.

- Silva, V., Soares, F., Esteves, J. S., & Pereira, A. P. (2019). PlayCube: Designing a Tangible Playware Module for Human-Robot Interaction. Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-030-02053-8_80.
- Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, 2, 497–503.
- Silva, V., Soares, F., Esteves, J. S., Pereira, A. P., & Matos, D. (2020). Adequacy of Game Scenarios for an Object with Playware Technology to Promote Emotion Recognition in Children with Autism Spectrum Disorder. Advances in Intelligent Systems and Computing.
- Silva, V., Soares, F., Esteves, J., Pereira, A., Leão, C., & Queirós, S. (2021). Hybrid Approach to Promote Social Interaction with Children with Autism Spectrum Disorder. 7th International Conference on Information and Communication Technologies for Ageing Well and E-Health, 69–77.
- Silva, V., Soares, F., Sena Esteves, J., & Vercelli, G. (2021). Human action recognition using an image-based temporal and spatial representation. 2020 12th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT), 1–6.
- Amaro, B., Silva, V., Soares, F., & Esteves, J. S. (2019). Building a behaviour architecture: An approach for promoting human-robot interaction. Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-3-319-91334-6_6.
- Azevedo, J., Silva, V., Soares, F., Pereira, A. P., & Esteves, J. S. (2018). An application to promote emotional skills in children with autism spectrum disorders. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11243 LNCS. https://doi.org/10.1007/978-3-030-02762-9_30.
- Santos, P., Silva, V., Sena Esteves, J., Pereira, A. P., & Soares, F. (2022). HiZeca: A Serious Game for Emotions Recognition. In Innovations in Mechatronics Engineering (pp. 393–405). Springer. https://doi.org/10.1007/978-3-030-79168-1_35.
- Santos, P., Silva, V., Soares, F., & Simões, A. (2019). Facial Virtual Tracking: A System to Mirror Emotions. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-30244-3_7.

1.5 Thesis Structure

The thesis is organized in 7 chapters, as follows:

- Chapter 2 offers a comprehensive overview of the interrelated fields of emotional processes, affective computing, and the utilization of social robots with children diagnosed with ASD. It begins by delving into the topic of emotions and theories surrounding emotional processes. Then, it covers the area of affective computing, particularly focusing on head pose data and human action recognition for affective sensing. The chapter also sheds light on the challenges faced by individuals with ASD, particularly in social interaction and emotion recognition. Finally, it highlights studies that have used robots to interact with children with ASD, with a focus on how robots can be used to enhance social interaction and emotion recognition. Projects focusing on the introduction of some degree of autonomy in a supervised manner, that enables the robot to perform some autonomous behaviours are highlighted. The information in this chapter serves as a foundation for the research presented in the thesis.
- Chapter 3 details the design of the system, including its hardware components, the strategies employed for promoting emotion recognition, and the methodology used. The hardware framework is outlined first, followed by a description of the game scenarios and the configuration of the room used during the experiments. Finally, the chapter outlines the methodological considerations taken into account in conducting the present work.
- Chapter 4 focus on the development of two objects with playware technology. Several research methods ranging from quantitative and qualitative approaches were used in the design, conception, and development of these devices. Pilot studies were conducted with the target group in order to assess the use of these devices as an add-on to a HRI configuration in a triadic way (child-robot-experimenter).
- Chapter 5 outlines the structure and features of the supervised behavioural framework, which consists of three key components: Perception, Reasoning, and Actuation. The chapter details the functions of each layer, starting with the Perception layer, which uses machine learning algorithms to identify nonverbal actions in real-time. Next, the Reasoning layer is discussed, with a focus on the study conducted with professionals to define how the robot could adapt its behaviour to the child. Finally, the actuation layer is covered, including a pilot study that evaluated the feasibility and limitations of the system in a school setting with the target population.

- Chapter 6 presents a study aiming to evaluate the use of the developed approach to promote social interaction and socio-emotional skills development in children with ASD. The study analyses the children's joint attention time, verbal and non-verbal behaviours as well as their performance in the game scenarios.
- Chapter 7 draws the conclusion of the work described in this thesis and provides some outlook for the future use of robotics and OPT in support for children with ASD.

References

Ambady, N., & Rosenthal, R. (1992). *Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis* (pp. 256–274).

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2009.0191

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2). https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger Syndrome. *Visual Cognition*, *4*(3). https://doi.org/10.1080/713756761

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013a). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013b). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Dautenhahn, K., & Werry, I. (2007). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & CognitionPragmatics and Cognition*. https://doi.org/10.1075/pc.12.1.03dau

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2). https://doi.org/10.1037/0012-1649.40.2.271

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. In *Research in Autism Spectrum Disorders* (Vol. 6, Issue 1, pp. 249–262). Elsevier Ltd. https://doi.org/10.1016/j.rasd.2011.05.006

D'Mello, S., & Calvo, R. (2013). Beyond the basic emotions: what should affective computing compute? *... Abstracts on Human Factors in Computing ...*, 2287–2294. https://doi.org/10.1145/2468356.2468751

Ekman, P., Friesen, W. V, O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. In *Journal of Personality and Social Psychology* (Vol. 53, Issue 4, pp. 712–717). American Psychological Association. https://doi.org/10.1037/0022-3514.53.4.712

el Kaliouby, R., Picard, R., & Baron-Cohen, S. (2006). Affective computing and autism. *Annals of the New York Academy of Sciences*, *1093*, 228–248. https://doi.org/10.1196/annals.1382.016

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*(2). https://doi.org/10.1007/s10803-005-0057-y

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion: a further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(February), 671–680. https://doi.org/10.1111/j.1469-7610.1986.tb01836.x

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049. https://doi.org/10.1007/s10803-012-1645-2

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism.

Archives of General Psychiatry, *59*(9), 809. https://doi.org/10.1001/archpsyc.59.9.809

Michaud, F., Laplante, J.-F., Larouche, H., Duquette, a., Caron, S., Letourneau, D., & Masson, P. (2005). Autonomous spherical mobile robot for child-development studies. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 471–480. https://doi.org/10.1109/TSMCA.2005.850596

Nachson, I. (1995). On the Modularity of Face Recognition: The Riddle of Domain Specificity*. *Journal of Clinical and Experimental Neuropsychology*, *17*(2). https://doi.org/10.1080/01688639508405122

Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, *14*(Spitz 1945), 141–169.

Picard, R. W. (1997). Affective Computing. In *Rosalind W. Picard*. http://www.amazon.com/exec/obidos/ASIN/0262161702/acmorg-20

Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3575–3584. https://doi.org/10.1098/rstb.2009.0143

Ricks, D. J., & Colton, M. B. (2010). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. https://doi.org/10.1109/ROBOT.2010.5509327

Robinson, P., & el Kaliouby, R. (2009). Computation of emotions in man and machines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3441–3447. https://doi.org/10.1098/rstb.2009.0198

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Saleh, M. A., Hanapiah, F. A., & Hashim, H. (2021). Robot applications for autism: a comprehensive review. In *Disability and Rehabilitation: Assistive Technology* (Vol. 16, Issue 6, pp. 580–602). Taylor and Francis Ltd. https://doi.org/10.1080/17483107.2019.1685016

S.Messinger, D. (2014). *Affective Computing, Emotional Development, and Autism*. 516–536.

Solomon, R. C., & Griffiths, P. E. (1999). What Emotions Really are: The Problem of Psychological Categories. *The Philosophical Review*, *108*(1). https://doi.org/10.2307/2998275

Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., Drew, A., Charman, T., Rees, L., & Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial

stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(5). https://doi.org/10.1017/S0021963098002595

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2010). Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. *2010 10th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2010*. https://doi.org/10.1109/ICHR.2010.5686346

Weber, M. (2011). The Nature of Social Action. In *Max Weber: Selections in Translation*. https://doi.org/10.1017/cbo9780511810831.005

Wilson, G. (1996). Mindblindness—an essay on autism and theory of mind. *Personality and Individual Differences*, *20*(2). https://doi.org/10.1016/s0191-8869(96)80006-7

Yirmiya, N., Sigman, M. D., Kasari, C., & Mundy, P. (1992). Empathy and Cognition in High-Functioning Children with Autism. *Child Development*, *63*(1). https://doi.org/10.1111/j.1467-8624.1992.tb03603.x

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

2 LITERATURE REVIEW

Summary

This chapter starts by introducing the existing theories of emotions in humans. Then, it addresses the Autism Spectrum Disorder (ASD) with emphasis in difficulties in social interaction and emotion recognition skills. The field of Affective Computing is introduced with focus on affect sensing from head motion data and human action recognition, which are the main subfields present in this thesis. The field of Social Assistive Robotics is introduced by mentioning relevant projects that use social-assistive robots to interact with children with ASD. The remaining sections, present specific literature which supports the research presented in the next chapters (projects focusing on the introduction of some degree of autonomy in a supervised manner, that enables the robot to perform some autonomous behaviours).

2 Literature Review

2.1 Autism Spectrum Disorder

2.2 Affective Computing

2.3 Objects based on Playware Technology (OPT)

2.4 Human-Robot Interaction – Social Assistive Robotics and ASD

2.5 Summary and Conclusions

Emotion can be described as an indication of a subjective, affective state that occurs in response to something experienced (Cole et al., 2004). Over the years, many theories have been proposed to explain the origins and functions of emotions.

Charles Darwin proposed that emotions developed as an evolutionary advantage via natural selection, as they are useful for humans, in communication and also in sustaining their survival. For instance, feelings of fear compel people to either fight or flee the source of danger (Darwin, 1872). In addition, Darwin considered that emotional facial expressions are innate and, by being able to interpret them correctly, a person can respond and avoid danger. Darwin's main goal was to demonstrate how emotions in humans were comparable to those in animals, supporting his hypothesis of the existence of a common ancestor for humans and animals.

In the 1880s, a hypothesis was brought up by William James and Carl Lange, the James-Lange theory, about the nature of emotions and feelings (James, 1884; Lange, 1887). It describes the relationship between emotion and arousal by stating that emotions arise as result of physiological stimulation, meaning that in essence the self-perception of changes in the body produce emotional experiences (James, 1884; Lange, 1887). As claimed by this theory, a person laughs (as a consequence of a physiological response to a stimulus), and consequently he/she feels happy (an emotion). Thus the experience of an emotion happens by a chain of events comprised that starts with a stimulus, followed by a perception/interpretation, specific pattern of autonomic arousal, to then experience the particular emotion (James, 1884; Lange, 1887). This theory has some limitations, as evidenced by Walter Cannon, arguing that physiological arousal and emotional experience occur simultaneously, yet independently rather than one leading to the other as the James-Lange theory suggests (Cannon, 1927). By conducting different experiments with animals, like dogs or cats, Cannon discovered that emotion still occurs even if the brain was cut off from the information about bodily responses, such as heart rate or blood pressure (Cannon, 1927). Thus, physiological arousal does not have to precede emotional expression or experience. Additionally, (Bard, 1928), throughout his experiments also concluded that the experience of an emotion does not depend on the input from the body or on how it is responding.

The Schachter-Singer theory takes a cognitive approach by drawing from elements from both James-Lange and Cannon-Bard theories. According to the theory, physiological arousal occurs first and is then followed by a cognitive process in which an individual assigns meaning to the arousal (Schachter & Singer, 1962). Thus, emotions are the result of the interaction between two factors: physiological arousal and

cognition. This theory claims that physiological arousal is cognitively interpreted within the context of each situation, which ultimately produces the emotional experience. For example, a stimulus, such as a snake, triggers physiological arousal, and the cognitive interpretation of this arousal leads to the experience of an emotion, such as fear. Based on the person's past experience, these cognitive interpretations are formed. Furthermore, this theory suggests that similar physiological responses can produce varying emotions, depending on the environment that the person is at the moment.

The Opponent-process theory of emotion proposed by the psychologist Richard Solomon in the 1970s (R. L. Solomon & Corbit, 1974), suggests that emotions are experienced in two phases: the "excitatory phase" and the "inhibitory phase". During the excitatory phase, an individual experiences a primary emotion in response to a stimulus. This primary emotion is typically intense and short-lived. Following the excitatory phase, the individual enters the inhibitory phase, during which the opposite emotion is experienced. This emotion is less intense and longer-lasting than the primary emotion. The theory suggests that this process helps to balance and regulate emotions.

(Ekman, 1971), conducted research in several different cultures, including the United States, Japan, and Papua New Guinea, in order to assess whether certain facial expressions were recognized as indicating the same emotions across cultures. In particular, Ekman and his colleague Wallace Friesen presented participants with photographs of individuals making different facial expressions, and asked them to identify the emotion being expressed. They found that the same facial expressions were consistently recognized as indicating the same emotions across cultures, leading them to conclude that these facial expressions are biologically-based, and not just culturally-determined. Ekman's experiments also led to the conclusion that emotional expressions are related to specific physiological changes in the body, such as changes in heart rate, blood pressure, and muscle activity. The recognition of basic facial expressions is important for social interactions as it allows individuals to understand the emotional state of others. This ability is important for building and maintaining relationships, as well as for interpreting and responding to social cues.

Another theory, the socio-cultural theory, suggests that emotions are learnt and influenced by culture and society (D. et al., 1979). This theory emphasizes the role of social and cultural factors in shaping emotions. For example, some cultures may place a greater emphasis on the expression of certain emotions, such as shame or pride, while others may discourage the expression of certain emotions. This theory is supported by research. It shows that emotions vary across cultures and they can be learnt through socialization and cultural practices.

Damasio's research focuses on the relationship between emotions, the brain, and decision-making,

(Damasio, 1994). According to Damasio, the brain has a pre-organized system for processing emotional responses, which is based on the detection of certain characteristics of the stimulus, such as its shape, colour, texture, or other features. The somatic marker hypothesis, proposed by him, suggests that this emotional response is stored in the brain as a kind of "memory" of the experience. The memory of this emotional response is associated with a particular feeling in the body, which is the somatic marker.

For example, if a person has a bad experience with a certain type of food in the past, their brain may store an emotional response of disgust or nausea associated with that food. Next time the person encounters that type of food, his/her brain retrieves the associated emotional marker, and somatic marker, which is the feeling of nausea, as well as the emotional response, which is disgust, this will guide the person's behaviour to avoid that type of food.

As can be inferred from the theories discussed, there is currently no consensus on a global definition or understanding of emotional processes. Different theories have been proposed to explain the nature and function of emotions, and while some aspects of emotions have been well-established, there are many unanswered questions and ongoing debates in the field. Conversely, most research agrees that emotions are often accompanied by distinctive facial expressions, physiological changes in the body, gestures, among others. Therefore, emotions are typically described as complex psychological and physiological states that are triggered by certain events or situations.

2.1 Autism Spectrum Disorder

Initially defined by Kanner in 1943 as “autistic disturbance of affective contact” (Kanner, 1943), Autism Spectrum Disorder (ASD) is a group of lifelong disabilities that affect people’s ability to communicate and to understand social cues (American Psychiatric Association, 2013a). Kanner also observed that these individuals present an incapacity of relating to others, lack in the use of language, and obsessive desire of maintaining everything the same way (Kanner, 1943). Furthermore, these individuals were also regarded as having restricted activities or interests and stereotypical behaviours (Y. A. Wang et al., 2010). The “autism spectrum” was later defined by Lorna Wing and Judith Gould by conducting a study with 35.000 children (Wing & Gould, 1979). The conclusions addressed from this study allowed Wing and Gould to conclude that the children had difficulties in social interaction, associated to impairments in communication and lack of interest in activities. In general, this triad of difficulties in interpreting social situations, causes children with ASD to lose or miss information on what is happening or happened during the social exchange (Happé et al., 2001).

In the DSM-IV, ASD is defined as a global deficit in three main areas of development: social interaction, communication, and behaviour (Bennett & Goodall, 2016). Conversely, according to the current criteria in the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition), ASD is characterized by persistent impairment in reciprocal social communication and interaction, and restricted, repetitive patterns of behaviour or interests.

With regard to the area of social communication, it is highlighted the triad of deficits in social-emotional reciprocity, in nonverbal communicative behaviours, and, in developing, maintaining, and understanding relationships (American Psychiatric Association, 2013b). Additionally, others criteria includes insistence on sameness, highly restricted fixated interests, hyper- or hypo-reactivity to sensory input, and stereotyped or repetitive motor movements (American Psychiatric Association, 2013b; CDC, 2022b). These stereotyped or repetitive motor movements are defined as restricted and repetitive patterns that appear to the observer to be invariant in form and without any obvious eliciting stimulus (CDC, 2022b). Stereotypical actions are defined as being repetitive and sometimes inappropriate in nature (Cunningham & Schreibman, 2008). They are highly heterogeneous in presentation, being verbal or non-verbal, fine or gross motor-oriented, as well as simple or complex. Frequent examples of stereotyped behaviours are hand flapping/wave, body rocking, among others. They manifest in different ways, depending on individuals, settings, and time conditions (Cunningham & Schreibman, 2008).

In addition, children with ASD tend to have difficulty generalizing, which means they struggle to apply previously acquired knowledge to new situations or contexts (Myles, Brenda Smith; Swanson, Terri Cooper, & Holverstott, 2013).

Baron-Cohen refers that children with ASD have difficulty in understanding the perspective of others and predict behaviours, i.e., children with ASD have difficulty in sharing and showing empathy (Baron-Cohen et al., 1985). Empathy is a basic socio-emotional process of human development that involves the ability to perceive, share, and understand the emotional states of others (Harmsen, 2019).

Understanding the causes of ASD is intricate, as multiple theories intersect, oscillating between theories that aim to understand the symptoms of this disorder through underlying psychological and cognitive processes, and theories that focus on potential neurological causes (Marques, 1998).

Two terms are commonly used to describe individuals with ASD: low functioning ASD and high functioning ASD. High functioning refers to individuals with ASD who are considered to have a higher level of cognitive functioning and an intelligence quotient (IQ) greater than 70. However, these individuals may still present deficits in areas such as communication, emotion recognition and expression, and social interaction (Sanders, 2009).

The Centre for Disease Control and Prevention (CDC) estimates that about 1 in 44 8-year-old children have been identified with ASD in the US, with a higher prevalence in boys (4 times more) than in girls (CDC, 2022a). Concerning the European Union, prevalence estimates were very different among countries, ranging from 1 in 64 (in UK) to 1 in 806 in Portugal (Málaga et al., 2019).

There is a big contrast in the development of children with ASD when compared with typically developing children, and for that reason they need timely and suitable support that responds to all of their impaired areas of development.

2.1.1 Social Interaction and Deficits in Emotional Process

Social interaction is a fundamental aspect of human life and is critical for healthy development and well-being. Individuals with ASD often have difficulty with social interaction and communication, which can negatively impact their quality of life. Children with ASD exhibit a significant qualitative deficit in their use of multiple nonverbal behaviours, such as eye contact, facial expressions, body postures, and gestures, to regulate social interaction (American Psychiatric Association, 2013b). The deficit in social interaction is marked and persistent. In particular, one key marker of this deficit is Joint Attention Time (JAT) which refers to the ability to share attention with others, such as following another person's gaze or pointing to an object. This ability is considered to be an important aspect of social development, and can serve as an early indication of the presence in deficit in social interaction, as it enables individuals to share information and experience with others (Johnson et al., 2007).

In contrast to typically developing children, a child with ASD has problems following the parents' eye gaze, pointing, or looking for social sharing. A child with a typical development, tend to smile in response to a simile or verbalisation in the first months of life. By 12 months, in general, a child will already be able to follow the parents gaze (e.g.: towards an object of interest) when they verbalize "look", (Johnson et al., 2007).

Conversely and oftentimes, children with ASD exhibit a tendency to approach an object without making any request or directing their gaze towards the intended object, making it challenging for adults to comprehend the child's true intentions (Siegel, 2008). Therefore, joint attention time is a pivotal skill in ASD.

The challenges associated with interpreting social situations holistically and paying attention to details can make it difficult for children with ASD to understand and process the information being shared during social interactions. This can make it harder for them to interact with others, as they may not fully grasp the context or overall dynamics of the situation (Happé et al., 2001).

Children with ASD also present imitations impairments. In fact, typically developing children begin to imitate observed actions in the early stages of their lives, which demonstrates that learning by imitation is a fundamental tool to help in the learning process and to improve social skills (Piaget, 1976). Some studies have shown that children with ASD have imitation impairments, probably due to a malfunction of the mirror neuron system (Williams et al., 2001).

Besides the imitation impairments, children with ASD have difficulty in recognizing and understanding mental states in themselves and in others, i.e., the Theory of Mind (ToM) (Baron-Cohen et al., 1985). Presented by Baron-Cohen, ToM consists in the ability to understand that others have mental states, such as beliefs, desires, and intentions, that may differ from one's own. It is the ability to attribute mental states and emotions, to oneself and others and understand that these mental states can influence behaviour.

Developed by Simon Baron-Cohen and his colleagues at the University of Cambridge in 1985, the Sally-Anne test is sued to assess the development of a child's ToM, (Baron-Cohen et al., 1985).

The Sally-Anne test typically involves two dolls, Sally and Anne, and a box. The experimenter shows the child a scenario in which Sally puts a toy into the box, and then leaves the room. The experimenter then moves the toy to a different location, such as a bag, and then brings Sally back into the room. The child is then asked the Belief Question: "Where will Sally look for the toy?". The correct answer to the Belief Question is that Sally believes that the toy is in the box. This test tries to measure a child's ability to understand false beliefs.

Typically developing children by the age of four or five, pass this test by understanding that Sally will look for the toy in the box where she originally put it, because she does not know that it has been moved, and that the child should indicate that Sally will look for the toy in the box. Children with ASD, however, tend to fail this test, indicating that they have difficulty understanding false beliefs, and that they may have difficulty understanding the mental states of others.

ToM is closely related to emotion recognition as recognizing emotions in others requires an understanding of the mental states that give rise to those emotions. Individuals with ASD are less accurate at recognizing emotions in others, and they may have difficulty understanding the social cues that are associated with emotional expressions. They may have difficulty recognizing basic emotions such as happiness, sadness, anger, and fear. They may also have difficulty recognizing more complex emotions such as disgust and surprise. Additionally, individuals with ASD may have difficulty recognizing emotions in the eyes, which is considered a key feature of emotion recognition. For example, a study conducted by (Hobson, 1986), in which children with ASD and typically developing children had to match drawings of gestures to

videotaped vocalizations and facial expressions representative of four emotional states, showed that children with ASD had significantly more difficulty performing the task in comparison with typically developing children.

Another comparison study was conducted with children diagnosed with low- and high-functioning ASD, as well as typically developing children, who were matched by their verbal and non-verbal mental age, in an emotion recognition task (Loveland et al., 1997). The study consisted of having all participants watch video clips and identify the emotions expressed, whether verbally, non-verbally, or both. The emotions presented were happy, angry, sad, surprised, or neutral, and verbal expressions of emotion were either explicit, implicit, or neutral. Meanwhile, non-verbal expressions were animated or flat, with animated expressions clearly conveying happiness, sadness, anger, or surprise and flat expressions being neutral. The results revealed a difference between high-functioning and low-functioning groups. The performance of low-functioning participants indicated that they had difficulty determining how a person in the video clips felt based on what they said, especially when the emotion was not clearly stated. On the other hand, the performance of high-functioning participants suggested that they relied more on non-verbal information than verbal information to determine a speaker's emotion, except when the emotion was explicitly named (Loveland et al., 1997).

The use of social stories have been found to be an emerging support approach for individuals with ASD (Sansosti et al., 2004). According to (Gray & Garand, 1993), a social story is defined as a description of a situation, skill, or concept in terms of relevant social cues, perspectives, and common responses, presented in a specific style and format. Bader et al. conducted a study with the goal of increasing emotion recognition and labelling skills through the use of social stories, (Bader, 2006). The authors reported that the social stories were effective in increasing and labelling of four emotions and the results demonstrated some short-term maintenance of learning from the social story.

Additionally, the use of social stories has been shown to result in improvements in comprehension, an increase in requesting, unprompted spontaneous verbalizations, and effective communication interactions (Hanley-Hochdorfer et al., 2010; Hudock et al., 2011; Samuels & Stansfield, 2012).

2.1.2 Supporting Intervention Methods for Children with ASD

The management of ASD is intricate and generally employs a multifaceted approach.

Behavioural interventions, such as Applied Behavioural Analysis (ABA), have been shown to be effective in improving social and communication skills in individuals with ASD, (McDaniel et al., 2022). Children that participated in the research carried out in this thesis experiments have some support in terms of

behaviour through behavioural strategies of this model. Speech and language therapy can also help to improve language and communication skills. Occupational therapy can help individuals with ASD to develop the skills needed for daily living. It is important to note that the type of support and intervention that is most appropriate for an individual with autism will depend on their unique needs and abilities. In this sense, several support methodologies have been developed (McDaniel et al., 2022). Although there are many support methods available, a special focus will be given to the TEACCH method due to the fact that the children who are involved in the experiments in the research carried out in this thesis are being supported with the principles and practices of this model.

Developed by Eric Schopler and colleagues at the University of North Caroline, the TEACCH (Treatment and Education of Autistic and related Communication-Handicapped Children) methodology is an evidence-based approach for working with individuals with ASD (Schopler & van Bourgondien, 2020). It is based on the idea that individuals with ASD have unique strengths and weaknesses, and that their treatment should be tailored to their individual needs. This methodology is highlighted here, as it is the methodology used with the children in the present research.

One of the key elements of the TEACCH methodology is the use of structured teaching, which involves breaking down tasks into small, manageable steps and providing visual cues and other supports to help individuals with ASD understand and complete the tasks. This approach is based on the idea that individuals with ASD may have difficulty generalizing information from one situation to another, and that they need extra support to understand the relationship between different tasks.

Another key element of the TEACCH methodology is the use of visual supports, such as schedules, social stories, and visual prompts, to help individuals with ASD understand and manage their environment. Vision is one of the strongest skills of children with ASD and utilising visual supports is crucial in helping them understand requests and tasks (Lima & Levy, 2015). It can help individuals with ASD understand what is expected of them, and can also be used to help them understand the meaning of spoken language. In addition, other elements of the TEACCH methodology involve the organization of the physical space to decrease distracting factors and creation of one-on-one workspace.

Nowadays, the use of assistive technology, such as virtual reality and mobile apps, can also provide opportunities for individuals with ASD to practice recognizing emotions in a safe and controlled environment (Boucenna et al., 2014; Grynszpan et al., 2014; Knight et al., 2013). In fact, different approaches are used in intervention process with children with ASD. More recently, the use of affective computing together with social robots and Objects with Palyware Technology (OPT) has been explored in order to facilitate the support process of children with ASD.

2.2 Affective Computing

Affective Computing, firstly introduced by Rosalind Picard, is an interdisciplinary field that embraces computer science, psychology, and cognitive science (R. W. Picard, 1997b). It consists in the study and development of systems and devices that can recognize, interpret, process, and simulate human affects, being the main research motivation the ability to simulate empathy. Moreover, affect sensing, affect synthesis and the design of emotionally intelligent interfaces are fields encompassed in affective computing (Baltrusaitis, 2014). Technological gadgets with this capacity could gather cues, from a variety of sources, and use it to respond to user's emotions and other stimuli. Facial expressions, posture, speech, gestures, force, heart rate, and body temperature can all potentially signify emotional changes. Technologies such as speech, gesture, and facial expressions recognition are being explored and employed in affective computing applications (Baltrusaitis, 2014; Daily et al., 2017; Hudlicka, 2008).
Affective technology can be used in support sessions with children, helping them to develop and optimally have the potential to benefit society in the long term (S.Messinger, 2014). This technology can be applied in the detection and modelling of emotion, using affective software to detect behavioural signs of emotion and model emotional functioning and communication and the construction of software and hardware agents that interact with children. In the autism research, affective computing is used in automated measurement of facial expressions in infants and their parents during early interaction or in the automated detection of infant emotion from vocalizations (e.g. infant cries) (S.Messinger, 2014) or even in detecting and measuring stereotypical behaviour (e.g. hand flapping) (Goncalves et al., 2012b; Jazouli et al., 2019). Additionally, affective computing approaches can also be applied into modelling a system's "emotional response" to a user and to express emotion via personified conversational agents or robots (social robots) (R. W. Picard, 1997b). Although the potential development outcomes of interactions with social robots are only beginning to receive attention in the psychological literature (Kahn et al., 2013), there has been an increase in the presence of these robots in the intervention of children with ASD (S. Costa, 2014; Dautenhahn et al., 2009a; Pennisi et al., 2016). A number of studies have indicated that individuals with ASD tends to show a preference for robot-like features over non-robotics toys, such as robots or OPT (Lund, 2009; Pennisi et al., 2016), and in some situations, respond faster when prompted by a robotic movement than by human movement. Additionally, robotic systems offers certain advantages when compared with virtual agents, likely due to the capability of robotic systems to use physical motion in a manner not possible in screen technologies (S.Messinger, 2014). Despite of this, there have been few systematic and adequately controlled applications of robotic technology in investigating the success/impact during support sessions with children with ASD. Furthermore, the majority of these

robotic systems studied to date have been unable to perform autonomous closed-loop interaction, i.e., these systems are controlled using the Wizard-of-Oz (WOZ) setup, needing human input in every step of the interaction (Pennisi et al., 2016).

For effective ASD support, innovative interventional approaches using robot systems should have the ability to perceive the environment and users' behaviours, states, and activities.

2.2.1 Affect Sensing – Facial Tracking, Head Pose Estimation

In the context of Affective Computing, Facial Tracking and Head Motion Analysis is often used to detect and interpret various facial expressions (such as happiness, sadness, anger, and surprise) and estimate engagement level. Head Pose Estimation can be used to detect other non-verbal cues such as attention or engagement, which can be used to determine if a person is paying attention or becoming disengaged. It is also used to track facial expressions and movements to understand how a person is feeling, what they are thinking, and how they are behaving.

Head Pose Estimation remains challenging due to different light conditions, occlusions, and various facial expressions that may affect the algorithms accuracy (Murphy-Chutorian & Trivedi, 2009). Several approaches have been employed over the years consisting on the use of more or less invasive techniques. These techniques range from the use of Inertial Measurement Units (IMUs) embedded in virtual reality (VR) headsets (HTC, 2023; Meta, 2023), or full-body motion capture systems (OptiTrack, 2023; Vicon, 2023), to mark-less non-wearable approaches consisting on the use of RGB or RGB-D cameras with machine learning techniques in order to estimate the head pose (Mehta et al., 2017, 2020; Shotton et al., 2013; Toshev & Szegedy, 2014; Ye et al., 2011). More recently, vision researchers are more focused on the latter approach, i.e., to obtain the pose estimation using few RGB or RGB-D images and employing key-points-based approaches or non-key-points-based approaches.

The key-points-based approach can be divided in three steps (Dhingra, 2022): face detection (i.e. determines if there is a face in the image (H. Wang et al., 2018)), face landmarks localisation (i.e. identifies the location of the facial landmarks on images or video frames (Wu & Ji, 2019)), and head angle estimation (i.e., 3D orientation of the head, usually Euler angles). Several works in the literature employ this approach with deep learning techniques to predict 3D face models (Baltrusaitis et al., 2018; Cao et al., 2021), usually performing better than the traditional approaches (Dhingra, 2022).

Regarding non-key-points-based approaches, they have the benefit of not requiring the acquisition of the landmarks which makes the task less intensive (Dhingra, 2022). Recent techniques have commonly employed the use of a single RGB image to determine head postures (Ruiz et al., 2018; Yang et al., 2019;

H. Zhang et al., 2020). These works combine the use of deep learning models (such as: ResNet50, Convolution Neural Network (CNN) based networks, among others) with multi-loss techniques and attention mechanisms in order to estimate the head pose angles, (Dhingra, 2022).

Following this trend, there are now available commercial and open-source solutions aiming on detect and track the user face and estimate the head position, (Affectiva, 2023; Baltrusaitis et al., 2018; Cao et al., 2021; Nuitrack, 2019).

Concerning the open-source solutions available, OpenFace is the most advanced solution (Baltrusaitis et al., 2016). It offers a set of algorithms that can track and extract facial cues in real-time (15-20 fps depending on the computing device) by using an RGB camera. It is an open source solution capable of detecting the user face and outputs the facial landmarks, the head pose and eye gaze data, and facial action units, Figure 2.1.

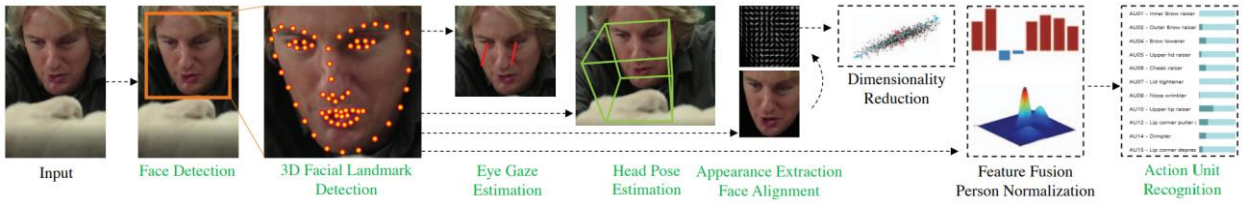


Figure 2.1: OpenFace facial behaviour analysis pipeline, including: landmark detection, head pose and eye gaze estimation, facial action unit recognition. The outputs of all of these are highlighted in green, (Baltrusaitis et al., 2018).

It uses Convolutional Experts Constrained Local Model (CE-CLM) for facial landmark detection and tracking. By using CE-CLM technique, it is able to extract the head pose, as CE-CLM internally uses a 3D representation of facial landmarks and projects them into the image using orthographic camera projections, Figure 2.2. Additionally, it is able to estimate the eye gaze by using a Constrained Local Neural Field (CLNF) landmark detector to detect the eyelids, iris, and the pupil.

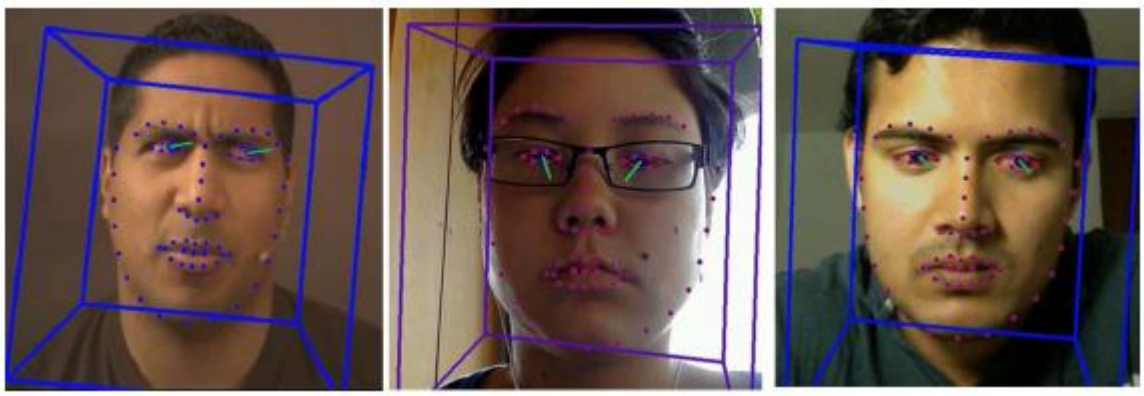


Figure 2.2: Head pose and eye gaze estimation on video sequences. Eye gaze estimation is highlighted in green and heap pose highlighted with the blue 3D box (Baltrusaitis et al., 2018).

This toolbox can be used through the OpenFace 2.0: Graphical User Interface (for Windows), and

command line (for Windows, Ubuntu, and Mac OS X). The source code can be integrated in any C++, C#, or MATLAB based project.

Another open source solution available is OpenPose (Cao et al., 2021), a real-time multi-person system to jointly detect human body, hand, facial, and foot key-points (in total 135 key-points) on single images, Figure 2.3. In more detail the main functionalities are: 2D real-time multi-person key-point detection with 15, 18, or 25-keypoint body/foot key-point estimation, 2x21-key-point hand landmark estimation, and 70-key-point facial landmark estimation; 3D real-time single person key-point detection with 3D triangulation from multiple single views and synchronization with Flir cameras available. In order to achieve this, the authors proposed a method that employs a non-parametric representation called Part Affinity Fields (PAFs) in combination with a multi-stage CNN architecture to learn to associate body parts with individuals in the image, (Cao et al., 2021).



Figure 2.3: Output of OpenPose, detecting body, foot, hand, and facial landmarks, retrieved from (Cao et al., 2021).

Runtime can vary on the number of detected people as well as on the hardware used. In fact, in order to be able to run the framework in real-time, it is highly recommend to use a machine with a dedicated Nvidia GPU with at least 1.6GB of dedicated VRAM plus 2.5GB of free RAM and a CPU with at least 8 cores as minimum requirements (Hidalgo et al., 2022).

Regarding the frameworks available in the commercial space, Nuitrack SDK (Nuitrack, 2019) offers similar features to OpenFace and OpenPose. Nuitrack is a 3D tracking cross-platform middleware capable of full-body skeletal tracking (up to 19 joints), with an agnostic 3D sensor independent by supporting the Microsoft Kinect V1 and V2, Orbbec™ Astra, Intel™ RealSense, among other 3D devices. The solution can track up to 6 skeletons at a time. In more detail, it can detect and track the user's face, estimate the head pose and landmark positions, and detect facial expressions, Figure 2.4.

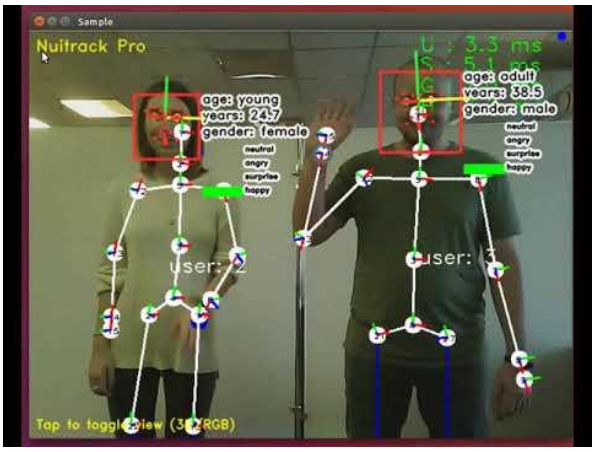


Figure 2.4: Nuitrack main features including facial tracking, head pose estimation, and body key-points estimation, (Nuitrack, 2019).

Conversely, pupil detection and eye gaze estimation is not included in the set of capabilities of Nuitrack SDK as is it included in OpenFace. The real-time performance can be up to 45 fps depending on the hardware configuration used (Nuitrack Frame Rate, 2018).

As it can be inferred, there are several approaches concerning the facial tracking and head pose estimation, some available as opensource tools and others available in the commercial space.

2.2.2 Human Action Recognition and Detection of Typical and Stereotypical Behaviours

Human Action Recognition (HAR) is a broad key field in computer vision. It is employed in video surveillance, human-machine interactions, and robot vision. There are numerous research directions in HAR. Traditionally, studies in HAR have mainly focused on the use of hand-crafted features (Dollar et al., 2005; Laptev et al., 2008) that can be provided by 2D cameras (RGB data) or 3D sensors (RGB + Depth information). Works with RGB data only, usually start by foreground extraction to detect the action region, to then extract features, by employing techniques such as Cuboids (Afsar et al., 2015) or HOG/HOF (Laptev et al., 2008). Probabilistic graphical models such as Hidden Markov Models (HMM) or the Dynamic Time Warping (DTW) algorithm, which measures the similarity between two temporal sequences, are used to classify the action (Afsar et al., 2015; Ahmad & Lee, 2006). (Afsar et al., 2015) proposed a method to automatically detect a subject and perform action recognition using a Hidden Markov Model and a bag of Words from RGB images from surveillance cameras. The Gaussian mixture model was used to perform back-ground subtraction. The model achieved an accuracy of 97.02% on their dataset with two classes (‘Sit’ and ‘Walk’). The authors in (Abdul-Azim & Hemayed, 2015) applied a novel method for action recognition using trajectory-based feature representation, tracking spatial-temporal interest points (STIPs) using a Cuboid detector to then build trajectories using SIFT-matching between

STIPs in consecutive frames. A Support Vector Machine (SVM) was used to classify the actions. The method was tested on the UCF sports dataset (Rodriguez et al., 2008), achieving an accuracy of 89.97%. Nonetheless, one of the major drawbacks of the 2D data is the absence of 3D structure from the scene (Pham et al., 2018). Furthermore, by only using the RGB information, these approaches tend to be more influenced by the environment lighting conditions.

With the rapid development of 3D sensing technology, several works propose the use of this hardware for different approaches (Aggarwal & Xia, 2014) in order to tackle the HAR problem. Furthermore, devices such as Microsoft Kinect (Microsoft, 2017) and Orbec Astra (ORBBEC, 2019) offer, through their SDKs, real-time skeleton detection, and tracking algorithms, an effective way to accurately describe actions. The main advantage of skeleton-based representations is the lower dimensionality of the data in comparison with representations based on RGB/RGB-D, al-lowing the action recognition systems to be computationally faster and less complex. Thus, several skeleton-based action recognition methods have been proposed (J. Wang et al., 2012; Xia et al., 2012), being a promising research direction. The authors in (B. Liu et al., 2017) proposed an approach that consisted of computing hand-crafted features by combining 3D joints moving trend and geometry. These features were used as input to a Linear SVM classifier. In that case, the method achieved an overall accuracy of 91.3% on a public dataset. Following this trend, some works in the literature have been using skeleton-based methods in order to detect some stereotypical behaviours in children with ASD (Goncalves et al., 2012a). (Goncalves et al., 2012a) used the Microsoft Kinect V1 and the DTW algorithm to detect the hand flapping or hand wave gesture, achieving an overall test accuracy of 51.0%. However, these approaches suffer from a lack of automation because of the dependency on hand-crafted features, requiring pre-processing the input data.

More recently, approaches based on Deep Learning techniques have been used and have yielded remarkable results in many recognition tasks. Most popular approaches use Recurrent Neural Networks with Long Short-Term Memory units (RNN-LSTM) for skeleton-based action recognition (Du et al., 2015; J. Liu et al., 2016; Shahroudy et al., 2016), achieving high-level performance and showing superior results when compared to classical hand-crafted approaches. However, they tend to overemphasize the temporal information and lose the spatial information of skeletons (Pham et al., 2019). Additionally, the amount of training time is considerably more when training LSTMs than CNN models, mainly because GPUs are optimized to process 2D data with extreme parallelism and speed, which CNNs utilize. (Bai et al., 2018) show that a CNN could achieve similar or even better performance than an RNN in many tasks such as speech recognition (Y. Zhang et al., 2016), some tasks of NLP (Gehring et al., 2017), among others. Following this trend, recent works in the literature have been employing Convolutional Neural Networks

(CNNs) to tackle the HAR problem (J. Wang et al., 2019). Many studies also indicate that CNNs are better than hand-crafted approaches since they have the ability to learn complex motion features (Du et al., 2015; C. Li et al., 2017; Song et al., 2017). However, most existing CNN-based methods use only the RGB component, only the depth component, or a combination of both components (RGB-D) as the input to learning models. Indeed, RGB-D images are informative for action recognition, but due to the large dimension of the input features, their use increases the computation complexity of these models rapidly. Additionally, the computational burden can be increased by adding more layers, partially affecting the real-time response during the test and training phases, especially when fine-tuning the CNN models. Thus, resulting in more complex and slower models, less practical for solving large-scale real-time problems. In order to tackle the dimensionality problem and computation complexity, works in (Ludl et al., 2019; Pham et al., 2018, 2019; Tasnim et al., 2021) use a skeleton-based representation with deep learning techniques for the recognition of different behaviours. The work proposed by (Ludl et al., 2019) consists of using skeleton-based representation data obtained from a 2D camera by using OpenPose (Cao et al., 2021) for skeleton detection in order to recognize three main actions: Idle, Wave, and Walking. The pipeline consists of first detecting the human by using a YOLO V3 (Redmon & Farhadi, 2018) followed by skeleton detection and tracking to finally recognizing the action. The method achieved an overall accuracy of 97.1% in their use case, recognition of three actions in a parking lot scenario to be employed in automated vehicles. Although it is a step forward to reduce the model complexity and running time, this approach requires dedicated hardware (a dedicated GPU to run the models was used). Thus, future improvements in the pipeline configuration are necessary to optimize system performance. Another approach, (Pham et al., 2018), consisted of using a 3D sensor, the Microsoft Kinect V2, to extract the user joints coordinates. Then, the 3D data was converted into image space. This method speeds up the overall pipeline since the Kinect SDK detects and extracts in real-time the user joints coordinates. The authors used different configurations of the ResNet architecture by varying the deepness of the model. The architecture was evaluated on a dedicated NVIDIA GTX 1080 Ti GPU achieving an overall accuracy of 99.9% and 99.8% on the MSR action 3D (W. Li et al., 2010) and KARD (Gaglio et al., 2015) datasets, respectively. Although the accuracy is high and the classification time for a skeleton sequence is around 0.128 s, it requires the use of a dedicated GPU with a final model configuration of 44 layers. The solutions proposed in (Ludl et al., 2019; Pham et al., 2018, 2019; Tasnim et al., 2021) used dedicated hardware, having a considerable computational burden in both training and testing phases, compromising real-time performance. Table 1 summarizes some recent approaches that employ skeleton-driven action classification.

Table 2.1: Some approaches of the state of the art that employs skeleton-driven action classification.

Author	ToC[1]	Hardware	Real-Time	CA[2]	NoA[3]	AA[4]	RR[5]
(Goncalves et al., 2012a)	Kinect V1	—	Yes	DTW	1 stereotyped action	HCI and ASD	51.0%
(C. Li et al., 2017)	Kinect V1	—	—	CNN	60 actions [34] 27 actions [48]	HCI	76.2% 88.1%
(B. Liu et al., 2017)	Kinect V2	—	—		Three sets of 8 actions each	Human-Computer Interaction (HCI)	91.3%
(Ke et al., 2017)	Kinect	—	—	CNN + MTLN	60 actions [34] 8 actions [50] 45 actions [51]	HCI	79.6% 93.6% 93.2%
(Pham et al., 2018)	Kinect V2	NVIDIA GTX 1080 Ti GPU	Yes	ResNet	20 actions [46] 18 actions [47] 60 actions [34]	HCI	99.9% 99.8% 78.2%
(Jazouli et al., 2019)	Kinect V1	—	Yes	$P Point-Cloud Recognizer	5 stereotyped actions	HCI and ASD	94.0%
(Ludl et al., 2019)	RGB Camera	Intel i7-8700 6 core CPU NVIDIA GTX 1080 GPU	Yes	CNN	3 actions	Autonomous Driving	97.1%
(Laraba et al., 2019)	Kinect V2	Intel i7-7800X 2 × NVIDIA GTX 1080 Ti	—	Pre-trained architectures of CNN and RNNs	49 actions [34]	HCI	82.1%
(Lee & Ahn, 2020)	RGB camera	NVIDEA JETSON XAVIER with dedicated Volta GPU	Yes (14 fps)	CNN	15 actions [34]	Human-Robot Interaction (HRI)	71.0%

ToC[1] – Type of Camera, CA[2] – Classification Algorithm, NoA[3] – Number of Actions, AA[4] – Application Area, RR[5] – Recognition Rate

Systems as mentioned could be helpful in making progress in improving the capacity for complex social interaction. Following this trend, it is possible to observe a recent paradigm shift where more research is focusing on investigating the efficacy of robots in ASD intervention (Begum et al., 2015; S. Costa, 2014)

and how to increase the robots autonomy (E. T. Bekele et al., 2013b; Chevalier et al., 2017).

The next subchapters will focus on describing some projects that reported consistent research regarding the use of robots as well as presenting research projects that used OPT to interact with children with ASD.

2.3 Objects based on Playware Technology (OPT)

Playware is defined as intelligent technology for children's play and playful experiences for the user. Henrik Lund suggested the term "playware" as a combination of intelligent hardware and software that aims at producing play and playful experiences among users (Lund et al., 2005). This technology emphasizes the role of interplay between morphology and control using processing, input, and output. The OPT attempts to combine the features of traditional playground and computer games, improving the child's health condition by augmenting the children's ability to engage in social and fantasy play. A system that adopts this type of technology can keep the children engaged by adapting the game to each child's personal preferences.

Few research projects have been using playware technology objects as a tool intervention with children with ASD. The project conducted by Henrik Lund consisted in designing a modular robotic playware, which is flexible in both set-up and activity building for the end-user, allowing easy creation of games. The development of this system consisted in building interactive quadratic shapes tiles Figure 2.5 with self-contained energy source and wireless communication (local and global), and with different games.



Figure 2.5: Interactive tiles as a playware object (Lund et al., 2009).

An initial study (Yannakakis et al., 2006) was conducted with the goal to determine the best way to create a fun experience to the user by focusing on the game opponent's behavior contributions to the real-time entertainment value of the game, Figure 2.6. A Bug-Smasher game was utilized to acquire data of human judgement on entertainment, being played by twenty eight children with ages between 8 and 10 years old (Yannakakis et al., 2006).



Figure 2.6: Children playing the activity with the developed tiles (Yannakakis et al., 2007).

A follow-up study was conducted involving children with ASD, (Lund, 2009). It consisted in mixing the tiles in different ways in order to produce new colours. The basic idea consisted in having three source tiles with predefined and fixed colours (red, green, and blue) and using the other 12 tiles, with the property that they can change their colours accordingly to their local neighbourhood. For example, if a normal tile is connected to a red source tile, the normal tile will become red just as its neighbour but with a lower intensity. If a blue source tile is connected to the last two tiles, the normal tile will change its colour to purple, blending the two colours (red and blue).

A set of experiments was performed with a group of 7 children with ASD, Figure 2.7. The authors concluded that the results provided by the research offers an interesting novel research direction to investigate playware as playful tools for cognitive challenged children, giving the children a playful experience and automatically investigate the playful interaction to provide insight (and possible a diagnosis).



Figure 2.7: Children with ASD playing with modular tiles (Lund et al., 2009).

Research has also been conducted in the field of biofeedback technology using social playware, by utilizing a wearable device (authors called bibs, Figure 2.8) that has an IMU and an RGB LED as feedback (Suzuki,

2014). Seven children diagnosed with autism, ranging in age from 7 to 13 years, participated in this study. The duration of time each child spent in group activities, such as holding or throwing a ball, was recorded. The authors observed that once the LEDs light up, the duration that the children spent in groups of two increased much longer than that of the period without being lit up the bibs. The use of illuminated bibs provided opportunities for children to initiate interactions and increased the likelihood of perceptual crossing.

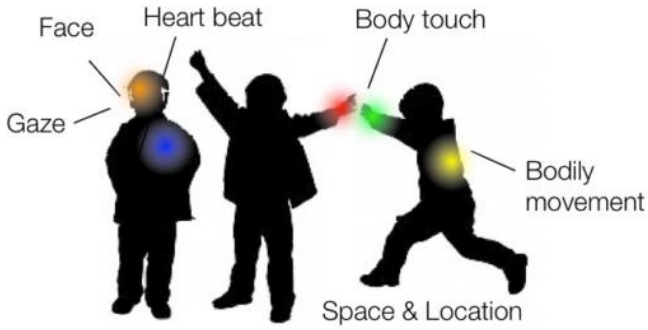


Figure 2.8: Social Playware: Cyber-physical system (Suzuki, 2014).

Another study consisted in the development of a serious game which utilizes a playware object similar to Lego-like building blocks augmented with electronic modules, (Barajas et al., 2017). The intended goal is to be used as a therapy tool to improve social and cognitive skill in children with ASD. An experiment involving nine children with ASD between the ages of 6 and 15 was conducted to evaluate the effects of computer-assisted play on solitary and collaborative play time. The results of the study indicated that the participants spent less time in solitary play when using a computer and had an increase in collaborative play time. Additionally, their performance in the computer game was significantly improved.

A project developed by (Azevedo et al., 2018) involves a serious game and a playware object, which is a physical component that acts as the game controller, enhancing the interactivity and allowing the user to manipulate a virtual object (serious game) through the use of a physical one, Figure 2.9. The playware object has six pressure buttons, each one showing an emoji with a specific facial expression. The six facial expressions of happiness, sadness, fear, anger, surprise, and neutral/normal were utilized in the three game activities of imitation, recognition, and storytelling.



Figure 2.9: The developed approach consisting on a serious game (running on a tablet) and the OPT developed (Azevedo et al., 2018).

An experiment comprising of four sessions was conducted with six children diagnosed with ASD. The aim of the study was to assess the effectiveness of utilizing a serious game in conjunction with a playware object as a tool for teaching emotional skills. The results of the study revealed that the children had a good understanding of the game's objective and instructions, and found the playware object to be an attractive and intuitive tool for interaction. Although the sample size was small, the results indicated that this type of tool could be an effective mediator for teaching emotional skills to children with ASD, as improvements were observed in the children's responses and engagement.

2.4 Human-Robot Interaction – Social Assistive Robotics and ASD

During support sessions, therapists usually use props/toys when interacting with the child. More recently, the use of social robots have been explored as they can provide a stimulus for children belonging to this spectrum, promoting social interaction (Pennisi et al., 2016; Ricks & Colton, 2010b). The robots used in ASD research possess different embodiments, varying from a non-humanoid appearance, to an animal like, and humanoid, Figure 2.10. Using different robot designs, projects like AuRoRA (Dautenhahn & Werry, 2004), LABO-1 (Dautenhahn & Werry, 2004), Robota (Dautenhahn & Billard, 2002), KASPAR (Dautenhahn et al., 2009b), Keepon (Kozima et al., 2008), and IROMEC (Ferrari et al., 2009), studied the potential role of robots as therapy tools for children with ASD.

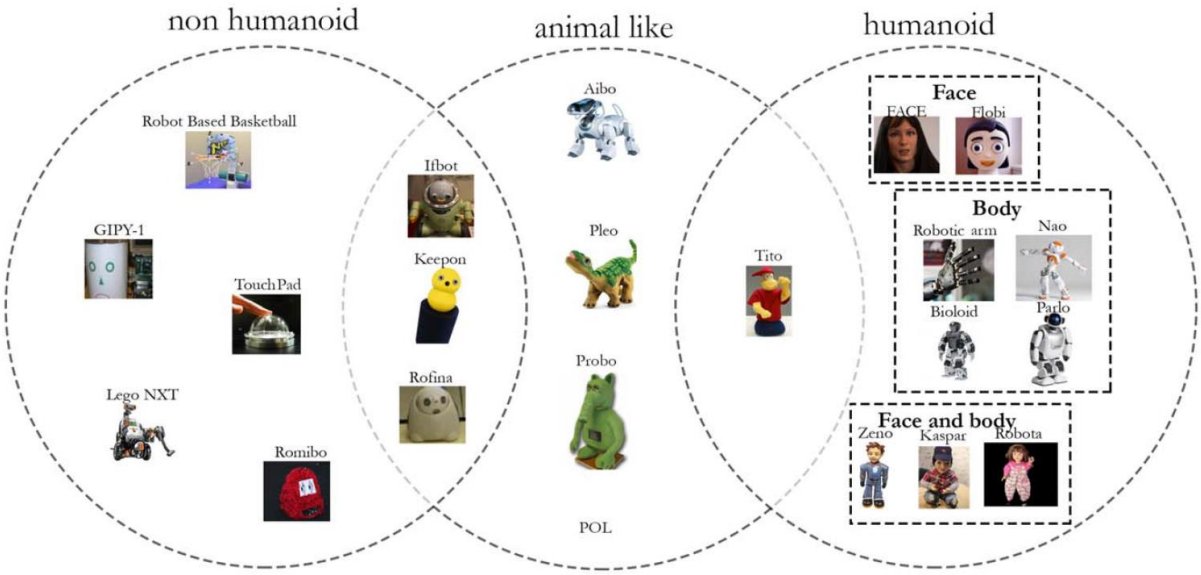


Figure 2.10: Social robots used in ASD research (Pennisi et al., 2016).

Project AuRoRA (AUtonomous RObotic platform as a Remedial tool for children with Autism) studied the potential role of mobile robots as therapy tools for children with ASD, (Dautenhahn & Werry, 2004).

The authors of the study examined the potential of utilizing mobile robots to foster social behaviours, including eye contact, joint attention, and imitation, which are crucial for human cognitive and developmental processes. They aimed to design robots that can be utilized by educators, caregivers, and parents to support children with ASD in educational settings or at home. During the study, the children were given the freedom to interact with the robot in any way they desired, with minimal interference from the researchers, only intervening if the child was at risk of damaging the robot or if the child turned off the robot.

During the trials, the children interacted freely with the mobile robot LABO-1, Figure 2.11, that used basic spatial approach and avoidance movements in turn-taking games. The results demonstrated that the children exhibited more attention towards the robot in terms of gaze and touch, and were more engaged in interactions compared to a non-robotic toy.

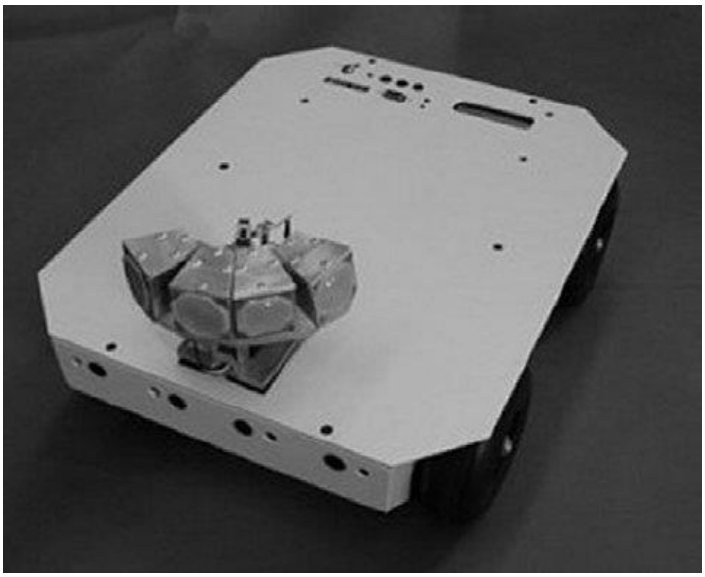


Figure 2.11: The robotic platform used, LABO (Dautenhahn & Werry, 2004).

The authors came to the conclusion that the mobile robot, LABO-1, was a safe option for children to interact with and demonstrated great adaptability in various contexts. Observations showed that children were motivated to engage in interactions with the robot for extended periods of time, and were more intrigued by the robot's reactive mode as opposed to its rigid and repetitive non-interactive behaviour. However, the main limitation of the non-humanoid robot used in the study was the limited number and type of interactions that could occur, specifically being limited to spatial approach and avoidance turn-taking games.

Inside the same project, the authors developed the humanoid robot Robota (Dautenhahn & Billard, 2002), offering additional interaction capabilities including mimicking movements of various body parts, such as hands and head, as well as more complex interactions involving sequences and combinations of actions. Robota, in the form of a doll-shaped robot, utilizes a motion tracking system to replicate upward movements of the user's left and right arm. The robot was controlled via Wizard-of-Oz (WoZ), as it needed to respond with very high precision movements of the children in unconstrained scenarios, in order to facilitate spontaneous interactions.

The objective of the study was to investigate the potential of humanoid robots, through the use of simple imitation and turn-taking games, to enhance social interaction skills in children with ASD. Additionally, the researchers sought to determine the ability of the robot to act as a facilitator of social interaction between children with ASD and their typically developing peers, by serving as an object of shared attention.



Figure 2.12: On the left: Several prototypes of Robota. On the right: Robota mimicking the user by lifting up its arm (Dautenhahn & Billard, 2002).

The findings of this study indicate that the level of engagement and interaction with the humanoid robot, Robota, increased over time in terms of eye gaze, touch, and imitation. Additionally, the results suggest that the robot was able to elicit imitative behaviour in children with ASD. The authors advocate for the use of non-invasive interfaces when interacting with the robot, highlighting the importance of the robot's appearance in eliciting a positive response from the child. Furthermore, the study suggests that limiting the robot's behaviours to a specific set would be more beneficial when testing with children with ASD, as it allows for clearer quantification of the child's reactions, particularly as children with ASD tend to have unique and individualized responses towards the robot (Robins et al., 2004).

Another project by the same authors is KASPAR (Figure 2.13), a child-sized humanoid robot with a minimalistic facial expression and functional arms capable of performing gestures. It has been used in several studies with children with ASD (Robins et al., 2005, 2008, 2009; Zaraki et al., 2018) as well as with typically developing children (Kose-Bagci et al., 2009; Wood et al., 2013). The purpose of KASPAR is to facilitate predictable and reliable interactions for children with ASD in educational and therapeutic settings. Additionally, the authors aimed for this robot to be adaptable to various levels of the autism spectrum, offering multi-modal embodied interactions tailored to the individual child's needs and preferences, such as varying levels of tactile interaction.



Figure 2.13: The KASPAR robot (Dautenhahn et al., 2009b).

The results of studies involving children with ASD and the humanoid robot KASPAR indicated that the children displayed a higher frequency of gaze switching during dyadic collaborative games when playing with the robot. Additionally, the children were more engaged and invested in the game, and collaborated better with their partners. Overall, the use of KASPAR was found to not only demonstrate important social interaction abilities, but also to elicit a high level of direct, physical engagement from the children. Furthermore, it was observed that the children seemed to generalize this behaviour to the experimenter. Keepon, Figure 2.14, was created to engage in nonverbal communication with children, with the aim of examining and evaluating psychological theories of the growth of social intelligence (Kozima et al., 2008).

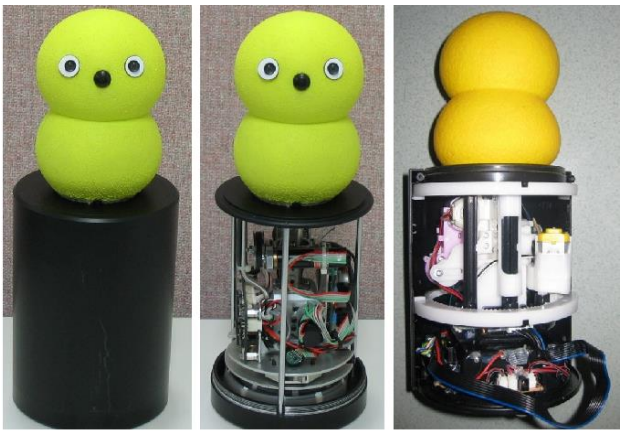


Figure 2.14: The Keepon robot with the details to the internal and external structures (Kozima et al., 2008).

The authors of the study used Keepon in a playroom at a day-care centre to observe and study the development of social intelligence in children with ASD and typically developing children. They also included the children's parents and professionals in the interactions, both in an unstructured setting and in organized groups. The goal of the experiments was to gather information on robot design, cognitive theory, psychological experimental design and intervention, and pedagogical practice by observing

Keepon's interactions with the children in a natural setting, and observing how the robot responded to significant social interactions made by the child.

The authors observed that Keepon simple design and predictable responses resulted in spontaneous and engaging dyadic interactions. They found that children were able to expand the interaction into interpersonal communication, with Keepon serving as a mediator for triadic interactions with adults and peers. Their main conclusion was that robots with minimal expressiveness can facilitate natural exchanges of mental states in children with ASD.

The IROMEC (Interactive Robotic Social Mediators as Companions) project had the goal to understand how play can aid in the development of children, specifically focusing on children who have difficulty engaging in play, (Ferrari et al., 2009). The project aimed to utilize interactive robotic social mediators as companions to encourage play and interaction by using elements such us visual and tactile perception, spatial awareness, and proprioception, Figure 2.15.



Figure 2.15: The IROMEC robot in the horizontal and vertical configurations with the main screen providing emotional feedback, (Ferrari et al., 2009).

In partnership with experts, ten scenarios were developed that utilize the features implemented in the IROMEC robot and its various modules. Through an iterative process with both therapists and teachers, for each scenario, a set of intervention and educational objectives were established. Once a play scenario was selected, the IROMEC robot was able to operate autonomously, taking into account the children's play needs and addressing various special needs such as cognitive, physical, and developmental impairments.

The robot adapted its behaviour to assist children in exploring and achieving increasingly intricate goals vital for social interaction dynamics. The authors aimed to utilize play to assist children in developing skills for social interactions, emotional expression, and engagement through the robot ability to adapt its

behaviour. The goal was to promote positive interactions and foster interpersonal relationships in the social-emotional domain.

The authors of the study found that the IROMEC system, which is programmable, was capable of offering various stimuli such as movement or sound. The behaviour of the robot could be adjusted based on the child's interaction with it. The robot was used as a shared point of focus that could foster interaction among children and adults, and it also helped to raise the level of complexity in the interaction, promoting a range of learning abilities (Ferrari et al., 2009).

The research discussed above provides an overview of relevant projects that have yielded significant findings and positive outcomes in the use of social robotics for children with ASD.

Since the present work uses a humanoid robot that is capable of synthesize facial cues as the main interaction tool, the next sections will address research projects that presented consistent research regarding the use of facial expressive robots when interacting with children with ASD and projects that use humanoid robots with capabilities to adapt the interaction with a child in some way.

2.4.1 Facial Expressive Robots

The use of facial expressive social robots in working with children with ASD has shown to be a promising area of research with many studies indicating positive results in terms of social interaction, communication and learning outcomes, (S. Costa et al., 2013; Mazzei, Lazzeri, et al., 2012; Soares et al., 2019). Figure 2.16 shows the humanoid robots used in ASD research with capability to display facial cues in order to promote the recognition of facial expressions with children with ASD.

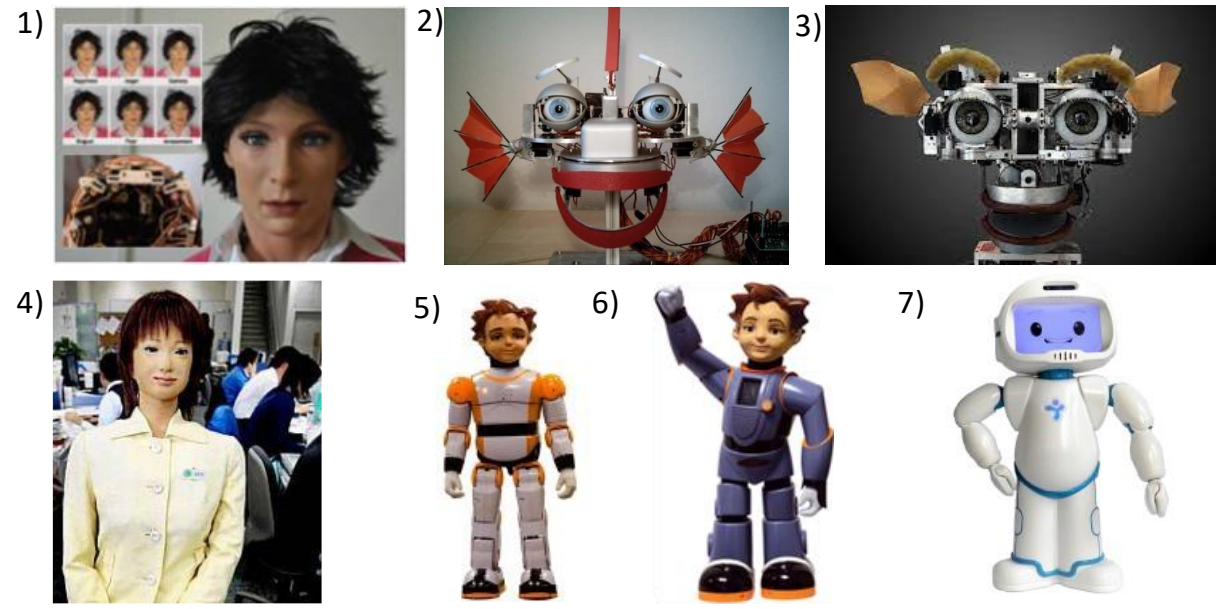


Figure 2.16: Humanoid robots used in ASD research – 1) FACE (Mazzei, Lazzeri, et al., 2012); 2) EDDIE (Sosnowski et al., 2006); 3) Kismet (Breazeal, 2000); 4) SAYA (Hashimoto et al., 2011); 5) Zeno R50 (S. Costa, 2014); 6) Zeno R25 (Robots4Autism, 2018); 7) QT (LuxAI, 2023).

FACE (Figure 2.16 – 1), developed by (Mazzei, Lazzeri, et al., 2012), is a FACE is a female android that is actuated by 32 servo motors moving the artificial skin, which allows human facial expressions to be re-created. Initial research consisted in evaluate the emotions (happiness, sadness, anger, disgust, fear, and surprise – the six basic emotions) expressed by the robot. Thus, the system was tested with five children with ASD and fifteen typically developing children. The results demonstrated that happiness, sadness and anger were correctly labelled with high accuracy for both children with ASD and typically developing children. Conversely, fear disgust, and surprise had not been labelled correctly, particularly by participants with ASD.

The work developed by (Sosnowski et al., 2006) consisted in developing a robotic head with 23 degrees of freedom (DoF), EDDIE (Figure 2.16 – 2). Each actuator was assigned to the particular action units of the Facial Action Coding System (FACS). The system was evaluated by twenty-four participants: eight children from five to eight years old and sixteen adults from twenty-five to forty-eight years old. The average recognition rate for all emotions was 57.0%.

Similar to EDDIE, Kismet (Breazeal, 2000) is able to engage humans in expressive social interaction. It has 15 DoF and it was designed with the possibility to process a variety of social signals from visual and audio channels and deliver social signals to the human with whom it was interacting (Figure 2.16 – 3). The values of the recognition rate were higher than those obtained with EDDIE.

SAYA (Hashimoto et al., 2011), a tele-operated android robot used in the role of a teacher, is capable of

express human-like facial expression and perform some communicative functions with its head and eye movements, Figure 2.16 – 4. The face has 19 DoF for generating facial expressions. In order to evaluate the designed facial expressions, the system was initially tested by twenty adults that watched videos of SAYA performing each of the six basic emotions. The results showed a high recognition rate, 97.3%.

ZECA (S. Costa et al., 2013), Zeno Engaging Children with Autism (Figure 2.16 – 5), is a humanoid robot that is used in *Robótica Autismo* project at University of Minho. This project, *Robótica Autismo*, emerged from a partnership between University of Minho and APPACDM association in Braga (*Associação Portuguesa de Pais e Amigos do Cidadão Deficiente)*, and seeks to use robotic platforms to improve social skills of individuals with ASD. ZECA was used in a study with the purpose to analyse the use of a humanoid robot as a tool to teach recognition and labelling of emotions. This humanoid robot developed by *Robokind* has the ability to simulate facial expressions, since the face has 11 Degrees of Freedom (DoF) and it is covered with a polymeric material called *Frubber*.

A perceptual study was conducted with the intention of providing the perceptions both from typically developing children and adults on the facial expressions displayed by ZECA, (S. Costa et al., 2013). Two groups were created – Group A with 42 participants aged between 8 and 10 years old and Group B with 61 participants aged between 18 and 59 years old. The participants watched videos of ZECA displaying only facial expressions and then displaying facial expressions and gestures corresponding to basic emotions. The results showed that the adding the gesture to the facial expression helped the recognition of the associated emotion, in particular for fear (45.2% to 73.8%) and anger (26.2% to 47.6%). Further results are detailed in Table 2.2.

(Soares et al., 2019), presented a child-robot interaction study with 45 children, where they had to recognize, and label emotions displayed by ZECA in three different game scenarios, involving imitation, recognition, and storytelling activities. The goal of these scenarios was to help a child with ASD to acquire knowledge about different emotions and to improve their recognition skills. The Imitate Me game scenario consisted in ZECA displaying a facial expression, and then asking the child for showing that same facial expression. The experimenter used a wireless keypad to classify the answer as either successful or unsuccessful. In the Recognize game scenario, ZECA would display a facial expression and the child had to choose the correct racquet that matched ZECA facial expression. In the Storytelling game scenario, social stories were told by ZECA, and the child had to identify the emotion felt by the character, in this case ZECA, in each story.

The 45 children were divided in three groups – G1: they interacted with the robot; G2 – the activities were done with the experimenter; G3 – had no support sessions. All children participated in the pre- and post-

test sessions, performed without the robot and it consisted in matching cards on which a man or a woman was showing one of five different emotions (happiness, sadness, anger, surprise, and fear).

The results showed that the children who interacted with the robot, presented better emotion recognition and higher abilities to imitate facial emotion expressions compared with the other two groups, particularly in the Storytelling game scenario. The authors concluded that this improvement might indicate that the robot helped the children understand the perspective of the character in the story, in this case ZECA.

The Zeno R25 is a compact and child-like humanoid robot with similarities to Zeno R50 as both were developed by the same company, (Robokind, 2023). The Zeno R25, similar to the Zeno R50, possesses the capability of creating realistic simulated facial expressions, (Figure 2.16 – 6).

A study conducted by (Schadenberg et al., 2017), consisted in using the humanoid robot Zeno R25 to quantify the emotion recognition accuracy with 28 typically developing children aged between 9 and 12. The facial expressions include joy, sadness, fear, anger, surprise, and disgust. The overall recognition accuracy was 61.5%. The authors conclude that not all facial expressions were recognised well with fear being often confused with surprise.

Another study, (Lecciso et al., 2021), consisted in using Zeno R25 to promote emotion recognition skills with children with ASD. A total of 12 children participated in the study, where 6 interacted with the robot (group 1) and the rest did the activity on a computer (group 2). The second group of children was trained using pre-record videos of typically developing peer performing the facial expressions. A total of four facial expressions were displayed by the robot, namely happiness, sadness, fear, and anger. The goals of the study were to compare the efficacy of the two support methods (robot vs. computer) and to see if there was a significant interaction between the two technological- based support methods (robot- vs. computer-based). The results showed that there is a similar impact by using a computer based support method and the robot. There was a similar level of engagement of the children with the robot and the video on the computer. However, the results showed a significant interaction effect regarding the expression of fear and anger for the children who interacted with the robot.

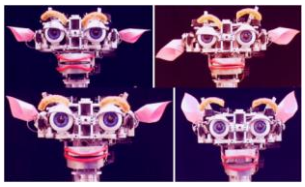
The QTRobot (Figure 2.16 – 7) is a recently developed social robot by LuxAI, (LuxAI, 2023). It has the ability to move its arms with multiple DoF. Different from the previous projects, QTRobot does not have moving parts on its face but a display is used instead. Additionally, it has the ability to move its head with some DoF.

As being recently developed there are not many studies exploring the use of this particular robot as a mediator in support sessions with children with ASD, (A. P. Costa et al., 2018; Zou et al., 2021).

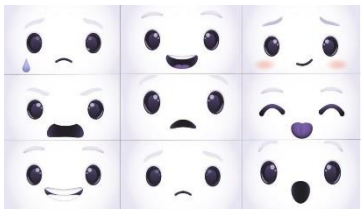
(A. P. Costa et al., 2018) have evaluated the usefulness of QTrobot by assessing children's attention, imitation, and presence of repetitive and stereotyped behaviours. A total of 15 children, aged between 4 and 14, all diagnosed with ASD participated in the study. It was compared the interaction of the children with a person and the robot. Children's gaze towards the person and the robot was assessed as a measure of attention, as well as children's imitation of the person or the robot. The presence of repetitive stereotyped behaviours was also quantified. The results showed the children directed more attention towards the robot, imitated the robot as much as with a person, and engaged in fewer repetitive or stereotyped behaviours with the robot in comparison with a person.

Table 2.2 summarizes some of the research projects that use facial expressive robots to promote social interaction and emotion recognition skills with children with ASD.

Table 2.2: Summary of the facial recognition rates for all the referred projects.

Robot	Sub	H	Sa	Su	D	F	A	N	Avg
FACE (Mazzei, Lazzeri, et al., 2012) 	TdC	93.3%	86.7%	40%	20%	46.7%	93%	-	61.1%
	ASD_C	100%	100%	40%	20%	0%	100%	-	60.1%
EDDIE (Sosnowski et al., 2006) 	TdC and Ad	58%	58%	75%	58%	42%	54%	-	57%
Kismet (Breazeal, 2000) 	TdC and Ad	82%	82%	82%	71%	47%	76%	-	73%
SAYA (Hashimoto et al., 2011) 	Ad	100%	100%	100%	92%	100%	92%	-	97.3%
ZECA (S. Costa et al., 2013) 	TdC	83.3%	97.6%	76.2%	-	45.2%	26.2%	85.7%	69%
	Ad	91.8%	91.8%	86.6%	-	77.1%	24.6%	90.2%	77%
	TdC[1]	81%	83.3%	69%	-	73.8%	47.6%	78.6%	72.2%
	Ad[1]	98.4%	88.5%	83.6%	-	93.4%	70.5%	91.8%	77%

*Sub – Subjects, TdC – Typically developing children, ASD_C – children with ASD, Ad – Adults, [1]with gestures, H – Happiness, Sa – Sadness, Su – Surprise, D – Disgust, F – Fear, A – Anger, N – Neutral, and Avg – Average

Robot	Sub	H	Sa	Su	D	F	A	N	Avg
Zeno R25 (Schadenberg et al., 2017) 	TdC	100%	46.0%	54.0%	46.0%	38.0%	85.0%	-	61.5%
QTRobot (LuxAI, 2023) 	-	-	–	-	-	-	-	-	-

*Sub – Subjects, TdC – Typically developing children, ASD_C – children with ASD, Ad – Adults,[1]with gestures, H – Happiness, Sa – Sadness, Su – Surprise, D – Disgust, F – Fear, A – Anger, N – Neutral, and Avg – Average

2.4.2 Social Assistive Robots and Autonomy

Most of the interactions with robots with children with ASD are often rigid, i.e., therapist or researcher fully controls the robot – Wizard of Oz (WoZ) method. Despite of being a successful method (Huijnen et al., 2016), it requires an additional operator other than the therapist that is engaging with the child in the triadic setup. However, some configurations (Soares et al., 2019), use a keypad where the researcher/therapist controls the interaction. Although it does not need an additional operator, this approach imposes a cognitive load on the researcher/therapist during the intervention. Additionally, systems that employ the WoZ method usually do not record the child's behaviour (body posture, facial expressions, eye gaze, among others) which might not be suitable to use in live HRI scenarios (Zaraki et al., 2018). Therefore, it is paramount to introduce some degree of autonomy, enabling the robot to perform some autonomous behaviours whilst keeping track of the interaction data. Following this trend, there has been a concerning in developing more adaptive approaches to interact with children with ASD. Some of the works in the literature (E. Bekele et al., 2014; E. T. Bekele et al., 2013a; Mazzei, Lazzeri, Billeci, Igliozzi, Mancini, Ahluwalia, Muratori, & de Rossi, 2011), use a combination of hardware, wearable devices and software algorithms to measure the affective states (e.g. eye gaze attention, facial expressions, vital signs, skin temperature, and skin conductance signals) of the child in order to adapt

the robot behaviour.

The work developed by (Mazzei, Lazzeri, Billeci, Igliozzi, Mancini, Ahluwalia, Muratori, & De Rossi, 2011) consisted in controlling the robot (FACE Figure 2.16-1) reactions and responses, by using a combination of hardware, wearable devices (e-shirt with built-in sensors), and software algorithms to measure the affective states (e.g. eye gaze attention, facial expressions, vital signs, skin temperature, and skin conductance signals) of individuals with ASD. The therapeutic room is equipped two with motorized pan/tilt/zoom cameras, mounted at opposite corners of the room. Six individuals aged between 15 and 22 were involved in the experiments with the goal to explore three specific robot-subject interactions: spontaneous behaviour and reactions of the participants to therapist presses in correlation with physiological data, focusing of the attention towards FACE's eye movements, and comprehension and imitation of FACE's expressions.

Preliminary results demonstrate that the platform was well accepted by the target group and can be consequently used as novel therapy for social skills training.

Bekele et al. (E. Bekele et al., 2014; E. T. Bekele et al., 2013a) developed and later evaluated a humanoid robotic system capable of intelligently managing joint attention prompts and adaptively respond based on measurements of gaze and attention. The experimental setup was constrained to a room with a network of cameras, used to infer gaze by tracking a hat wore by the participants with built-in infrared LEDs, Figure 2.17. The robot used was NAO, (Softbank Robotics, 2006), Figure 2.17.

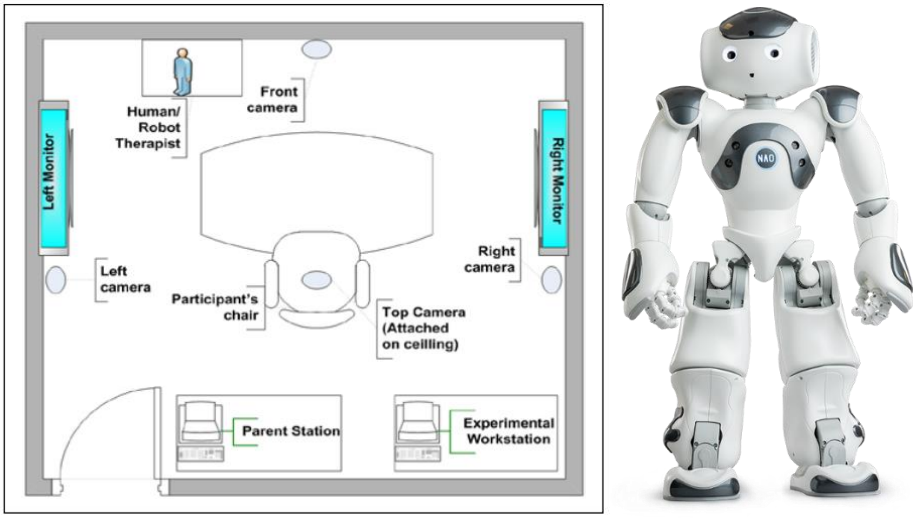


Figure 2.17: On the left: the room apparatus. On the right: the humanoid robot NAO used in the experiments with individuals with ASD (E. Bekele et al., 2014).

The authors concluded that the children with ASD directed their gaze towards to the robot when prompted with a question by the robot. Furthermore, the authors suggested that robotic systems endowed with

enhancements for successfully captivating the child attention might be capable to meaningfully enhance skills related to coordinated attention. However, the completion rate of the activity was 60% for the ASD group, mainly due to the fact of the willingness of these participants to wear the LED cap even for a brief interval of time (i.e. less than 15 min). This wearable device was a crucial part of the system since it allowed to track the children gaze during the activity. Thus, there is a need for the development of non-invasive systems and the use of such technologies with children with ASD with common sensory sensitivities (Rogers & Ozonoff, 2005).

In order to overcome the use of wearable devices, other projects (Esteban et al., 2017; Koutras et al., 2018) use an array of cameras and depth sensors that allows the robot to perform tasks autonomously. These sensory devices are usually precisely fixed, having into account the background and the lighting conditions, in a static structure around a table where the robot is placed, being limited to a controlled environment, Figure 2.18.

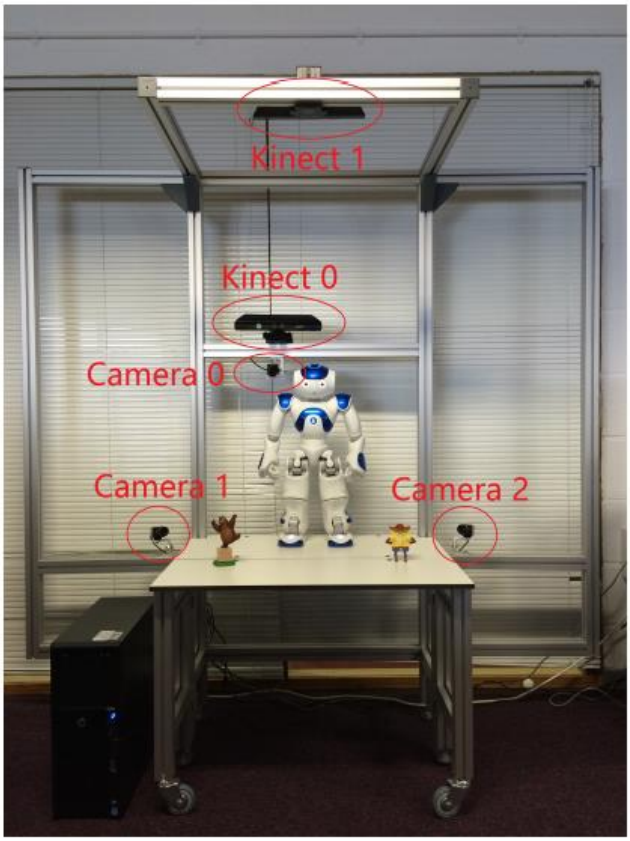


Figure 2.18: Experimental setup configuration with the fixed array of sensory devices (Esteban et al., 2017).

Single case experiments were conducted with 21 children with ASD, where the robot was used to initiate behaviours like arm movements (for imitation purposes), taking turns and triggering joint attention episodes.

Children directly interacted with the robot and the therapist mediated the interaction between the child.

The results indicated that scores have improved for imitation and for each of the turn-taking tasks but not for joint attention. The authors concluded that there is a need for personalized interventions that match each child's skill level and that a fully autonomous robot is not desirable, and aiming to achieve it is unrealistic, Figure 2.19. For this reason, a supervised autonomy approach is preferred (Esteban et al., 2017).

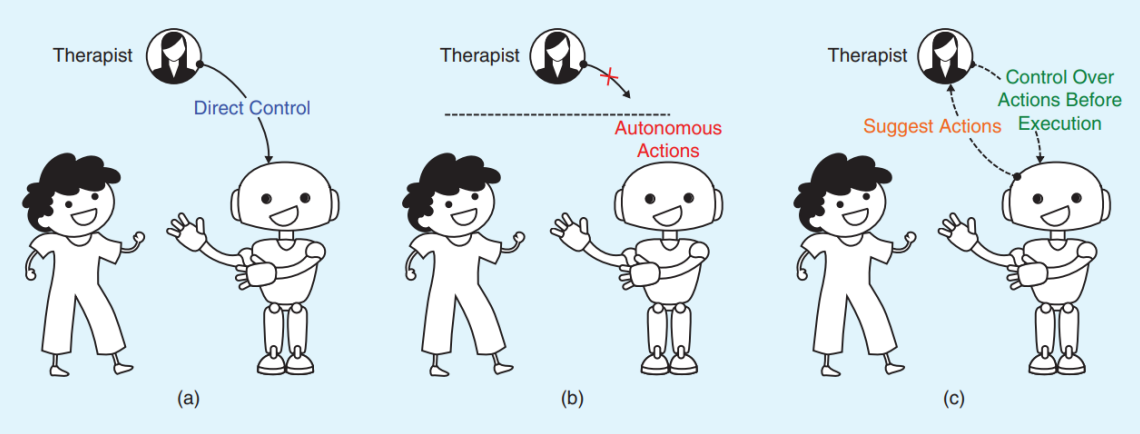


Figure 2.19: The different robot control paradigms: (a) Wizard of Oz, (b) full autonomy, and (c) supervised autonomy, (Esteban et al., 2017).

Another project consisted developed by (Melo et al., 2019), consisted in a room setup with mounted sensory devices, namely 3D cameras (Kinect) on the ceiling and walls to track the child, Figure 2.20. The children interacted with the robot and the therapist was the mediator that verbally controlled the robot behaviour with build detailed scripts for the different activities. That way the robot would react accordingly and adequately to the children's responses. In a study spanning 4 weeks, the robot participated in 121 therapy sessions with 18 children with ASD. The robot functioned autonomously while being overseen by two operators – one monitoring the robot's perception and the other responsible for its actions. The activities played consisted in rearrange blocks, completing puzzles, obstacle avoidance, among others. Concerning the results, all the children successfully completed more than 50% of the tasks and shorter sessions would correspond to higher rates of completions. The authors found difficult to tract the children's eye gaze since the setup includes a robot that moves constantly with the child. Nevertheless, the preliminary studies have revealed that a greater amount of eye gaze directed towards the robot is not necessarily indicative of a desirable behaviour. In fact, higher percentages of eye gaze were observed in sessions where children would stare at the robot with their faces close to touching, indicating inappropriate social interactions. Most of the children's speech was directed at the robot, which was a great indication of both engagement with the robot and enjoyment. The authors concluded that the study provided insight into the potential for a versatile system to be used in the therapy of children with ASD.

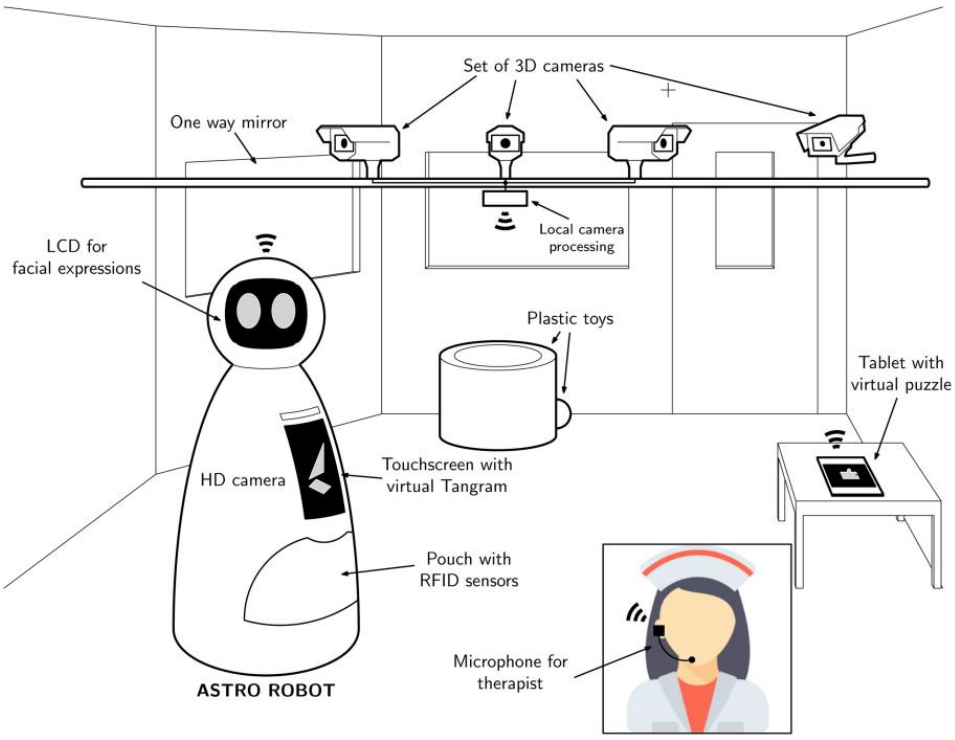


Figure 2.20: Diagram illustrating the room setup configuration with the robot used (ASTRO), (Melo et al., 2019).

Although the previous approaches are less invasive, they still are constrained to a room and laboratorial settings are usually not suitable for children with ASD, since it can be stressful, taking a considerable time for them to adapt to a new environment. Therefore, another approach (Zaraki et al., 2018) proposes the use social robots and how to introduce some autonomy in non-clinical environments. The authors use the Kaspar robot with framework called Sense-Think-Act that allows the robot to operate with some autonomy (under human supervision) with children in the real-world school settings, Figure 2.21. Contrary to most of the previous approaches the whole software framework runs on a laptop, which ensured the portability of the system. A Kinect sensor was used to determine the children body gestures and head posture.

The study included four children with ASD – two males and two females, with a mean age of 13 - who had varying levels of ability. These children participated in two different types of play with the Kaspar robot: two children played two games individually (dyadic child-robot interaction) and two children played two games in pairs (triadic child-robot interaction). The games were designed to promote the development of Visual Perspective Taking (VPT) skills in children with ASD. The experimenter would supervise the robot, overriding or permitting its behaviours by pressing a button on a Bluetooth key.



Figure 2.21: The trial involved children with ASD in a school setting playing dyadic and triadic games with the semi-autonomous Kaspar robot, (Zaraki et al., 2018).

The authors concluded that the system was successful in providing the robot with appropriate control signals to operate in a semi-autonomous manner. They further concluded that the architecture appears to have promising potential.

Another study examined a one-month, at-home intervention that used an autonomous social robot to improve the social communication skills of 12 children with ASD between the ages of 6 and 12, (Scassellati et al., 2018). The children participated in 30-minute sessions every day, engaging in activities such as emotional storytelling, perspective-taking, and sequencing with the assistance of a caregiver and the robot (Jibo).

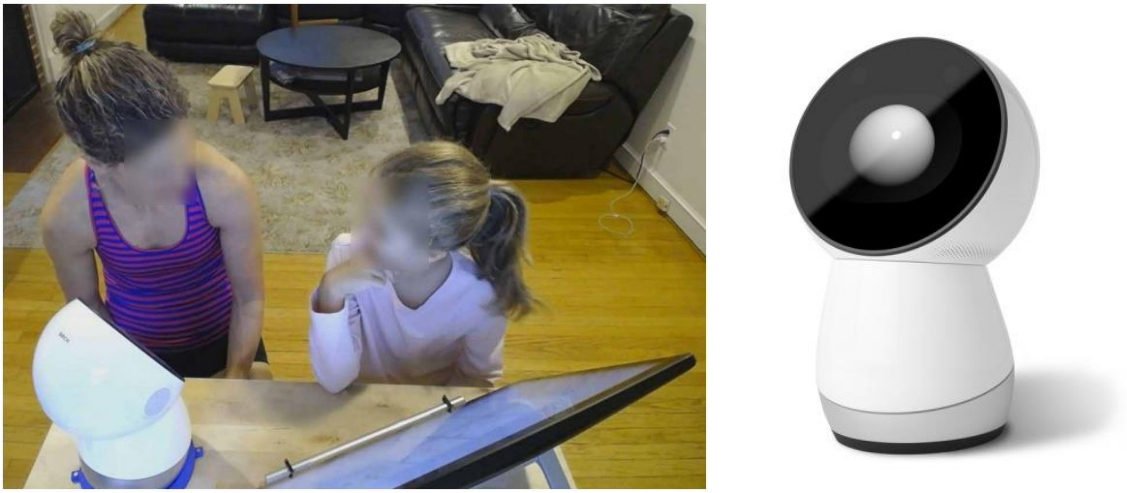


Figure 2.22: An example of a typical interaction between the robot (Jibo on the right), the child, and the caregiver, (Scassellati et al., 2018).

The system adjusted the level of difficulty of practice games based on the children's performance in the games. Additionally, an attention-tracking subsystem was used, which continuously estimated the children's attentional targets in the environment by analysing input from an RGB camera stream, estimating the children's head pose and orientation.

The main conclusions were: the robot helped to increase engagement, tailored the difficulty of the

activities to the child's abilities, and modelled positive social skills. Over the course of the month-long intervention, children showed improvement in their joint attention skills with adults when not in the presence of the robot, and these results were supported by caregiver evaluations. Caregivers reported that over time, they needed to provide less prompting, and overall communication had improved.

Although the studies on the of social assistive robots that can in some way tailor the interaction to the child's actions are still limited, and more research is needed, it offers a promising potential as it can be seen by the mentioned studies.

2.5 Summary and Conclusions

Social robots are becoming increasingly popular as a tool for supporting individuals with ASD. These robots are designed to provide social interaction and support in an interactive manner. They can be used to teach social skills, provide emotional support, and promote social engagement, (Scassellati, 2007). Most common approach to control the interaction flow of a support session with a robot and children with ASD is still the WoZ method. However, more studies, with different approaches, have been researching the introduction of some level of autonomy in a supervised manner in these systems in order to tailor the interaction to each child. The common relation between these studies is the triadic interaction: robot-child-therapist. The therapist or experimenter supervising the interaction flow of the support sessions. Despite of this advance, more research is needed. Moreover, none of the approaches in the literature explores the use of an adaptative hybrid approach, using a social robot and an OPT, to interact with children with ASD.

Following this idea, the present work proposes a supervised behavioural system architecture using a hybrid approach to allow the detection of the child behaviour and consequently adapt the robot to the child's action. The goal of this approach is to introduce some level of automation in a supervised manner. Additionally, the present work consists in the development of an OPT to be used as an add-on to the HRI with children with ASD in emotion recognition activities.

In more detail, the research this thesis deviates from the research discussed in the previous sections in the following ways:

- Using OPT in a HRI scenario for emotional skills training (Chapter 4). An initial phase of the research consisted in the development of a new OPT device and a pilot study was conducted in order to understand if the introduction of a new device to the interaction could be beneficial and stimulating. Then, a four-step study, consisting of a focus group, think aloud sessions, an online

questionnaire, and a pilot study was conducted. This study allowed the development of a new OPT and the redesign of the social stories scenarios taking into account the amount of stimuli, object disposition, and the colour palette.

- Developing a supervised behavioural framework with three main components – the perception, reasoning, and actuation layers – that allow to modify the interaction flow accordingly to the child's actions (Chapter 5). . A deep learning model was trained on non-verbal behaviour data in order to detect these behaviours. The framework was developed taking into account the portability functionality as the system should be as light, and unconstrained as possible in order to be transportable.
- Following a quantitative study involving children with high-functioning ASD (Chapter 6). The children performed the game scenarios with the supervised hybrid system (robot and the OPT), aimed to promote social interaction and emotion recognitions skill. The studies discussed in the chapter not only evaluated the performance in the activity (number of successful, unsuccessful answers, and unanswered prompts), but also considered behaviours that indicate social engagement, as described in the literature.

References

Abdul-Azim, H. A., & Hemayed, E. E. (2015). Human action recognition using trajectory-based representation. *Egyptian Informatics Journal*, *16*(2), 187–198. https://doi.org/10.1016/j.eij.2015.05.002

Affectiva. (2023). *Solutions - Affectiva : Affectiva*. https://www.affectiva.com/what/products/

Afsar, P., Cortez, P., & Santos, H. (2015). Automatic human action recognition from video using Hidden Markov Model. *Proceedings - IEEE 18th International Conference on Computational Science and Engineering, CSE 2015*, 105–109. https://doi.org/10.1109/CSE.2015.41

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. In *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2014.04.011

Ahmad, M., & Lee, S. W. (2006). HMM-based human action recognition using multiview image sequences. *Proceedings - International Conference on Pattern Recognition*, *1*, 263–266. https://doi.org/10.1109/ICPR.2006.630

American Psychiatric Association. (2013a). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic and Statistical Manual of Mental Disorders 4th Edition TR.*, 280.

https://doi.org/10.1176/appi.books.9780890425596.744053

American Psychiatric Association. (2013b). Diagnostic and Statistical Manual of Mental Disorders, DSM5. In *Arlington* (5th. Ed.). APA. https://doi.org/10.1176/appi.books.9780890425596.744053

Azevedo, J., Silva, V., Soares, F., Pereira, A. P., & Esteves, J. S. (2018). An application to promote emotional skills in children with autism spectrum disorders. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11243 LNCS*. https://doi.org/10.1007/978-3-030-02762-9_30

Bader, R. (2006). *Using Social Stories to Increase Emotion Recognition and Labeling in School-Age Children with Autism*. Alliant International University.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*. http://arxiv.org/abs/1803.01271

Baltrusaitis, T. (2014). Automatic facial expression analysis. In *Pattern recognition*. University of Cambridge.

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477553

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*. https://doi.org/10.1109/FG.2018.00019

Barajas, A. O., al Osman, H., & Shirmohammadi, S. (2017). A Serious Game for children with Autism Spectrum Disorder as a tool for play therapy. *2017 IEEE 5th International Conference on Serious Games and Applications for Health, SeGAH 2017*. https://doi.org/10.1109/SeGAH.2017.7939266

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology*, *84*, 490–515.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Begum, M., Serna, R. W., Kontak, D., Allspaw, J., Kuczynski, J., Yanco, H. A., & Suarez, J. (2015). Measuring the Efficacy of Robots in Autism Therapy. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 335–342. https://doi.org/10.1145/2696454.2696480

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of

an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013a). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013b). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Bennett, M., & Goodall, E. (2016). A Meta-analysis of DSM-5 Autism Diagnoses in Relation to DSM-IV and DSM-IV-TR. *Review Journal of Autism and Developmental Disorders*, *3*(2). https://doi.org/10.1007/s40489-016-0070-4

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., & Chetouani, M. (2014). Interactive Technologies for Autistic Children: A Review. *Cognitive Computation*, *6*(4), 722–740. https://doi.org/10.1007/s12559-014-9276-x

Breazeal, C. (2000). Sociable machines: Expressive social exchange between humans and robots. In *Expressive Social Exchange Between Humans and Robots*.

Cannon, W. B. (1927). The James-Lange theory of emotions: a critical examination and an alternative theory. *The American Journal of Psychology*, *39*, 106–124. https://doi.org/10.2307/1415404

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

CDC. (2022a). *Autism and Developmental Disabilities Monitoring (ADDM) Network | CDC*. https://www.cdc.gov/ncbddd/autism/addm.html

CDC. (2022b). *Diagnostic Criteria | Autism Spectrum Disorder (ASD) | CDC*. https://www.cdc.gov/ncbddd/autism/hcp-dsm.html

Chevalier, P., Li, J. J., Ainger, E., Alcorn, A. M., Babovic, S., Charisi, V., Petrovic, S., Schadenberg, B. R., Pellicano, E., & Evers, V. (2017). Dialogue Design for a Robot-Based Face-Mirroring Game to Engage Autistic Children with Emotional Expressions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10652 LNAI*, 546–555. https://doi.org/10.1007/978-3-319-70022-9_54

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. In *Child Development* (Vol. 75, Issue 2). https://doi.org/10.1111/j.1467-8624.2004.00673.x

Costa, A. P., Charpiot, L., Lera, F. R., Ziafati, P., Nazarikhorram, A., van der Torre, L., & Steffgen, G. (2018). More Attention and Less Repetitive and Stereotyped Behaviors using a Robot with Children with Autism. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2018.8525747

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Costa, S., Soares, F., & Santos, C. (2013). Facial expressions and gestures to convey emotions with a humanoid robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 542–551. https://doi.org/10.1007/978-3-319-02675-6_54

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders*, *2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

D., C. P., Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., Souberman, E., & Wertsch, J. v. (1979). L. S. Vygotsky: Mind in Society. The Development of Higher Psychological Processes. *The American Journal of Psychology*, *92*(1). https://doi.org/10.2307/1421493

Daily, S. B., James, M. T., Cherry, D., Porter, J. J., Darnell, S. S., Isaac, J., & Roy, T. (2017). Affective Computing: Historical Foundations, Current Applications, and Future Trends. In *Emotions and Affect in Human Factors and Human-Computer Interaction*. https://doi.org/10.1016/B978-0-12-801851-4.00009-4

Damasio, A. (1994). Descartes' error: Emotion, rationality and the human brain. In *New York: Putnam*.

Darwin, C. (1872). The expression of the emotions in man and animals. *London, UK: John Marry*, 374. https://doi.org/10.1037/h0076058

Dautenhahn, K., & Billard, A. (2002). Games children with autism can play with Robota, a humanoid robotic doll. *Universal Access and Assistive Technology*, 179–190. https://doi.org/10.1007/978-1-4471-3719-1_18

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009a). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3–4), 369–397. https://doi.org/10.1080/11762320903123567

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009b). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3–4), 369–397. https://doi.org/10.1080/11762320903123567

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, *12*(1), 1–35. https://doi.org/10.1075/pc.12.1.03dau

Dhingra, N. (2022). LwPosr: Lightweight Efficient Fine Grained Head Pose Estimation. *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022*. https://doi.org/10.1109/WACV51458.2022.00127

Dollar, P., Rabaud, V., Cottrell, G., & Belongie, S. (2005). Behavior recognition via sparse spatio-temporal features. *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 65–72. https://doi.org/10.1109/VSPETS.2005.1570899

Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2015.7298714

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, *19*.

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H. L., Coeckelbergh, M., Costescu, C., David, D., de Beir, A., Fang, Y., Ju, Z., Kennedy, J., Liu, H., Mazel, A., Pandey, A., Richardson, K., Senft, E., Thill, S., … Ziemke, T. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn*, *8*(1), 18–38. https://doi.org/10.1515/pjbr-2017-0002

Ferrari, E., Robins, B., & Dautenhahn, K. (2009). Therapeutic and educational objectives in robot assisted play for children with autism. *IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man)*, 108–114. https://doi.org/10.1109/ROMAN.2009.5326251

Gaglio, S., Re, G. lo, & Morana, M. (2015). Human Activity Recognition Process Using 3-D Posture Data. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2014.2377111

Gehring, J., Auli, M., Grangier, D., & Dauphin, Y. N. (2017). A convolutional encoder model for neural machine translation. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 123–135. https://doi.org/10.18653/v1/P17-1012

Goncalves, N., Rodrigues, J. L., Costa, S., & Soares, F. (2012a). Automatic detection of stereotyped hand

flapping movements: Two different approaches. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2012.6343784

Goncalves, N., Rodrigues, J. L., Costa, S., & Soares, F. (2012b). Preliminary study on determining stereotypical motor movements. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1598–1601. https://doi.org/10.1109/EMBC.2012.6346250

Gray, C. A., & Garand, J. D. (1993). Social Stories: Improving Responses of Students with Autism with Accurate Social Information. *Focus on Autism and Other Developmental Disabilities*, *8*(1). https://doi.org/10.1177/108835769300800101

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. In *Autism*. https://doi.org/10.1177/1362361313476767

Hanley-Hochdorfer, K., Bray, M. A., Kehle, T. J., & Elinoff, M. J. (2010). Social stories to increase verbal initiation in children with autism and asperger's disorder. *School Psychology Review*, *39*(3). https://doi.org/10.1080/02796015.2010.12087767

Happé, F., Briskman, J., Frith, U., Happé, F., & Frith, U. (2001). Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: II. Real-life skills and preferences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(3), 309–316. https://doi.org/10.1017/S0021963001006916

Hashimoto, T., Kobayashi, H., & Kato, N. (2011). Educational system with the android robot SAYA and field trial. *IEEE International Conference on Fuzzy Systems*, *8*(3), 766–771. https://doi.org/10.1109/FUZZY.2011.6007430

Hidalgo, G., Cao, Z., Simon, T., Wei, S.-E., Raaj, Y., Joo, H., & Sheikh, Y. (2022). *CMU-Perceptual-Computing-Lab/openpose - System Requirements*. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/installation/0_index.md#operating-systems-requirements-and-dependencies

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion: a further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(February), 671–680. https://doi.org/10.1111/j.1469-7610.1986.tb01836.x

HTC. (2023). *VIVE - VR Headsets, Games, and Metaverse Life | United States*. https://www.vive.com/us/

Hudlicka, E. (2008). Affective Computing for Game Design. *Scientist*, 5–12.

Hudock, R. L., Kashima-Ellingson, Y., & Bellini, S. (2011). Increasing Responses to Verbal Greetings in

Children With Autism Spectrum Disorders: A Comparison of the Effectiveness of Social Story and Generic Story Interventions. In *School Psychology Forum* (Vol. 5, Issue 3).

Huijnen, C. A. G. J., Lexis, M. A. S., Jansens, R., & de Witte, L. P. (2016). Mapping Robots to Therapy and Educational Objectives for Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-016-2740-6

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205. http://www.jstor.org/stable/2246769

Jazouli, M., Majda, A., Merad, D., Aalouane, R., & Zarghili, A. (2019). Automatic detection of stereotyped movements in autistic children using the Kinect sensor. *International Journal of Biomedical Engineering and Technology*. https://doi.org/10.1504/IJBET.2019.097621

Johnson, C. P., Myers, S. M., Lipkin, P. H., Cartwright, J. D., Desch, L. W., Duby, J. C., Elias, E. R., Levey, E. B., Liptak, G. S., Murphy, N. A., Tilton, A. H., Lollar, D., Macias, M., McPherson, M., Olson, D. G., Strickland, B., Skipper, S. M., Ackermann, J., del Monte, M., … Yeargin-Allsopp, M. (2007). Identification and evaluation of children with autism spectrum disorders. In *Pediatrics* (Vol. 120, Issue 5, pp. 1183–1215). https://doi.org/10.1542/peds.2007-2361

Kahn, P. H., Gary, H. E., & Shen, S. (2013). Children's Social Relationships With Current and Near-Future Robots. *Child Development Perspectives*, *7*(1), 32–37. https://doi.org/10.1111/cdep.12011

Kanner, L. (1943). Autistic disturbances of affective contact. In *Nervous Child* (Vol. 2, pp. 217–250). https://doi.org/10.1105/tpc.11.5.949

Ke, Q., Bennamoun, M., An, S., Sohel, F., & Boussaid, F. (2017). A new representation of skeleton sequences for 3D action recognition. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 4570–4579. https://doi.org/10.1109/CVPR.2017.486

Knight, V., McKissick, B. R., & Saunders, A. (2013). A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. In *Journal of Autism and Developmental Disorders* (Vol. 43, Issue 11, pp. 2628–2648). https://doi.org/10.1007/s10803-013-1814-y

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, *23*(14). https://doi.org/10.1163/016918609X12518783330360

Koutras, P., Potamianos, G., Efthymiou, N., Maragos, P., Filntisis, P. P., & Tsiami, A. (2018). *Multi3: Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots*. https://doi.org/10.1109/icra.2018.8461210

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2008). Keepon. *International Journal of Social Robotics*,

1(1), 3–18. https://doi.org/10.1007/s12369-008-0009-8

Lange, C. (1887). *Ueber gemuthsbewgungen* (3rd ed., Vol. 8).

Laptev, I., Marszałek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587756

Laraba, S., Tilmanne, J., & Dutoit, T. (2019). Leveraging Pre-trained CNN Models for Skeleton-Based Action Recognition. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11754 LNCS* (Issue November). Springer International Publishing. https://doi.org/10.1007/978-3-030-34995-0_56

Lecciso, F., Levante, A., Fabio, R. A., Caprì, T., Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., & Petrocchi, S. (2021). Emotional Expression in Children With ASD: A Pre-Study on a Two-Group Pre-Post-Test Design Comparing Robot-Based and Computer-Based Training. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.678052

Lee, J., & Ahn, B. (2020). Low-Cost RGB Camera and Mobile Robot Platform. *Sensors (Switzerland)*, *20*, 1–12. https://doi.org/doi:10.3390/s20102886

Li, C., Hou, Y., Wang, P., & Li, W. (2017). Joint Distance Maps Based Action Recognition with Convolutional Neural Networks. *IEEE Signal Processing Letters*, *24*(5), 624–628. https://doi.org/10.1109/LSP.2017.2678539

Li, W., Zhang, Z., & Liu, Z. (2010). Action recognition based on a bag of 3D points. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*. https://doi.org/10.1109/CVPRW.2010.5543273

Lima, C. B. de, & Levy, P. de Q. (2015). A Perturbação do Espetro do Autismo. In *Perturbações do Neurodesenvolvimento: Manual de orientações diagnósticas e estratégias de intervenção*.

Liu, B., Yu, H., Zhou, X., Tang, D., & Liu, H. (2017). Combining 3D joints Moving Trend and Geometry property for human action recognition. *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*. https://doi.org/10.1109/SMC.2016.7844262

Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016). Spatio-temporal LSTM with trust gates for 3D human action recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9907 LNCS*, 816–833. https://doi.org/10.1007/978-3-319-46487-9_50

Loveland, K. A., Tunali-Kotoski, B., Chen, Y. R., Ortegon, J., Pearson, D. A., Brelsford, K. A., & Gibbs, M. C. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and*

Psychopathology, *9*(3). https://doi.org/10.1017/s0954579497001351

Ludl, D., Gulde, T., & Curio, C. (2019). Simple yet efficient real-time pose-based action recognition. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*. https://doi.org/10.1109/ITSC.2019.8917128

Lund, H. H. (2009). Modular playware as a playful diagnosis tool for autistic children. *2009 IEEE International Conference on Rehabilitation Robotics, ICORR 2009*, 899–904. https://doi.org/10.1109/ICORR.2009.5209606

Lund, H. H., Dam Pedersen, M., & Beck, R. (2009). Modular robotic tiles: Experiments for children with autism. *Artificial Life and Robotics*, *13*(2), 394–400. https://doi.org/10.1007/s10015-008-0623-4

Lund, H. H., Klitbo, T., & Jessen, C. (2005). Playware technology for physically activating play. *Artificial Life and Robotics*, *9*(4), 165–174. https://doi.org/10.1007/s10015-005-0350-z

LuxAI. (2023). *LuxAI - Award winning social robots for autism and special needs education*. https://luxai.com/

Málaga, I., Lago, R. B., Antonio, H.-F., Álvarez-Álvarez, N., Oreña-Ansonera, V. A., & Baeza-Velasco, M. (2019). Prevalencia de los trastornos del espectro autista en niños en Estados Unidos, Europa y España: coincidencias y discrepancias. *Medicina (Buenos Aires)*, *79*, 1–6. http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0025-76802019000200003&lng=es&tlng=es.

Marques, C. (1998). *Perturbações do espectro do autismo: ensaio de uma intervenção construtivista e desenvolvimentalista com mães*. Universidade de Coimbra.

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., Muratori, F., & de Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2011*, 4515–4518. https://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., Muratori, F., & De Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Engineering in Medicine, Proceedings of the Annual International Conference of the IEEE*, *2011*, 4515–4518. https://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Hanson, D., & de Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and*

EMBS International Conference on Biomedical Robotics and Biomechatronics, 195–200. https://doi.org/10.1109/BioRob.2012.6290687

McDaniel, J., Brady, N. C., & Warren, S. F. (2022). Effectiveness of Responsivity Intervention Strategies on Prelinguistic and Language Outcomes for Children with Autism Spectrum Disorder: A Systematic Review and Meta-Analysis of Group and Single Case Studies. *Journal of Autism and Developmental Disorders*, *52*(11). https://doi.org/10.1007/s10803-021-05331-y

Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., Seidel, H. P., Rhodin, H., Pons-Moll, G., & Theobalt, C. (2020). XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera. *ACM Transactions on Graphics*, *39*(4). https://doi.org/10.1145/3386569.3392410

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H. P., Xu, W., Casas, D., & Theobalt, C. (2017). VNect: Real-time 3D human pose estimation with a single RGB camera. *ACM Transactions on Graphics*, *36*(4). https://doi.org/10.1145/3072959.3073596

Melo, F. S., Sardinha, A., Belo, D., Couto, M., Faria, M., Farias, A., Gambôa, H., Jesus, C., Kinarullathil, M., Lima, P., Luz, L., Mateus, A., Melo, I., Moreno, P., Osório, D., Paiva, A., Pimentel, J., Rodrigues, J., Sequeira, P., … Ventura, R. (2019). Project INSIDE: towards autonomous semi-unstructured human–robot social interaction in autism therapy. *Artificial Intelligence in Medicine*, *96*. https://doi.org/10.1016/j.artmed.2018.12.003

Meta. (2023). *Meta Quest Pro: Our Most Advanced New VR Headset | Meta Store | Meta Store*. https://www.meta.com/quest/quest-pro/

Microsoft. (2017). *Developing with Kinect*. https://developer.microsoft.com/en-us/windows/kinect/develop

Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(4). https://doi.org/10.1109/TPAMI.2008.106

Myles, Brenda Smith; Swanson, Terri Cooper, & Holverstott, J. (2013). Autism spectrum disorders: a handbook for parents and professionals. *Choice Reviews Online*. https://doi.org/10.5860/choice.45-5017

Nuitrack. (2019). *Nuitrack Full Body Skeletal Tracking Software - Kinect replacement for Android, Windows, Linux, iOS, Intel RealSense, Orbbec*. https://nuitrack.com/

Nuitrack Frame Rate. (2018). *Skeleton tracking specs ? 60 - 90 fps? - Nuitrack SDK - Nuitrack*. https://community.nuitrack.com/t/skeleton-tracking-specs-60-90-fps/916

OptiTrack. (2023). *OptiTrack - Motion Capture Systems*. https://optitrack.com/

ORBBEC. (2019). *Astra Series – Orbbec*. https://orbbec3d.com/product-astra-pro/

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. In *Autism Research* (Vol. 9, Issue 2, pp. 165–183). https://doi.org/10.1002/aur.1527

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2018). Exploiting deep residual networks for human action recognition from skeletal data. *Computer Vision and Image Understanding*. https://doi.org/10.1016/j.cviu.2018.03.003

Pham, H. H., Salmane, H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2019). Spatio–temporal image representation of 3D skeletal movements for view-invariant action recognition with deep convolutional neural networks. *Sensors (Switzerland)*, *19*(8). https://doi.org/10.3390/s19081932

Piaget. (1976). *Piaget and His School* (B. Inhelder, H. H. Chipman, & C. Zwingmann, Eds.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-46323-5

Picard, R. W. (1997). Affective Computing. In *Pattern Recognition* (Vol. 73). https://doi.org/10.1007/BF01238028

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. http://arxiv.org/abs/1804.02767

Ricks, D. J., & Colton, M. B. (2010). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. https://doi.org/10.1109/ROBOT.2010.5509327

Robins, B., Dautenhahn, K., & Dickerson, P. (2009). From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009*, 205–211. https://doi.org/10.1109/ACHI.2009.32

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., & Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *2005*. https://doi.org/10.1109/ROMAN.2005.1513864

Robins, B., Dickerson, P., Stribling, P., & Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction Studies*, *5*(2). https://doi.org/10.1075/is.5.2.02rob

Robins, B., Ferrari, E., & Dautenhahn, K. (2008). Developing scenarios for robot assisted play. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive*

Communication, RO-MAN. https://doi.org/10.1109/ROMAN.2008.4600663

Robokind. (2023). *Robokind | Advanced Social Robots*. https://www.robokind.com/

Robots4Autism. (2018). *Milo Robot*. https://robots4autism.com/

Rodriguez, M. D., Ahmed, J., & Shah, M. (2008). Action MACH: A spatio-temporal maximum average correlation height filter for action recognition. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587727

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Ruiz, N., Chong, E., & Rehg, J. M. (2018). Fine-grained head pose estimation without keypoints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2018-June*. https://doi.org/10.1109/CVPRW.2018.00281

Samuels, R., & Stansfield, J. (2012). The effectiveness of social stories™ to develop social interactions with adults with characteristics of autism spectrum disorder. *British Journal of Learning Disabilities*, *40*(4). https://doi.org/10.1111/j.1468-3156.2011.00706.x

Sanders, J. L. (2009). Qualitative or quantitative differences between asperger's disorder and autism? Historical considerations. *Journal of Autism and Developmental Disorders*, *39*(11). https://doi.org/10.1007/s10803-009-0798-0

Sansosti, F. J., Powell-Smith, K. A., & Kincaid, D. (2004). A Research Synthesis of Social Story Interventions for Children With Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities*, *19*(4). https://doi.org/10.1177/10883576040190040101

Scassellati, B. (2007). How social robots will help us to diagnose , treat , and understand autism. *Robotics Research*, 552–563. https://doi.org/10.1007/978-3-540-48113-3_47

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., & Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, *3*(21). https://doi.org/10.1126/SCIROBOTICS.AAT7544

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review.*, *69(5)*, 379–399. https://doi.org/10.1037/h0046234

Schadenberg, B. R., Heylen, D. K. J., & Evers, V. (2017). Affect bursts to constrain the meaning of the facial expressions of the humanoid robot Zeno. *CEUR Workshop Proceedings*, *2059*.

Schopler, E., & van Bourgondien, M. E. (2020). Treatment and education of autistic and related communication handicapped children. In *Autistic Adults at Bittersweet Farms*.

https://doi.org/10.4324/9781003075240-8

Shahroudy, A., Liu, J., Ng, T. T., & Wang, G. (2016). NTU RGB+D: A large scale dataset for 3D human activity analysis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 1010–1019. https://doi.org/10.1109/CVPR.2016.115

Shotton, J., Sharp, T., Fitzgibbon, A., Blake, A., Cook, M., Kipman, A., Finocchio, M., & Moore, R. (2013). Real-Time human pose recognition in parts from single depth images. *Communications of the ACM*, *56*(1). https://doi.org/10.1145/2398356.2398381

Siegel, B. (2008). *O mundo da criança com autismo: compreender e tratar perturbações do espectro do autismo.* Porto Editora.

S.Messinger, D. (2014). *Affective Computing, Emotional Development, and Autism*. 516–536.

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

Softbank Robotics. (2006). *NAO Robot*. https://www.ald.softbankrobotics.com/en/robots/nao

Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Review*, *81*(2). https://doi.org/10.1037/h0036128

Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2017). An end-to-end spatio-temporal attention model for human action recognition from skeleton data. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*.

Sosnowski, S., Kuehnlenz, K., & Buss, M. (2006). EDDIE - An Emotion Display with Dynamic Intuitive Expressions. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3113–3118. https://doi.org/10.1109/IROS.2006.282208

Suzuki, K. (2014). Social Playware: Device-mediated social interaction for therapeutic activities. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*. https://doi.org/10.1109/ROMAN.2014.6926232

Tasnim, N., Islam, M. K., & Baek, J. H. (2021). Deep learning based human activity recognition using spatio-temporal image formation of skeleton joints. *Applied Sciences (Switzerland)*, *11*(6). https://doi.org/10.3390/app11062675

Toshev, A., & Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

https://doi.org/10.1109/CVPR.2014.214

Vicon. (2023). *Vicon | Award Winning Motion Capture Systems*. https://www.vicon.com/

Wang, H., Hu, J., & Deng, W. (2018). Face Feature Extraction: A Complete Review. *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2017.2784842

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang, J., Liu, Z., Wu, Y., & Yuan, J. (2012). Mining actionlet ensemble for action recognition with depth cameras. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2012.6247813

Wang, Y. A., Yu, X., Silverman, P. M., Harris, R. L., & Edward, H. (2010). *NIH Public Access*. *385*(1), 22–29. https://doi.org/10.1016/j.jmb.2008.10.054.The

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. In *Neuroscience and Biobehavioral Reviews* (Vol. 25, Issue 4, pp. 287–295). https://doi.org/10.1016/S0149-7634(01)00014-8

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29. https://doi.org/10.1007/BF01531288

Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., & Syrdal, D. S. (2013). Robot-Mediated Interviews - How Effective Is a Humanoid Robot as a Tool for Interviewing Young Children? *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059448

Wu, Y., & Ji, Q. (2019). Facial Landmark Detection: A Literature Survey. *International Journal of Computer Vision*, *127*(2). https://doi.org/10.1007/s11263-018-1097-z

Xia, L., Chen, C. C., & Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3D joints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2012.6239233

Yang, T. Y., Chen, Y. T., Lin, Y. Y., & Chuang, Y. Y. (2019). Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*. https://doi.org/10.1109/CVPR.2019.00118

Yannakakis, G. N., Lund, H. H., & Hallam, J. (2007). Modeling children's entertainment in the playware playground. *Proceedings of the 2006 IEEE Symposium on Computational Intelligence and Games, CIG'06*, 134–141. https://doi.org/10.1109/CIG.2006.311692

Yannakakis, G. N., M, O., Hallam, J., & Lund, H. H. (2006). Comparative Fun Analysis in the Innovative Playware Game Platform. *In Proceedings of the 1st World Conference for Fun 'n Games*.

Ye, M., Wang, X., Yang, R., Ren, L., & Pollefeys, M. (2011). Accurate 3D pose estimation from a single depth image. *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2011.6126310

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

Zhang, H., Wang, M., Liu, Y., & Yuan, Y. (2020). FDN: Feature decoupling network for head pose estimation. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i07.6974

Zhang, Y., Pezeshki, M., Brakel, P., Zhang, S., Laurent, C., Bengio, Y., & Courville, A. (2016). Towards end-to-end speech recognition with deep convolutional neural networks. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *08-12-Sept*, 410–414. https://doi.org/10.21437/Interspeech.2016-1446

Zou, J., Gauthier, S., Gargot, T., Archambault, D., Cohen, D., & Anzalone, S. M. (2021). Development of a wizard-of-Oz based framework for QTRobot to facilitate the learning of handwriting in children with dysgraphia. *CEUR Workshop Proceedings*, *2806*.

3 SYSTEM DESIGN: ARCHITECTURE, ETHICS, AND METHODOLOGY

Summary

The following chapter describes the system architecture based on a hybrid approach that uses a humanoid robot (ZECA) and a device as an Object with Playware Technology (OPT) that can adapt its behaviour to the child's action. Furthermore, it addresses the study ethical concerns and methodological considerations.

3 System Architecture

3.1 Introduction

3.2 Experimental Set-Up

3.3 Ethical concerns

3.4 Methodological Considerations

The content of this chapter is part of the following publications:

- Silva, V., Soares, F., Esteves, J. S., & Pereira, A. P. (2018). Building a hybrid approach for a game scenario using a tangible interface in human robot interaction. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-02762-9_25.
- Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, 2, 497–503.
- Silva, V., Soares, F., Esteves, J., Pereira, A., Leão, C., & Queirós, S. (2021). Hybrid Approach to Promote Social Interaction with Children with Autism Spectrum Disorder. 7th International Conference on Information and Communication Technologies for Ageing Well and E-Health, 69–77.

3.1 Introduction

A system based on a hybrid approach using a robot capable of expressing emotional cues (ZECA) and a device as an Object with Playware Technology (OPT) that can adapt its behaviour to the child's action was developed to be used in emotion recognition activities. Thus, the system architecture consists of a humanoid robot, and OPT device (PlayBrick), and a set of potential sensors to detect child non-verbal behavioural cues.

In order to promote a more organic interaction/interplay between the child and the robot, the system has to be capable of deriving non-verbal behaviours of the child in real-time. In order to derive that information, it is necessary to use external wearable or non-wearable devices. Wearable devices were indeed employed to derive, for example, the child's gaze towards a robot, by using a cap with embedded sensors. However, child with ASD tend not to tolerate to wear wearable devices even for a brief interval of time (E. Bekele et al., 2014; E. T. Bekele et al., 2013a; Mazzei, Greco, et al., 2012). Thus, non-wearable devices are chosen as they offer a less invasive approach (Esteban et al., 2017; Koutras et al., 2018). Additionally, conducting experiments in laboratorial settings can be stressful for children with ASD, taking a considerable time for them to adapt to a new environment (Zaraki et al., 2018). Therefore, it is necessary to conduct the experiments in a setting that is natural and already known by the child, for example in schools.

Having the above statements in mind, the defined pre-requisites of the developed system were: to be portable with the intent to bring this system to the real-world school settings; and to operate with some level of autonomy and in real-time (under human supervision) while promoting recognition skills with children with ASD.

3.2 Experimental Set-Up

In order to conduct an effective child-robot interaction in a supervised manner it is important for the system to be able to characterize the participants to further adapt the support process. Therefore, it is paramount to have a framework able to extract different sensory data. Taking this into account the proposed framework (Figure 3.1) is composed of a humanoid robot capable of displaying facial expressions, a computer, an OPT device (PlayBrick), and a 3D sensor.

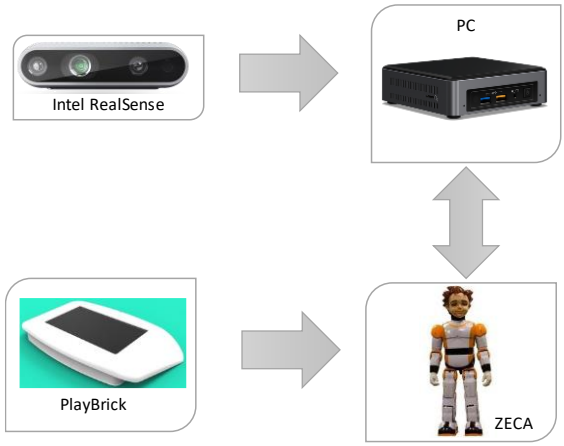


Figure 3.1: The proposed system. Starting from the left bottom: PlayBrick, Intel RealSense D435, computer Intel NUC, and the humanoid robot ZECA.

ZECA is the robotic platform used. The robot is capable of expressing facial cues thanks to the servo motors mounted on its face and a special material, Frubber, which looks and feels like human skin, being a major feature that distinguishes Zeno R50 from other robots.

Concerning the OPT (described in more detail in section 4.3), the PlayBrick (20cm×11cm×3cm) has a 5.0-inch touch resistive display which allows to display more information. As internal components it has an Inertial Measurement Unit (IMU), a small development board (ESP32) that already has built-in Bluetooth and Wi-Fi communication, an RGB LED bar, a Linear Resonant Actuator (LRA), and a Li-Po battery. Interacting with the PlayBrick just means, touching the physical object and manipulating it via natural gestures by tilting it.

The Intel RealSense 3D sensor, 90mm x 25 mm x 25mm, (Intel 2019) is a USB-powered device that contains a conventional RGB full HD camera, an infrared laser projector, and a pair of depth cameras.

The computer used is an Intel® NUC 10 performance Kit with a six core Intel® Core™ i7-10710U Processor (12M Cache, up to 4.70 GHz) with 32 GB of RAM and 1TB of SSD storage, (Intel, 2022). As connectivity it has 2x front (Type-A, Type-C) and 3x rear USB 3.1 Gen2 (2x Type-A, Type-C), HDMI 2.0b and USB-C (DP1.2).

3.2.1 Depth Sensing Device

Concerning the analyses of non-verbal behavioural cues, there are several approaches in the literature that use wearable and non-wearable devices. Wearable approaches are more invasive and children with ASD can have low tolerance for such devices (Rogers & Ozonoff, 2005). Thus, the implementation of the non-verbal behaviour analysis first consisted in choosing a device that would allow the analysis of such behaviours in a non-invasive way, i.e., by using RGB cameras or depth sensing devices. Depth sensing devices offers an extra dimension of information in comparison to traditional RGB cameras, being less sensitive to illumination and subject texture (Esteban et al., 2017).

There is a set of available 3D capable sensing devices with built-in RBG cameras on the market such as Orbbec™ Astra (Orbbec, 2020), Kinect™ V2 (Microsoft, 2021), Intel RealSense™ (Intel RealSense, 2020), among others. Table 3.1 shows the main characteristics of each device.

Table 3.1: 3D capable sensing devices specifications.

Sensor	Dimensions[1] and weight	Technology and Range	Depth Resolution	RGB Resolution	Connectivity
Kinect™ V2 (Microsoft, 2021)	249 x 66 x 67 mm 970 g	Time-of-Flight 0.5 to 4.5 m	512 × 424 @ 30 fps	1920 × 1080 @30 fps	USB 3.1 gen 1
Orbbec™ Astra (Orbbec, 2020)	165 x 30 x 40 mm 300 g	Structured Light and Coded Light 0.6m – 8m	640 × 480 @30fps	640 × 480 @30fps	USB 2.0
Intel™ RealSense (Intel RealSense, 2020)	90 × 25 × 25 mm 72 g	Stereo depth cameras with infrared projector 0.3 m to 3 m	Up to 1280 × 720 @ 90 fps	1920 × 1080 @ 30 fps	USB-C 3.1 Gen 1

[1]Dimenssions – Length × Depth × Height

The present work uses the Intel RealSense model D435 along with the Intel RealSense SDK. The Intel™ RealSense D435 was selected due to its small dimension (one of the pre-requisites for the final system portability), minimum range of 0.3m and novelty in the market.

3.2.2 This sensor was manly chosen due to its smaller size The Robot

The Zeno R50 humanoid child-like robot ZECA (Portuguese acronym that stands for **Z**eno **E**ngaging **C**hildren with **A**utism) is a robotic platform that has 34 degrees of freedom, Figure 3.2, that was developed by Robokind (Robokind, 2023). The robot is capable of expressing facial cues thanks to the servo motors mounted on its face and a special material, Frubber, which looks and feels like human skin, being a major feature that distinguishes Zeno R50 from other robots. It measures 60 cm tall, weighs less than 6 kg, is low power, and battery operated.

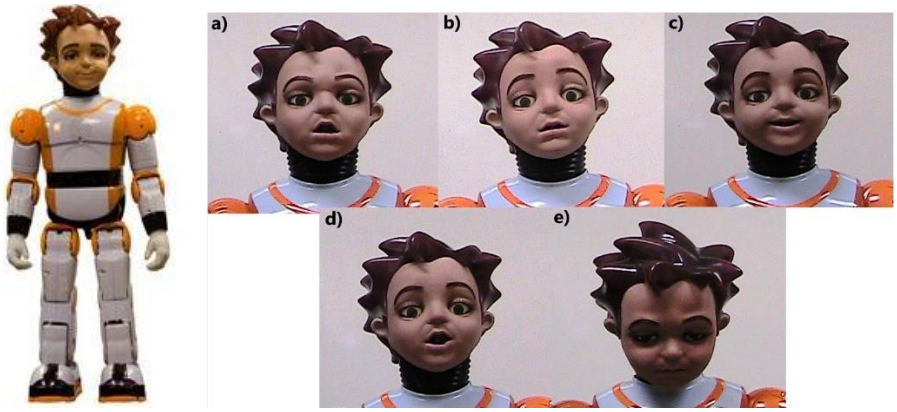


Figure 3.2: On the left: Humanoid robot ZECA. On the right: set of facial expressions displayed by ZECA - a) anger, b) fear, c) happiness, d) surprise, and e) sadness.

Table 3.2 summarizes the main features of the Zeno R50, including the degrees of freedom (DoF), as well as, the onboard sensors and connectivity.

Table 3.2: Principal characteristics of the Zeno R-50 robot from Robokind.

Physical Characteristics		**Multimedia**	1 Loud Speaker
Height	69 cm		3 Microphones
Weight	5,7 kg		2 CMOS Digital Hi-Def Cameras (720p)
Degrees of Freedom (DOF)		**Network Access**	
Total	36	Wi-Fi (IEEE 802.11a/b/g/n)	
Head	11	Ethernet Connection (Gigabit/100/10)	
Arm	4	**Motherboard**	
Waist	1	*Intel Atom* x86 Z530 1,6 GHz processor	
Leg	6	RAM 1GB DDR2	
Actuators		*Flash* memory 4 GB + 16 GB micro SD	
Dynamixel	(x10) RX – 64	**Available ports**	
	(x11) RX – 28	Audio: *Stereo* Line (*in* and *out*)	
PWM Servo	(x11)	(2x) USB 2.0	
Sensors		Video: (1x) HDMI (*out*)	
Available Types	1x gyro meter 3 axes	**Software Compatibilities**	
	1x accelerometer 3 axis	OS	Ubuntu Linux (32 bit x86)
	1x compass 3 axes		
	2x bumpers	Programing Language	Java
	2x ground contact		
	2x cliff	Software	Windows and Linux compatible
	1x IR proximity		
	21x Potentiometer		

Technical drawings of the robot can be found in the Appendix A.1. Hereafter, the robot is going to be referred as ZECA (Zeno Engaging Children with Autism).

3.2.3 The Activities

The activities played are focused on emotion recognition. Therefore, the two game scenarios aiming on improving the children emotion recognition skills are the Recognize and Storytelling activities (Soares et al., 2019). The Recognize and Storytelling activities played during the interaction process with the child-robot-experimenter were validated in a previous work (S. Costa, 2014).

Figure 3.3 depicts the steps that happen during both activities. In the Recognize game scenario, ZECA randomly performs a facial expression and its associated gestures, representing one of the five basic

emotions (happiness, sadness, anger, surprise, and fear). The child has to choose the correct facial expression matching the emotion. Concerning the Storytelling activity, ZECA randomly tells one of the fifteen available stories that are associated with an emotion and the child has to choose the correct facial expression matching the emotion. A visual cue of the story is shown to the child, as vision can be one of the strongest skills of children with ASD (Lima & Levy, 2015). Thus, a visual support can be fundamental to help children with ASD to understand requests and tasks. The visual cue consists of an image showing scenario that describes a situation allusive to a certain emotion (such as happiness, sadness, anger, surprise, and fear) in respect to a story.

In both game scenarios, the child manipulates the OPT by tilting it back or forward in order to scroll through the facial expressions displayed on the OPT screen (detailed in sections 4.2 and 4.3). The child selects an answer, by touching the top touch sensitive surface of the OPT screen, ZECA verifies if the answer is correct and prompts a reinforcement accordingly to the correctness of the answer. Simultaneously, the OPT provides visual and/or haptic feedback accordingly to the child's answer.

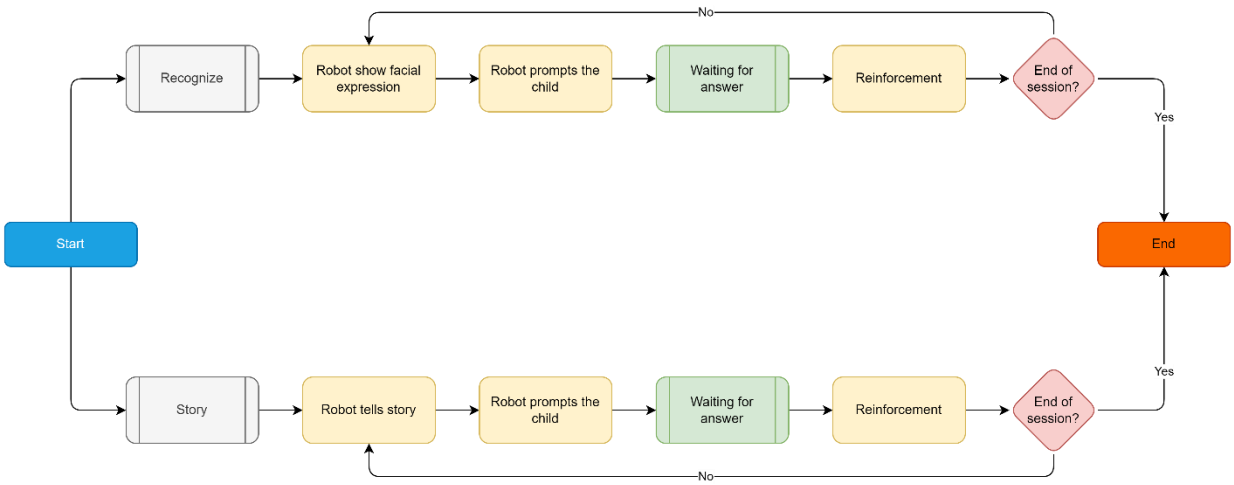


Figure 3.3: Activities flow chart concerning the Recognize and Storytelling game scenarios.

In both tasks, the system will display reinforcements (positive or negative) accordingly to the correctness of the child answer to ZECA prompts. The reinforcements feedback is configurable and displayed by ZECA and the OPT.

3.2.4 Room Setup

The experimental room set-up, Figure 3.4, consists in the child seating in front of the robot, on the child's line of sight, at approximately 85 cm, (Silva et al., 2019). Behind the robot, two cameras to video record the sessions are positioned. Camera A records only the child and camera B records the overall session.

Next to the child is the researcher/therapist responsible for maintaining him/her attention to the task. The PlayBrick, or PlayCube, is in the hands of the child since the beginning of the session.

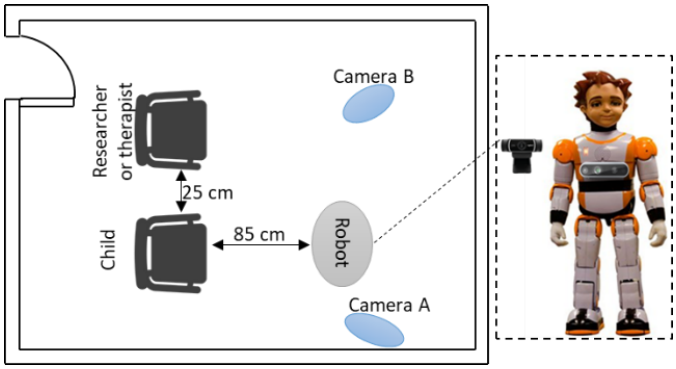


Figure 3.4: The proposed experimental design with a triadic configuration, where the child is approximately 25cm from therapist and 85cm from the robot.

The sessions are video recorded for further analysis, using cameras A and B (Figure 3.4).

This layout is proposed in order to establish a basis of comparison between the participants along the sessions, since the experiments are conducted in an unconstrained (in this case a school) but familiar setting for the child. It also follows the guidelines of works in the literature (S. Costa, 2014; Silva et al., 2019; Zaraki et al., 2018).

3.3 Ethical concerns

The research presented in chapters 4, 5, and 6 focuses on the use of technology and robotics with children with ASD. The studies were conducted in compliance with ethical guidelines set by the Ethics Committee of the University of Minho and the Portuguese National Committee for Data Protection. Additionally, the present research was approved by the national *Direção Geral de Educação* (inquiry n° 0741100002). Thus, the following issues were ensured to meet the ethical concerns:

- **Protocols:** A protocol was established between the participating schools and clinics, and the university where the research was developed. Before the experiments, a meeting was held at the school and/or clinic to address any questions from the professionals who regularly interact with the children. The professionals were also asked to complete a questionnaire to provide information about the children's profiles (e.g. name, diagnosis, characteristics of the child, among others, Appendix A.6);
- **Parents' consent:** The parents/guardians of the children gave their consent for their participation in the research by signing an informed consent form found in Appendix A.4. This

consent form included information about the research objectives, potential risks, and benefits, and also ensured that the parents/guardians had the option to withdraw their child from the study at any point.

- **Privacy:** The personal information of participants in the research is kept confidential and handled in accordance with data privacy regulations as outlined by the Portuguese National Committee for Data Protection. Anonymousness is maintained throughout the research project, and only the researcher and professionals working with the children on a daily basis have access to this information. Videos are only shared at scientific communication events with prior approval from the child's parent or guardian.

In the appendix A.7 is available the approval of the study by the *Direção Geral de Educação* (inquiry n° 0741100002) as well as the guidelines by the Portuguese National Committee for Data Protection that were followed in the present work approved by the University of Minho Ethical Committee.

3.4 Methodological Considerations

A hybrid approach using a robot and an OPT device as the main technological tools is the focus of this research. Instead of evaluating the technology in a one-on-one interaction between the child and the robot, the aim is to foster triadic interactions between the child with ASD, an adult, and the robot plus OPT as objects shared of attention. The research encompasses various subjects, including Affective Computing, Assistive Technology, Objects with Playware Technology, Social Robotics, research on ASD, and developmental psychology.

Children with ASD are characterized by challenges with social interaction, communication, and repetitive behaviours and/or interests. They may also display difficulties with sensory processing, perspective taking, and adapting to changes in routine, (American Psychiatric, 2013). As such several research strategies are used along the research carried out in this thesis.

In this thesis, both qualitative and quantitative research are conducted through the use of pilot studies and quantitative study. Whenever possible, the results are presented as quantitative data and analysed statistically to describe and evaluate the final outcomes by focusing on data and behavioural changes.

Different research methods (including focus group, think aloud sessions, semi-structured interviews, pilot studies) are employed for example in the development of the OPT devices and in the design of the interaction in a system that have levels of autonomy implemented in a supervised manner.

A focus group is a type of research method that involves gathering a small, diverse group of people to

discuss and provide feedback on a specific topic. Focus groups are often used in market research, product development, and social sciences to gather qualitative data and insights through group discussion and interaction. The participants are selected based on certain criteria and are led by a moderator to explore their attitudes, opinions, and experiences related to the topic of discussion. This technique enables the researcher to analyse a topic in depth and gives emphasis to the interaction between participants and their opinions (Amado, 2014; McQuarrie et al., 1991).

With Think-aloud sessions, a type of qualitative research method, participants are asked to verbalize their thought process on a specific matter, (Güss, 2018). The purpose of this method is to gain insight into how people think, solve problems, make decisions, and interact with a particular product, service, or system. In a think-aloud session, the participant is instructed to think out loud while performing the task, and the researcher records and transcribes his/her thoughts, feelings, and observations. This method is often used in fields such as psychology, HCI, and design research, and provides valuable insights into the user experience and decision-making processes.

Online questionnaire is used to assess the impressions of the story scenarios developed. This method has the advantage to reach a large and diverse sample of participants in a short amount of time, allowing the automation of the data collection and analysis.

Semi-structured interviews are a type of qualitative research method that involves a flexible interview guide to explore the experiences, opinions, and perspectives of the participants, (Adams, 2015). Contrariwise to structured interviews, which follow a set of pre-determined questions, semi-structured interviews allow for some elasticity in the line of questioning based on the responses of the participants. The interviewer can follow up on particular points or explore new paths of inquiry as they emerge. This type of interview provides a more in-depth and exploratory understanding of the topic being studied. In the research carried out in this thesis, this method is used for example in the design of the interaction process in a triadic setup child-robot-experimenter where the system has autonomous capabilities that were implemented with the specialists feedback .

Pilot studies are conducted in several steps of the research (chapters 4 and 5) in order to test some of the developed modules of the system with the target group. A pilot study is a preliminary version of a full-scale research project, typically conducted to test and refine the research design, methodology, and procedures before the implementation of the full study. The purpose of a pilot study is to identify and resolve any issues that may arise during the main study, and to ensure that the study will be conducted smoothly and efficiently, (Leon et al., 2011).

A quantitative study was conducted in order to answer the research questions presented in the previous

Chapter 1. This research approach consists of gathering and analysing data to test hypotheses and to understand relationships between variables. It typically involves the use of structured data collection methods, such as surveys or experiments, and statistical analysis to identify patterns and make inferences about the population being studied, (Almeida & Freire, 2010).

The studies described in this thesis follow a structured experimental approach for data collection. This includes forming a formal partnership with the school or clinic where the experiments took place, consulting with professionals who work with the children regularly, and completing a questionnaire to profile each child. The next step involves planning the experiments to align with the study's objectives, including a familiarisation phase for the experimenter to get to know the children, which is crucial as it can be difficult for the children to accept new people into their routine.

When the experiments were conducted, video recordings were one of the sources of information and data. All videos, more specifically in chapter 6, were analysed using a pre-determined list of behaviours that were established based on the specific objectives of the study. To maintain inter-rater reliability, a second, separate coder re-analysed 15% of the videos.

References

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation: Fourth Edition*. https://doi.org/10.1002/9781119171386.ch19

Almeida, L., & Freire, T. (2010). *Metodologia da investigação em psicologia e educação* (6ª). Psiquilíbrios.

Amado, J. (2014). *Manual de investigação qualitativa em educação* (2ª, Ed.). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0879-2

American Psychiatric, A. (2013). Dsm-5. In *Diagnostic and statistical manual of mental disorders, DSM-5*. https://doi.org/doi.org/10.1176/appi.books.9780890425596.dsm17

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H. L., Coeckelbergh, M., Costescu, C., David, D., de Beir, A., Fang, Y., Ju, Z., Kennedy, J., Liu, H., Mazel, A., Pandey, A., Richardson, K., Senft, E., Thill, S., … Ziemke, T. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn*, *8*(1), 18–38. https://doi.org/10.1515/pjbr-2017-0002

Güss, C. D. (2018). What is going through your mind? Thinking aloud as a method in cross-cultural psychology. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/fpsyg.2018.01292

Intel. (2022). *Intel NUC 10 Performance kit NUC10i7FNKN Product Specifications*. https://ark.intel.com/content/www/us/en/ark/products/214614/intel-nuc-10-performance-kit-nuc10i7fnkn.html

Intel RealSense. (2020). *Depth Camera D435 – Intel® RealSense™ Depth and Tracking Cameras*. https://www.intelrealsense.com/depth-camera-d435/

Koutras, P., Potamianos, G., Efthymiou, N., Maragos, P., Filntisis, P. P., & Tsiami, A. (2018). *Multi3: Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots*. https://doi.org/10.1109/icra.2018.8461210

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5). https://doi.org/10.1016/j.jpsychires.2010.10.008

Lima, C. B. de, & Levy, P. de Q. (2015). A Perturbação do Espetro do Autismo. In *Perturbações do Neurodesenvolvimento: Manual de orientações diagnósticas e estratégias de intervenção*.

Mazzei, D., Greco, A., Lazzeri, N., Zaraki, A., Lanata, A., Igliozzi, R., Mancini, A., Stoppa, F., Scilingo, E. P., Muratori, F., & de Rossi, D. (2012). Robotic social therapy on children with autism: Preliminary evaluation through multi-parametric analysis. *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*. https://doi.org/10.1109/SocialCom-PASSAT.2012.99

McQuarrie, E. F., Stewart, D. W., & Shamdasani, P. N. (1991). Focus Groups: Theory and Practice. *Journal of Marketing Research*. https://doi.org/10.2307/3172875

Microsoft. (2021). *Azure Kinect DK Windows comparison | Microsoft Docs*. https://docs.microsoft.com/en-us/azure/kinect-dk/windows-comparison

Orbbec. (2020). *Astra | Orbbec 3D*. https://shop.orbbec3d.com/Astra

Robokind. (2023). *Robokind | Advanced Social Robots*. https://www.robokind.com/

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. *ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics*, *2*, 497–503. https://doi.org/10.5220/0007958004970503

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

4 DEVELOPMENT OF OBJECT WITH PLAYWARE TECHNOLOGY

Summary

The following chapter describes the developments of devices as Objects with Playware Technology (OPT) that can be used as an add-on in HRI support sessions. It presents the studies conducted concerning the development of the OPT.

4 Development of Object with Playware Technology

4.1 Introduction

4.2 Development of the Object with Playware Technology – PlayCube

4.3 Development of the Object with Playware Technology – PlayBrick

The content of this chapter is part of the following publications:

- Silva, V., Soares, F., Esteves, J. S., & Pereira, A. P. (2018). Building a hybrid approach for a game scenario using a tangible interface in human robot interaction. Lecture Notes in Computer Science.
- Silva, V., Soares, F., Esteves, J. S., & Pereira, A. P. (2019). PlayCube: Designing a Tangible Playware Module for Human-Robot Interaction. Advances in Intelligent Systems and Computing.
- Silva, V., Soares, F., Esteves, J. S., Pereira, A. P., & Matos, D. (2020). Adequacy of Game Scenarios for an Object with Playware Technology to Promote Emotion Recognition in Children with Autism Spectrum Disorder. Advances in Intelligent Systems and Computing.
- Silva, V., Soares, F., Esteves, J., Pereira, A., Leão, C., & Queirós, S. (2021). Hybrid Approach to Promote Social Interaction with Children with Autism Spectrum Disorder. 7th International Conference on Information and Communication Technologies for Ageing Well and E-Health, 69–77.

4.1 Introduction

In order to test the introduction of an additional device to the Human-Robot Interaction (HRI), as a first step it was developed one prototype (PlayCube – section 4.2). This first approach was evaluated in two studies conducted in a school environment following the setup described in section 3.2.4. With the results of this first stage, a four-step study was conducted in order to design and develop another prototype for the game scenarios described in section 3.2.3, with a special focus on the Storytelling activity, adding the capability to the OPT device to provide visual support cues during the interaction. A pilot study was conducted with the goal to assess the approach obtained from the four-step study.

4.2 Development of the Object with Playware Technology – PlayCube

In order to test the idea of the introduction of an additional device to the HRI, it was developed the initial OPT prototype PlayCube (Figure 4.1). It consisted of a cube shaped device (7cmx7cmx7cm), with an OLED RGB display, Inertial Measurement Unit (IMU), a small development board (ESP32) that already has built-in Bluetooth and Wi-Fi communication, an RGB LED ring, a Linear Resonant Actuator (LRA), and a Li-Po battery. Additionally, the top face of the cube is a touch sensitive surface. Thus, interacting with the PlayCube just means, touching the physical object and manipulating it via natural gestures (e.g. rotation, shake, tilt, among others).

In order to design the play experience in all its fullness, feedback is a key feature in guiding the children through the play activity. Furthermore, the immediate feedback feature can be a very important factor specially when designing OPT devices for children with impairments (Schuetze et al., 2017).



Figure 4.1: PlayCube prototype. On the left: 3D design. On the middle and right: real images of the prototype.

For ensuring immediate feedback for the PlayCube, a ring with a total of sixteen multicolour and equal spaced LEDs is used. A haptic driver is used to enable haptic control of an LRA actuator. These actuators

can provide haptic and/or visual feed-back to the user. Additionally, the display can also provide visual feedback.

4.2.1 Results

In order to evaluate the PlayCube approach, two studies were conducted in a school environment. The goal of the pilot study with typically developing children was to detect the system constraints in a support session. Concerning the pilot study with children with ASD, the main goal was to verify if the system can implement a procedure that makes the children able to interact in a comfortable and natural way. The experiments were performed individually in a triadic setup, i.e., child-robot-researcher with a duration of five minutes. The activity played was the recognize game scenario described 3.2.3. As quantitative measures, the number of right/wrong and no answer was quantified as well as the children mean response time, in seconds, and standard deviation (SD). It is worth to point out that if the response time exceeded 60 seconds, the child's answer was accepted as not answering the robot prompt.

Preliminary experiments with typically developing children

A set of preliminary experiments were carried out involving eight children aged between six and seven years old. Initially, a questionnaire was presented to the children in order to evaluate the "emojis" used in the cube. In this questionnaire, each child was asked to label the "emoji" with the emotion it was expressing, Figure 4.2.

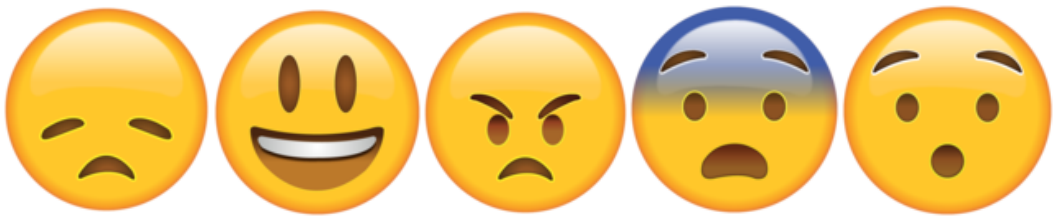


Figure 4.2: Set of emojis used in the OPT device. Starting from the left: sadness, happiness, anger, scared, and surprised.

Table 4.1 presents the results of the questionnaire conducted during the exploratory study with eight typically developing children. The children correctly labelled, with a success rate of 100%, the happiness, sadness, anger, and surprised emotions. However, they had more difficulties in labelling the scared (50%).

Table 4.1: Children's answers matching score.

Emotion	Score
Happiness	100%
Sadness	100%
Anger	100%
Surprised	100%
Afraid	50%

Then, in the next activity, the child was asked to scroll through the cube and find an emotion. This test had the main goal to evaluate the interaction with the PlayCube. As qualitative remarks, two groups of different tilt gestures – up/down and left/ right – allowing the children to scroll through the different faces displayed by the cube were evaluated. In general, it was observed that the up/down tilt gestures were more natural and easier for the children then the left/right tilt gestures when manipulating the cube.

Finally, a pilot study was performed individually in a triadic setup where the children played the recognize activity. The results obtained with the eight typically developing children are presented in Table 4.2.

Table 4.2: Children answers to the robot prompts where C1 to C8 is the number of the child.

	C1	C2	C3	C4	C5	C6	C7	C8
Right	7	3	3	6	2	4	5	7
Wrong	2	1	2	3	1	1	1	2
NoAnswer	0	3	3	0	4	2	2	0

Children 1, 4, 6, 7, and 8 showed an overall better performance. Children 2, 3 and 5 manifested more difficulties. Nevertheless, it is interesting to notice that the children answered the robot prompts giving a strong indication that in general the participants understood the game, and consequently interact with the robot by successfully manipulating the OPT (PlayCube). Furthermore, the mean response time in seconds (and the standard deviation) for the unsuccessful answers was higher when comparing this value for the successful answers – 42.79 (8.73) and 37.66 (2.70) seconds, respectively. This higher value might be related to the children thinking and considering all options that they have available. Additionally, the children response time to the robot prompts decreased along the session – from 46.38 (8.36) to 39.09 (8.42) – indicating that the children responded faster to the robot prompts and were able to manipulate the OPT.

Pilot study with children with ASD

A pilot study with three high-functioning children with ASD (two females and one male) aged between 6 and 9 years old was carried out during four sessions in a school environment with the goal of evaluating the suitability and comprehension of the game scenario and the OPT. Analysing each participant performance (Table 4.3) it is possible to conclude that child A performance improved more along the four sessions when compared with the other participants. Furthermore, the mean response time of child A decreased along the session which can indicate that she understood the activity and how to manipulate the OPT in order to answer the robot prompts.

Table 4.3: Children answers and mean response time to the robot prompts during the four sessions (SD): RT – response time (s), RA – right answer, WA – wrong answer, NA – no answer.

Child	Session	RT	RA	WA	NA
A	S1	45.6(7.9)	3	4	0
	S2	44.9(14.3)	5	1	0
	S3	36.1(5.6)	8	1	0
	S4	28.5(3.1)	9	2	0
B	S1	43.6(14.0)	4	3	1
	S2	54.4(13.9)	3	1	2
	S3	50.5(10.2)	4	2	1
	S4	51.7(13.0)	4	1	1
C	S1	42.0(7.7)	1	5	0
	S2	36.3(8.6)	3	3	0
	S3	34.5(3.8)	4	5	0
	S4	34.3(4.7)	7	2	0

Concerning child B, she does not present an evolutive pattern in the first two sessions. However, in the last two the number of wrong answers decreased, maintaining the number of right answers. It is also worth to point out that the response time, for this participant, is longer. The results from the last two sessions suggests that this participant may need more sessions, which is consistent with the difficulties that these children present when discovering new activities or interests (American Psychiatric Association, 2013a). Even though child C had more difficulties answering the prompts, his performance improved in the last session and the child mean response time also decreased. When comparing the children mean response time along the four sessions and the number of total prompts (Table 4.3), it is possible to

observe that in general the number of prompts increased along the four sessions as the children mean response time decreased along the sessions, considering that the session time is always five minutes. Additionally, it is worth to mention that none of the children that participated in this study abandoned the game. All children repeatedly touched gently the robot and in general manipulated correctly the OPT when prompted by the robot. Furthermore, as they answered the robot prompt they were particularly attentive to the cube feedback, lights and the images displayed in the screen for correct and incorrect answers.

4.2.2 Summary and Conclusions

By analysing the results of the typically developing children it is possible to conclude that they interacted/responded well to the robot and understood the mechanics of the OPT and the game. One of the system constraints detected was the placement of the robot support that caused the children to always have to look up towards the robot face. In the pilot study with children with ASD this was corrected by placing the robot in a similar height to the children face. In general, the children with ASD reacted positively to the activity which can indicate that the developed approach allowed the children to interact in a comfortable and natural way with the system.

Although these initial results are encouraging, the PlayCube can only display small visual cues due to the reduced screen dimension, which limits the size of the visual cues that can be displayed. Visual cues can be fundamental to help the children comprehend the requests and tasks as vision can be one of the strongest skills of children with ASD (Lima & Levy, 2015). For example, when considering the Storytelling game scenario described in section 3.2.3, it can be helpful to show a visual cue, in this case an image card portraying the scene of a story that ZECA is telling to the child. Thus, it is needed to add an extra element – the visual story cards – in order to adequately show the visual cue as it cannot be properly displayed in the small screen size of the PlayCube. This might alter the attention span of the child since he/she will need to focus on an additional element at the same time, besides the robot, and the cards containing the answers (S. Costa, 2014).

4.3 Development of the Object with Playware Technology – PlayBrick

Having in mind the initial findings of the study with the PlayCube as well as its limitations, a new prototype was drafted. In this new configuration, the OPT device provided the visual support cues during the interaction, especially during the Storytelling activity.

Following this idea, a four-step study was conducted, by using qualitative (focus group and think aloud sessions) and quantitative (online questionnaire and pilot study) research methodologies.

4.3.1 Methods

With the present study the research group aimed to evaluate the adequacy of a set of objects and scenarios used for the recognition of emotions in threefold activities (robot, child, and professional) with children with ASD. The scenarios under analysis were created and adapted by the research team according to a group of stories created by the same research team and published in (S. Costa, 2014), described in 3.2.3.

Different research techniques were considered in order to evaluate the scenarios and the role of this new OPT device in conjunction with the robot during the interaction process with the child. A four steps study (Figure 4.3) was conducted: focus group, think aloud session, online questionnaire, and a pilot study.

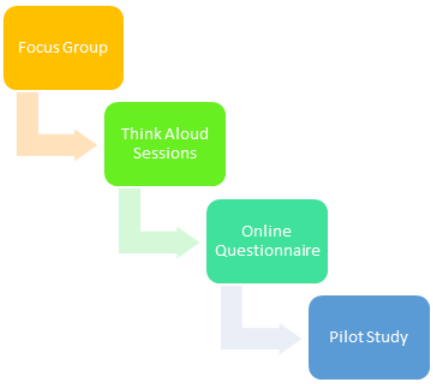


Figure 4.3: Four-step study consisting of a focus group session, think aloud sessions, an online questionnaire, and a pilot study.

In the following subsections are detailed how the different steps of the study were developed.

Focus Group

Focus group is considered the most fit data collection technique to accomplish the research goals set for this study. This technique enables the researcher to analyse a topic in depth and gives emphasis to the interaction between participants and their opinions (Amado, 2014; Stewart & Shamdasani, 2014). Goals were set for the focal group study, according to three main categories: the game scenarios, the OPT device, and the robot role. Concerning the scenarios, the goals were: 1) to analyse the number of stimuli; 2) to analyse the spatial location of the elements of each scenario; 3) to analyse the adequacy of the

colour palette and 4) to define the physical appearance of the main character. Regarding the OPT device, the following goals were set: 1) To define possible manipulation forms for the OPT; 2) To define the kind of feedback that is more suitable for children with ASD and 3) To define the size and format of the OPT device. Finally, the goal of how and what would be the robot role during the interaction was defined.

Participants were selected intentionally to maximize the detail of knowledge about the phenomenon under study, as stated by (Stake, 2010). A group of 10 participants integrated the focus group, and they were selected for being representative of the professionals who support children, aged six to nine years, with ASD, in Portugal. Participants were selected according to the following criteria of intentionality: professionals with different backgrounds in terms of training area (psychologists, occupational therapists, speech pathologists, teachers, and kindergarten teachers); who had a minimum experience of five years in supporting children with ASD; and who had an interest in and availability in participating in the focus group (Table 4.4). Informed consents were given and signed by each participant prior to the focus group session (see Appendix A.5).

Table 4.4: Professionals' characteristics.

ID	Age (years)	Gender	Profession	Expertise (years)
P1	51	Female	Early childhood special education	11
P2	31	Female	Psychologist	9
P3	34	Male	Psychologist	8
P4	53	Female	Occupational therapist	38
P5	32	Female	Speech pathologist	11
P6	45	Female	Special teacher	18
P7	40	Female	Special teacher	5
P8	58	Female	Special teacher	13
P9	39	Female	Special teacher	9
P10	58	Female	Speech pathologist	38

An interview guide was developed to establish a set of topics to be discussed in the focus group. Three types of questions were defined:

(1) general questions, to introduce the study to the participants - its goals and dynamics, and to promote members acquaintance;

(2) specific questions, directed to the three categories in analysis (scenarios, OPT device, and the robot role in the interaction), such as the scenarios' adequacy and amount of stimuli, the adequacy of the spatial location of the distinct elements that integrate the scenarios, the type

and adequacy of the colour palette used in the scenarios, the manipulation forms, the type of feedback, the format of the new OPT, the robot role and how the interaction should be conducted;

(3) closing questions allowing a reflection on the main ideas and a space for additional thoughts to share (for instance, participants were asked about aspects they found relevant but were not included in the discussion topics, and also about advices and suggestions they found relevant to share with the researchers).

A total of 15 scenarios portraying a different emotion were projected for participants to analyse; each scenario contained two sample images. In one sample image, the background used was based on real-world environments, identified as A. Conversely, in the other sample image, the general appearance was more cartoon alike, identified as B. Figure 4.4 shows an example of the two sample images for the scenario representing anger.

After each projection, participants were asked to analyse the dimensions related to the scenarios. Questions related to the OPT device and the robot role were also prompted to the participants, following the goals of each of the three main categories in the focus group study.



Figure 4.4: A sample of two sample images for the scenario representing the emotion anger. In scenario A (left), the background and the characters used were based on real-world environments. Scenario B (right) uses a more cartoon alike appearance.

Following the identification of the professionals who fulfilled the focus group inclusion criteria, they were contacted via email. In that email, the goals for the focus group were presented, and logistics - such as date, place, and time, were provided. An informed, clarified, and free consent request, which contained the ethical procedures to be followed in the focus group session, besides an acceptance statement about audio and video recording has also been sent to all participants, see Appendix A.5. The focus group took place at the university where the authors of the present study work, in a room with adequate conditions in terms of comfort, luminosity, and temperature. The focus group, with a duration of 90 minutes, was conducted by a moderator assisted by a co-facilitator who has training in this kind of data collection technique and so facilitating and ensuring that all topics were covered.

Data was transcribed verbatim and crossed-referenced with notes taken throughout the focus group session. The analysis process comprised: a pre-analysis stage, in which data were watched, heard, and transcribed; an exploration stage, in which data were coded into categories of analysis that had been previously defined according to the study goals; and an analysis stage, in which the thematic unities were identified. The categories of analysis defined were in coherence with the goals of the three main categories (the scenarios, the OPT device, and the robot role): 1) amount of stimuli; 2) spatial location; 3) colour palette; 4) manipulation and interaction forms; 5) type of feedback; 6) size and format of the OPT; and 7) physical appearance of the main character.

Transcripts were analysed line by line, broken into discrete parts and grouped together by similarity of content and/or meaning according to the goals of the study. To ensure data integrity multiple methods of data collection were used and more than one person was involved in the collection and analysis of the data.

Think Aloud Sessions

Think aloud sessions (a total of three) were conducted with two professionals (A1 and A2) with backgrounds in early childhood special education, and multimedia and product design (Table 4.5). The participants of the think aloud sessions gave their consent to participate in this study. These sessions were conducted with the main goal to define the appearance, interaction, and ergonomics of the final OPT prototype, as well as, to validate the final storytelling scenarios. In these sessions, it was presented to the participants the sketches elaborated for each scenario with the respective story by following the main conclusions of the focus group study. The participants were invited to give their impressions of each image scenario in terms of 1) amount of stimuli; 2) spatial location of the elements of each scenario; 3) adequacy of the colour palette.

The verbal reflections/responses of the participants were collected, and the data was analysed following the same three main categories previously defined. Additionally, designs of the OPT device were developed by following the reflections made by the participants of the think aloud sessions. This was an iterative process that culminate into a final design of the OPT device and the scenarios.

Table 4.5: Think Aloud Specialists Characteristics.

ID	Age (years)	Gender	Profession	Expertise (years)
A1	52	Female	Special education	30
A2	43	Male	Industrial designer	14

Online Questionnaire

An online anonymized questionnaire was conducted with a total of 138 voluntary participants, 69 of them adults (aged 20-67 years old, where 69.6% female and 30.4% male) and the remaining typically developing children (aged 7-11 years old, where 56.5% female and 43.5% male) (Silva et al., 2020). In the questionnaire, it was randomly presented the text and the image of each story from a total of 15 stories. The participant had to read the story, observe the image, and select the emotion (anger, fear, happiness, sadness, and surprise) that matches the story. The goal of the online questionnaire was to validate the image scenarios that were developed following the guidelines from the focus group and the think aloud sessions with the stories. The length of time required to complete the online questionnaire varies according to who answer, whether by an adult or a child, but it did not exceed 10 min on average.

Pilot Study

A small pilot study was conducted in a school setting with 4 children with ASD in order to assess the approach obtained from the present four-step study.

The selected children were aged between 6 to 10 years old (M=8.75; SD=0.96). Although ASD is more prevalent in boys (Christensen et al., 2016), the selected sample has gender-balanced representation). All children that participated in the study are verbal but their attempts to initiating interactions and make friends are difficult and typically unsuccessful (Level 1 of severity levels, DSM5, 2013). The participants are high functioning according to their diagnosis. A total of 3 sessions per child was performed.

The pilot study was conducted following the triadic setup (child-robot-researcher) described in section 3.2.4.

The task selected to be conducted was the storytelling game scenario in which randomly selected stories among 15 social stories are told by the robot, and the participant has to identify the emotion of the main actor, i.e. the robot. For these stories it was used the same scenarios that were designed with the inputs taken from the focus group and the thinking aloud sessions. As the robot starts telling the story, simultaneously an image is shown representing the social context of the story as a visual cue. Then, the child is prompted to answer how the robot felt in that story scenario. The child selects the answer by tilting back and forward the OPT device, scrolling through the facial expressions (common emoji) displayed by the OPT and touching the image. In parallel, when the answer is selected, a positive or negative reinforcement is prompted by ZECA and the developed OPT. The type of reinforcement is configurable on both the robot and the OPT according to the child preferences. In this present pilot study,

the types of reinforcements used was the same for three children (robot: verbal + movement + sound; OPT: visual + haptic) and different from one child (robot: verbal + movement; OPT: visual + haptic).
The goal of the pilot study was to detect the system´s constraints and verify if it allows implementing a procedure that makes the children able to interact in a comfortable and natural way during a support session.
The videos of the experiments were coded in terms of frequency and time in order to quantify different behaviours. For example, the frequency and the time that the children look at the OPT device was recorded. The number of times that the children needed help and the number of wrong usages of the OPT were counted. Additionally, the number of correct, incorrect, and unanswered answers as well as the number of total robot prompts during the sessions were quantified. The children's mean response time to the robot's prompts were also registered. At the end of each session, the robot asked the participant if he/she wanted to play more. This information was also recorded.
The non-parametric Wilcoxon signed-rank test (alternative to parametric paired t-test) was used to compare the children's attention during the storytelling game scenario using the children's mean duration per-gaze. This test will be reported by using the Z statistic.

4.3.2 Results

This section describes the results obtained from each research method employed to design the new story scenarios as well as the OPT PlayBrick.

Focus Group

The analysis was organized according to the three study categories.
Analysing the goals of the scenarios, the amount of stimuli in the scenarios, the focal group suggested to decrease the number of objects in the scenarios – *"...therefore, the first aspect would be to have only what matters, in this story, which is Zeca, Alice, and the game...focus the attention of the child with autism spectrum disorder on the face of the robot and the emotion that the robot will have;.. I think it would be better to have less objects ..."* (P5). Following the same idea, the participant P10 proposed – *"... take away most of the elements you might have."*. Moreover, in all scenarios, the participants recommended to direct the focus to the characters of the story and use a simple element to describe the scenes – *"Yes, leave only one element that identifies the site ... and to focus the child's attention on the character"* (P1). Thus, presenting only the materials necessary to complete the activity and eliminating

irrelevant and potential distracting materials can be a good visual arrangement (Hampshire & Hourcade, 2016). Finally, the objects placed in the scene should be similar to those the children are used to see and play – *"...I would use what kids are used to play..."* (P9).

Regarding spatial location in the scenarios, the objects that identify the scenes should be placed in the background relatively to the main character – *"... put the elements in the background there or somewhere in the picture, showing that this is a playground because it has a slide..."* (P10). Other participants also supported this claim – *"That allows contextualizing"* (P9). Participant (P3) further stresses this idea – *"Often here, the difficulty is this, this contextualization, but at the same time we do not want it to be the focus of the child's attention, but rather a global understanding of the situation."* This idea is further supported by (Happé & Frith, 2006) where the author says that: *"a person with ASD prefers a detail-oriented analysis and demonstrates great difficulty, or even inability to process global information"*. Furthermore, children with ASD have the facility to segment in parts, however they struggle when they have to deal with a global organization of the material (Mammarella et al., 2014). Additionally, since the scenarios are static images, the participants suggested to add a feeling of some movement in some of them in order to give the idea that something is happening. For example, in one of the scenarios there is a tower of blocks where the supporting character, Alice, takes it down, thus it was suggested to have some blocks falling – *"You can have the tower and the top with her hand, some falling and others on the ground."* (P9) In another scenario, it was suggested to add a crowd of people appearing to walk – *"Exactly, I would insert some movement, like people walking ..."* (P3)

Concerning the scenarios colour palette, the participants suggested the use of more vivid and contrasting colours, in order to accentuate the objects in the scenario and to appeal the child's attention since that usually the colours draws the child's attention – *"... I would tweak the colour palette in the room, I think (Hã) overall it's all very lively and I think the wall and floor should have a more beaten colour and the toys should have more brighten colours to help focus on (Hã), in the event..."* (P9). In the same sense – *"... there has to be a colour palette that gives a contrast like that, to differentiate."* (P2). Moreover, the participants suggested not to use white as a background colour since they considered that it is not a warm/cosy colour – *" ... the background wall may have a colour, not white, white will give an, an air ... it is not cosy, ..."* (P9).

Regarding the appearance of the main character (ZECA), Figure 4.5, participants agreed to use a more cartoon like avatar instead of the real robot image – *"Throughout the various scenarios that we see (Hã) I identified, at the level of facial expression, I identified more with the, the avatar than with the robot, yes. That's why I think here is easier to perceive the ZECA, but in cartoon."* (P6); Participant P3 further stresses

this opinion – *"It is because in the scenarios that we have been choosing, we have been defining, I think fits more, this cartoon alike image."*. In the same sense participant P9 complements this view – *"It looks like it's part of the picture."* (P9). Moreover, the participants recommended that the facial expressions of the main story character, ZECA, should translate the emotion in the story, i.e., if in the story ZECA was angry, its facial expression in the image should reflect the same emotion – *"...it would help if the robot showed some facial expression, in this scenario could be an angry facial expression..."* (P3). Regarding the other characters in the stories, they should display a neutral facial expression or the same expression as the main character, ZECA, when the feeling is mutual.



Figure 4.5: The physical appearance of the main character of the stories. On the right, a photo of the robot is shown. On the left, a visual cartoon representation of the same robot is shown.

Regarding the design of the OPT and in view of how the answer options should be prompted to the child, the participants suggested to only appear one option at the time. Since the device has a touch display and an Inertial Measurement Unit (IMU), the child could select the answer by touching the arrows, as shown in Figure 4.6, or tilting the device back and forward in order to scroll through the options, promoting a more dynamic way of interaction – *"This is that, this is the part of dynamism, I still realize, the idea of sweeping is more intuitive yes, but this question is also dynamic ... It is more elaborate, the gesture requires more competence, but it is possible to be even more motivating, more dynamic..."* (P3). He further stresses this idea – *"... I say again in terms of competence it requires more than just the sweep or the option to press the arrow, but it becomes perhaps more dynamic."* (P3).

The participants suggested that manipulation/interaction should be configurable accordingly to the child's capacity.

Relatively to the size of the overall new OPT device, the participants agreed to the use of a 5-inch size display – "Yes" (all participants). Furthermore, one participant expressed a concern regarding that a bigger screen size could be difficult for the child to interact with the tilt gestures, becoming less

comfortable – *"... I realize that having a bigger screen could be better in visual terms, but in terms of interaction, too big can be difficult to operate..."* (P3). This perspective is in line with the one advocated by P8 – *"I think they are so used to mobile phones, and all sorts of games, that they focus their attention on them, on devices like that, as the first one, and it's easier to handle something smaller"*.

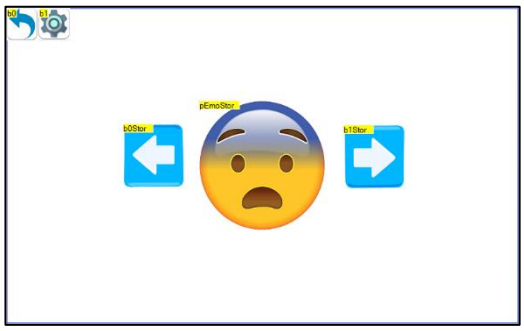


Figure 4.6: The interface that is prompted to the child on the OPT where he/she has to select the correct answer by pressing the arrows in order to scroll through the different emojis reflecting the emotion presented in the storytelling scenario.

Considering the types of feedbacks of the OPT, it was suggested to the participants to discuss the option of using visual (through the display and multicolour LEDs), sound, and haptic feedback. Firstly, the participants suggested the use of visual feedback by, for example, displaying an animation with fireworks – *"Or like those little things coming out, like fireworks, that's right ..."* (P8). However, another element of the group proposed the use of simple feedback images, without animation – *"I wouldn't put animation, put good in green, try again in yellow or ..."* (P5). With other participants agreeing with this idea – *"Ah! Yes. I agree with P5, the positive and negative gesture I find it simpler."* (P4). The participant P8 supports the same idea – *"It can also be."*. Additionally, it was discussed the inclusion of further stimulus such as the haptic, sound, and the use of multicolour LEDs, however some participants had some concerns about the use of several forms of feedback – *"It is a lot of confusion, a lot of stimuli, we already have the sound stimuli, the visual and we will still have the haptic..."* (P3). Conversely, other participants were more open to this idea – *"Yeah, the game can have this repertoire. And I don't know, try it because you can have a child that have fun with it."* (P10). For children with ASD, learning via reinforcement through feedback can be one of the most effective approaches to reinforcing desired behaviours since throughout learning, an association is formed between a suggestion or action and a reinforcer with some intrinsic motivational value (Schuetze et al., 2017). Following this idea, the participants agreed that with the use of all forms of the proposed feedbacks, but they must be configurable to each child – *"Yes you will have to configure it ..."* (P3). Therefore, the type of feedback must be configurable for different children as some types of feedback can be unenjoyable for some individuals, e.g., in the case of children with ASD, in general, a

sound feedback can be unpleasant for them (American Psychiatric Association, 2013a).

Considering robot role, in particular, the interaction procedure between the child and the robot, the participants agreed that when the robot should start telling a story, at the same time, the picture that represents that story should be displayed on the OPT, and when the robot prompts the question asking what the emotion was, the options should appear – *"The image of the story would disappear, I think so."* (P8). Furthermore, participant P9 also shares the same opinion –*"So the image would be there, but it would work the way you say, the answer options would only appear when you ask the question, after the question, and would not appear before."* (P9).

Regarding the scenarios, in general, all participants suggested the development of new ones taking into account the aspects discussed during the focus group. Most of the participants preferred the cartoon scenario (marked as B in Figure 3) but adding the robot with a cartoon appearance and some changes to these scenarios taking into account the number and spatial locations of objects and the colours used – *"Personally, I think that B has (Hã) a set of colours more appealing than A, no doubt, at the same time could also be, have more distracting elements"* (P3).

Table 4.6 shows the overall group conclusion on the ideas discussed during the focus group.

Figure 4.7 illustrates the interaction procedure between the robot and the child agreed by the participants in the focus group.

Table 4.6: Main conclusions of the focus group.

	Stimuli Amount	Spatial Location	Colour Palette
Scenarios	○ To decrease the number of objects in the scenarios. ○ To direct the focus to the characters of the story and use a simple element to describe the scenes. ○ The objects placed in the scene should be familiar to the child.	○ The objects that identify the scenes should be placed in the background relatively to the main character. ○ To add a feeling of some movement to the scenarios.	○ To use of more vivid and contrasting colours. ○ To use warmer colours instead of white backgrounds.
	Manipulation	**Display**	**Feedback**
OPT device	○ The manipulation/interaction of the OPT should be configurable accordingly to the child's capacity. ○ Tilt or touch gestures as form of interacting with the device;	○ A bigger display was suggested;	○ To use of simple images as visual feedback. ○ To use of multicolour LEDs and haptic feedback. ○ The type of feedback must be configurable according to the child preferences.

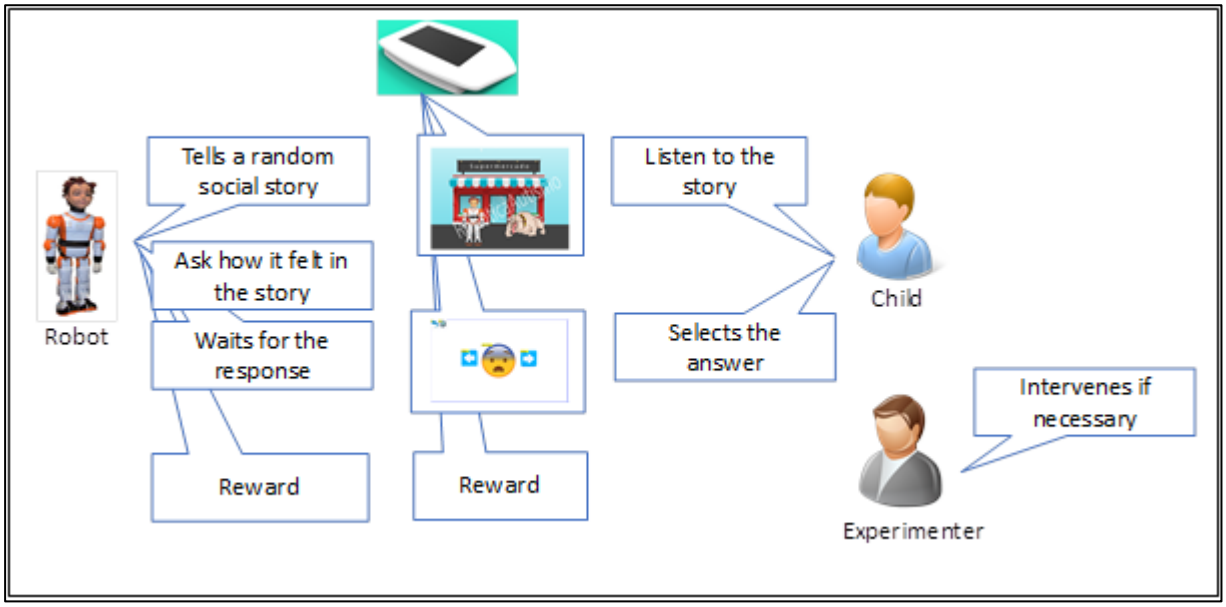


Figure 4.7: Interaction diagram for the storytelling game scenario.

Think Aloud

Concerning the scenarios, they were developed by following the main considerations of the participants of the focus group. Then, they were shown to the specialists where they were asked to verbalize their thoughts about the new scenarios. Therefore, Figure 4.8 shows an example of two scenarios that were designed and validated by following both the focus group and the specialists' inputs/guidelines. All the developed images for the Storytelling game scenario can be found the Appendix A.3. The stories can be found in Appendix A.2.



Figure 4.8: Sample images of two storytelling game scenarios. On the left it represents the anger emotion. On the right it represents the surprise emotion.

By following the main thoughts of the focus group and the suggestions of the specialists, the final design of the OPT was elaborated. Figure 4.9 shows the OPT final design. It has a 5-inch display, an ergonomic

design, without any sharp edges as suggest by the specialists.



Figure 4.9: Final design of the OPT device - PlayBrick.

Online Questionnaire

The information gathered with the online questionnaire allowed to verify the accuracy of the three scenarios per emotion. Figure 4.10 summarizes the results of the online questionnaire.

The chart shows the average matching accuracy (%) of the three scenarios per emotion for adults and children. For each emotion there are three stories. Thus, for each emotion and target group (adult and children) it was computed the average of the matching accuracy (%). In general, the average emotion matching accuracy was 91.8% for the adults and 89.9% for the typically developing children. The lower value obtained (70.1%) corresponds to the emotion Surprise for the children.

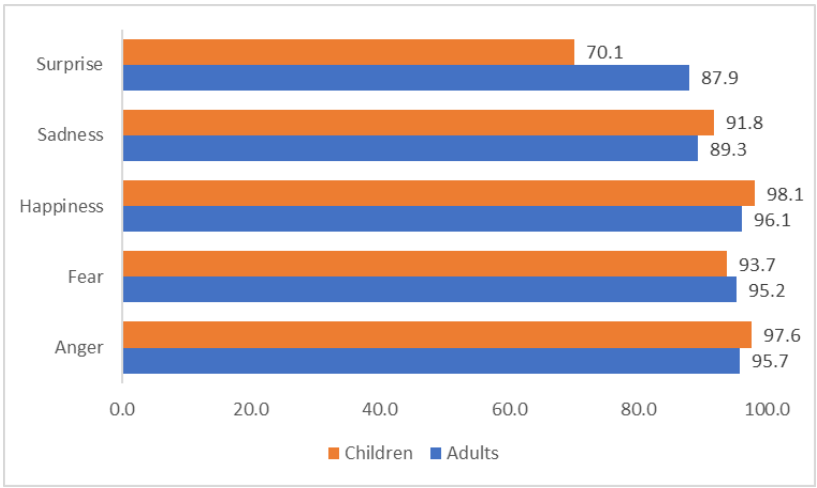


Figure 4.10: Average matching accuracy per emotion (%) for children and adults.

Pilot Study

A set of experiments were carried out in a school setting with four children with ASD where they interacted with the robot in a triadic setup by playing the storytelling activity. Regarding the data obtained from the

pilot study sessions, it was found, in general, that children gazed more towards to the OPT device than to the robot, Figure 4.11. In particular, the children's mean duration per gaze in the storytelling game scenario was significantly less for the gazes directed at the robot (M=5.28; SD=1.76) than those directed at the OPT (M=12.58; SD=2.97), Z=-3.059, $p<0.001$, Figure 4.12.

The number of right and wrong answers are shown in Figure 4.13. All participants answered the robot prompts and in general there was a positive evolution, regarding the number of successful answers along the sessions.

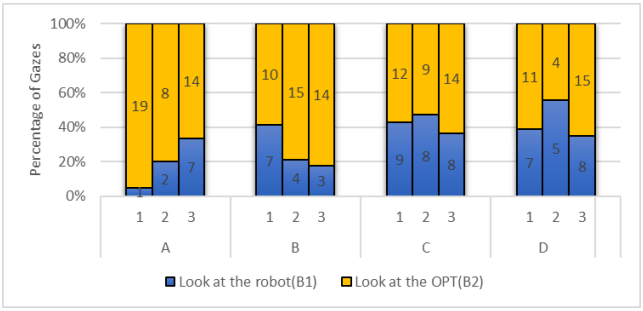


Figure 4.11: Number of gazes (A) towards the robot and the OPT for the 4 children (A, B, C, and D) during the sessions.

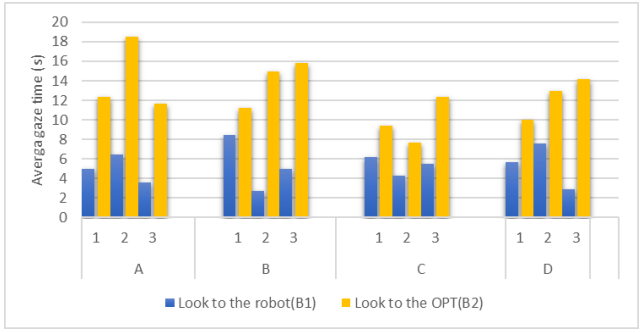


Figure 4.12: Mean gaze time (B) towards the robot and the OPT for the 4 children (A, B, C, and D) during the sessions.

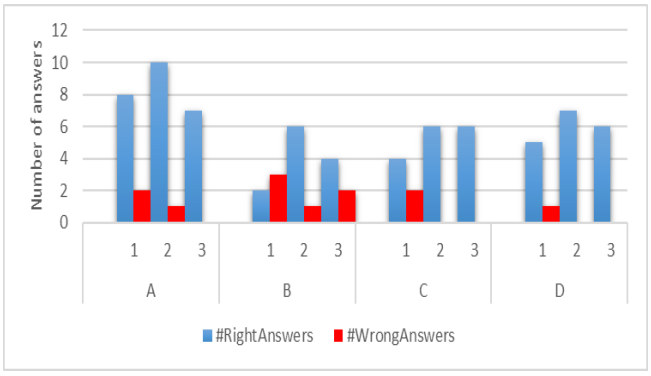


Figure 4.13: Answers of the children to the robot's prompts during the sessions.

The children response time to the robot prompts was in average the same, Figure 4.14. The times that the children needed help was relatively low over all sessions, Figure 4.15.

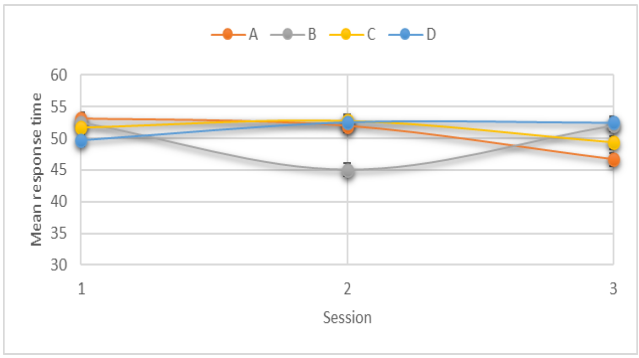


Figure 4.14: Children's mean response time in seconds (with confidence interval) to the robot prompts during the four sessions.

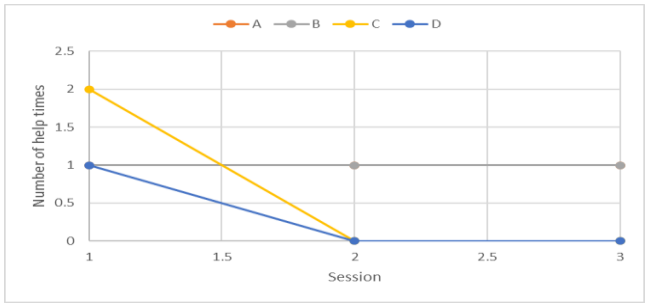


Figure 4.15: Number of times that the children needed help per session.

4.3.3 Summary and Conclusions

Regarding the scenarios, in general, all participants in the focus group suggested the development of new ones considering the aspects proposed. Most of the participants preferred the cartoon scenario (Figure 4.4 marked as B on the right) but adding the robot with a cartoon appearance and changing the number and spatial locations of the objects and the colour palette used.

Concerning the new OPT device, the participants agreed that the device should display different forms of feedback. Furthermore, these feedbacks should be tailored to each child, i.e., it should be customized. In the same sense, the members of the focus group also agreed that the manipulation and interaction, i.e. the use of gestures or arrows on the screen to select the answer, should be configurable accordingly to the child's interests and competences. Finally, the group firmly agreed that a 5-inch touch display would be the most adequate to be used for the device.

By conducting the think aloud sessions, it was possible to arrive to the final designs of the OPT and the scenarios following the feedback of the specialists and having in mind the outputs from the focus group. Finally, an online survey was conducted with 138 participants (child and adults) where, in general, the average emotion matching accuracy was 91.8% for the adults and 89.9% for the typically developing children. Thus, both children and adults, in general, correctly identified the emotion of each social story

(Silva et al., 2020).

Following these results, a pilot study was conducted with 4 children with ASD. The results allow to conclude that the children understood the mechanics of the game and successfully interacted with OPT. There was a significant difference in the mean gaze time towards the OPT, suggesting that the children used/relayed on the OPT during the activities. Moreover, in general, the children were keen to participate in the activities since they wanted to continue on playing. Additionally, they were also attentive to the OPT feedback, lights, haptic as well as the images for correct and incorrect answers displayed on the screen (Silva et al., 2021).

The use of the OPT eliminates the necessity of having several visual cards thus making the interaction seamlessly, as the interaction is dynamic. Additionally, the OPT device can offer a unique way to quantify the child behaviour without being too invasive (as for example the use of wearable sensors) and further feed the system in order for the robot to "know" the level of engagement of the child.

References

Amado, J. (2014). *Manual de investigação qualitativa em educação* (2ª, Ed.). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0879-2

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic and Statistical Manual of Mental Disorders 4th Edition TR.*, 280. https://doi.org/10.1176/appi.books.9780890425596.744053

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Christensen, D. L., Baio, J., van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L. C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., & Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance*

Summaries. https://doi.org/10.15585/mmwr.ss6503a1

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H. L., Coeckelbergh, M., Costescu, C., David, D., de Beir, A., Fang, Y., Ju, Z., Kennedy, J., Liu, H., Mazel, A., Pandey, A., Richardson, K., Senft, E., Thill, S., ... Ziemke, T. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn*, *8*(1), 18–38. https://doi.org/10.1515/pjbr-2017-0002

Hampshire, P. K., & Hourcade, J. J. (2016). Teaching Play Skills to Children with Autism Using Visually Structured Tasks. *TEACHING Exceptional Children*, *46*(3), 26–31. https://doi.org/10.1177/004005991404600303

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. In *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-005-0039-0

Intel. (2022). *Intel NUC 10 Performance kit NUC10i7FNKN Product Specifications*. https://ark.intel.com/content/www/us/en/ark/products/214614/intel-nuc-10-performance-kit-nuc10i7fnkn.html

Intel RealSense. (2020). *Depth Camera D435 – Intel® RealSense™ Depth and Tracking Cameras*. https://www.intelrealsense.com/depth-camera-d435/

Koutras, P., Potamianos, G., Efthymiou, N., Maragos, P., Filntisis, P. P., & Tsiami, A. (2018). *Multi3: Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots*. https://doi.org/10.1109/icra.2018.8461210

Lima, C. B. de, & Levy, P. de Q. (2015). A Perturbação do Espetro do Autismo. In *Perturbações do Neurodesenvolvimento: Manual de orientações diagnósticas e estratégias de intervenção*.

Mammarella, I. C., Giofrè, D., Caviola, S., Cornoldi, C., & Hamilton, C. (2014). Visuospatial working memory in children with autism: The effect of a semantic global organization. *Research in Developmental Disabilities*, *35*(6), 1349–1356. https://doi.org/10.1016/j.ridd.2014.03.030

Mazzei, D., Greco, A., Lazzeri, N., Zaraki, A., Lanata, A., Igliozzi, R., Mancini, A., Stoppa, F., Scilingo, E. P., Muratori, F., & de Rossi, D. (2012). Robotic social therapy on children with autism: Preliminary evaluation through multi-parametric analysis. *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*. https://doi.org/10.1109/SocialCom-

PASSAT.2012.99

Microsoft. (2021). *Azure Kinect DK Windows comparison | Microsoft Docs*. https://docs.microsoft.com/en-us/azure/kinect-dk/windows-comparison

Orbbec. (2020). *Astra | Orbbec 3D*. https://shop.orbbec3d.com/Astra

Robokind | Advanced Social Robots. (2015). http://www.robokindrobots.com/

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Schuetze, M., Rohr, C. S., Dewey, D., McCrimmon, A., & Bray, S. (2017). Reinforcement learning in autism spectrum disorder. In *Frontiers in Psychology* (Vol. 8, Issue NOV). https://doi.org/10.3389/fpsyg.2017.02035

Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. *ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics*, *2*, 497–503. https://doi.org/10.5220/0007958004970503

Silva, V., Soares, F., Esteves, J., Pereira, A., Leão, C., & Queirós, S. (2021). Hybrid Approach to Promote Social Interaction with Children with Autism Spectrum Disorder. *7th International Conference on Information and Communication Technologies for Ageing Well and E-Health*, 69–77. https://doi.org/10.5220/0010468600690077

Silva, V., Soares, F., Esteves, J. S., Pereira, A. P., & Matos, D. (2020). Adequacy of Game Scenarios for an Object with Playware Technology to Promote Emotion Recognition in Children with Autism Spectrum Disorder. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-27928-8_44

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

Stake, R. E. (2010). Qualitative Research: Studying How Things Work. In *Qualitative research: Studying how things work*. NY: Guilford Press.

Stewart, D. W., & Shamdasani, P. N. (2014). *Focus Groups: Theory and Practice* (Third). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/focus-groups/book239690

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic

System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

5 BEHAVIOURAL CONTROL ARCHITECTURE

Summary

The following chapter describes the supervised behavioural framework with its respective layers – Perception, Reasoning, and Actuation. It introduces each layer explaining its main functionalities. The perception layer is responsible for extracting data that is used to detect some of the child non-verbal actions. For the reasoning layer it was conducted a study with professionals aiming to define how ZECA could adapt its behaviour to the child. Finally, the actuation layer is addressed. Additionally, a Graphical User Interface was developed in order to provide an easier way to control the system.

5 Behavioural Control Architecture

5.1 Perception Layer – Hardware, OPT, Head Pose estimation, and body pose estimation

5.2 The Reasoning Layer

5.3 The Actuation Layer

5.4 Pilot Study

The content of this chapter is part of the following publications:

- Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, 2, 497–503.
- Silva, V., Soares, F., Leão, C. P., Esteves, J. S., & Vercelli, G. (2021). Skeleton driven action recognition using an image-based spatial-temporal representation and convolution neural network. Sensors, 21(13).
- Silva, V., Soares, F., Sena Esteves, J., & Vercelli, G. (2021). Human action recognition using an image-based temporal and spatial representation. 2020 12th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT), 1–6.

The software architecture includes three main subsystems where two are interconnected via a TCP/IP network and Bluetooth, Figure 5.1.

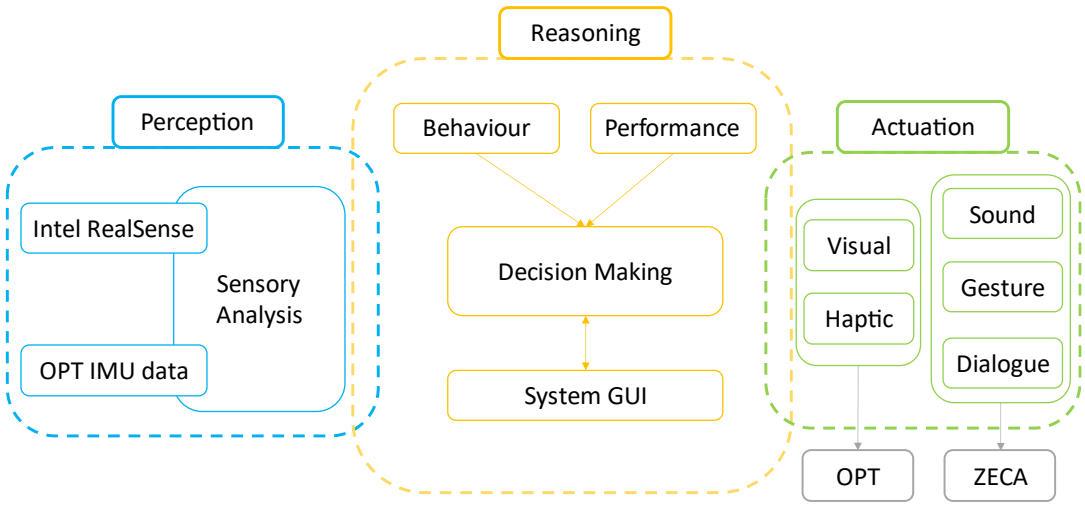


Figure 5.1: The behavioural architecture is composed of three main layers – the Perception (blue), Reasoning (yellow), and Actuation (green) layers.

The perception layer is responsible for sensing and processing of the data received from the sensors. Since the activities target the recognition of emotions, it is necessary to extract and analyse features that are required by these game scenarios such as: head orientation, the frequency and duration of possible movements done by the participant, and stereotypical behaviours (such as hands shaking).

The reasoning layer is influenced by the child behaviour and performance. Using the information acquired and computed by the perception layer, the system will decide the interaction flow of the intervention. The output of this layer will influence the dynamics of the next layer, the Action layer. Therefore, the behaviour displayed by the robot will be influenced by the interaction flow of the session. Additionally, the feedback given by the OPT (visual and haptic) will also be dependent of the output of the previous layer.

5.1 Perception Layer – Hardware, OPT, Head Pose estimation, and body pose estimation

As one of the elements of the perception layer, the PlayBrick sends IMU data in order to inform the system if the children is interacting with it. After processing, the information send is the detected tilt gestures (forward, backward), Figure 5.2

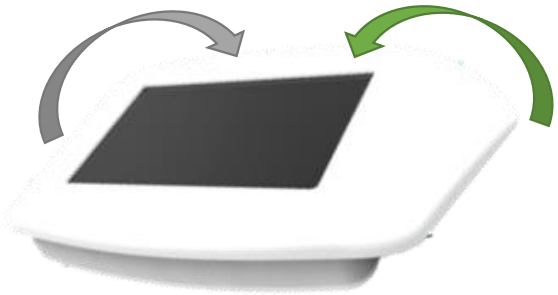


Figure 5.2: The PlayBrick and gestures that can be detected - forward tilt (grey arrow) and backward tilt (green arrow).

Head pose estimation and user skeleton tracking

Concerning the analyses of behavioural cues, there are several approaches in the literature that use wearable and non-wearable devices. As mentioned in section 3.2.1, it was chosen the Intel™ RealSense D435 as it offers the most suitable choice for the system based on the criteria listed in section 3.1.

In order to analyse the child non-verbal behaviours namely the head position and body gestures, there are open and commercial frameworks that have built-in algorithms to detect and track subjects. Concerning the open solutions available, OpenFace is the most advanced solution (Baltrusaitis et al., 2016). It offers a set of algorithms that can track and extract facial cues in real-time (15-20 fps depending on the computing device) by using an RGB camera. It is an open source solution capable of detecting the user face and outputs the facial landmarks, the head pose and eye gaze data, and facial action units, Figure 5.3.

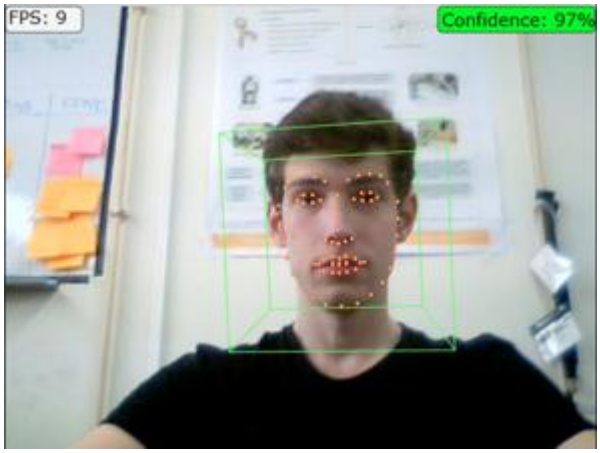


Figure 5.3: Facial features detection and extraction by using OpenFace with the frame rate (top left corner) and confidence of face detection (top right corner).

It accurately detects the user face, facial landmarks, and estimates the head position. However, the pupil detection as well as gaze estimation is inaccurate (OpenFace Pupil and Gaze Estimations, 2021; OpenFace Pupil Estimation, 2018). Furthermore, the performance in real-time was less than expected (Figure 5.3), with high CPU usage at above 60% using all cores available, same was found by other users of the framework (OpenFace CPU, 2016). Additionally, it is focused on only analysing facial features. Thus, another open source solution available that also can track and analyse body gestures is OpenPose (Cao et al., 2021). It is a real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints (in total 135 keypoints) on single images. Although the authors mention it can run in real-time, they highly recommend to use a machine with a dedicated Nvidia GPU with at least 1.6GB of dedicated VRAM plus 2.5GB of free RAM and a CPU with at least 8 cores as minimum requirements (Hidalgo et al., 2022). Additionally, even fulfilling the minimum requirements it does not guaranty real-time performance (*OOM with CUDNN Installed and GPU Requirements Met · Issue #1929*, 2021).

Considering these limitations and looking to what is available in the commercial space, Nuitrack SDK (Nuitrack, 2019) offers similar features to OpenFace and OpenPose. Nuitrack is a 3D tracking cross-platform middleware capable of full-body skeletal tracking (up to 19 joints), with an agnostic 3D sensor independent by supporting the Microsoft Kinect V1 and V2, Orbbec™ Astra, Intel™ RealSense, among other 3D devices. The solution can track up to 6 skeletons at a time. In more detail, it can detect and track the user's face, estimate the head pose and landmark positions, and detect facial expressions, Figure 5.4 (Nuitrack Face Tracking, 2019). Conversely, pupil detection and eye gaze estimation is not included in the set of capabilities of Nuitrack SDK as is it included in OpenFace. The real-time performance can be up to 45 fps depending on the hardware configuration used (Nuitrack Frame Rate, 2018). Additionally, it is compatible with a range of depth sensing devices including the Intel™ RealSense (Nuitrack, 2019).

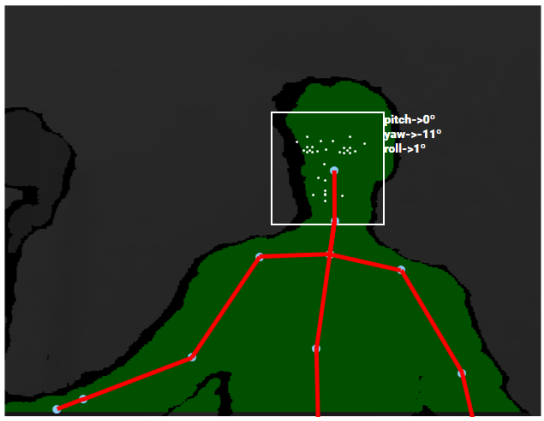


Figure 5.4: Nuitrack detection capabilities. It can track the user and its skeleton data plus it also offers facial tracking and pose estimation.

Based on the set of features and limitations of both frameworks as well as the real-time pre-requisite, the Nuitrack SDK was chosen.

5.1.1 Perception Layer – Detection of non-verbal behaviours for action recognition

One way to recognize different actions of the user is through skeleton pose data from depth sensors. To recognize in real-time the typical and stereotypical behaviours of children with ASD, it is important to take into account the hardware to be used (in this case the Intel™ RealSense with the Nuitrack SDK) as well as to select the behaviours to be detected. The following sections detail the datasets used and the procedure implemented in order to detected the child's non-verbal behaviours.

Selection of behaviours and datasets

Two datasets were used: the KARD dataset (Gaglio et al., 2015) and the Typical and Stereotypical Behaviours Dataset (TSBD) dataset developed with typical and stereotypical behaviors based on the previous works of the research team with children with ASD (S. Costa, 2014).

The KARD (Kinect Activity Recognition Dataset) (Gaglio et al., 2015) contains 18 activities, where each activity is performed 3 times by 10 different subjects. The activities are horizontal arm wave, high arm wave, two hand wave, catch cap, high throw, draw x, draw tick, toss paper, forward kick, side kick, take umbrella, bend, hand clap, walk, phone call, drink, sit down, and stand up. The skeleton joints used were head, neck, torso, right shoulder, right elbow, right hand, left shoulder, left elbow, left hand, right hip, right knee, right foot, left hip, left knee, and left foot.

The Typical and Stereotypical Behaviours Dataset (TSBD) used includes typical and stereotypical behaviours that children with ASD usually present. The stereotypical actions are usually manifested when

they are exposed to new environments or when there is a change in their daily routine and, as a result, they have a tendency to block the excessive feelings by exhibiting stereotyped behaviours, as hand-flapping (or hand wave), rocking, covering the ears, among others (CDC, 202 C.E.). Based on a previous work of the research group (S. Costa, 2014) with 15 children interacting with a humanoid robot, three stereotyped behaviours (hand wave, covering the ears, and rocking) were selected as the basis to develop the TSBD dataset, together with other 6 typical actions—idle, standing, clap, hand raise, pointing, and turn. These behaviours are also usually expressed by children with ASD during activity (S. Costa, 2014). An experimental setup was developed to collect the data. The Intel RealSense device was placed in three different configurations, as seen in Figure 5.5.

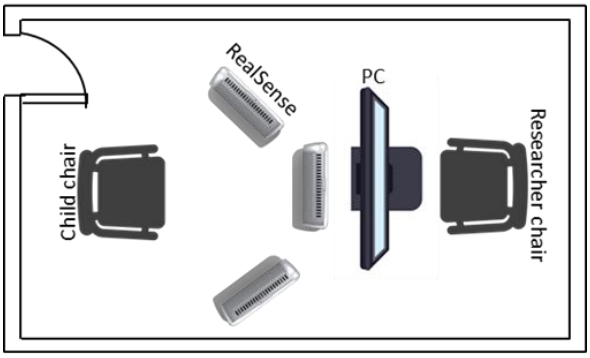


Figure 5.5: The experimental configuration used to extract the data for the dataset. The data was extracted by placing the Intel RealSense in three configurations: in front of the children, on his/her left, and his/her right.

The participants considered for the dataset construction were 32 typically developing children aged between 6 and 9 years old. Each child was seated in front of the Intel™ RealSense device and, for each action, performed 3 frontal samples and two side samples (from the left and right side). A total of 1440 samples were collected (160 samples per class).

The pipeline and processing

The overall system pipeline is presented in Figure 5.6. It starts by detecting the user and tracking the body joints. This process is done by using the Nuitrack SDK and the data from the Intel RealSense depth device.

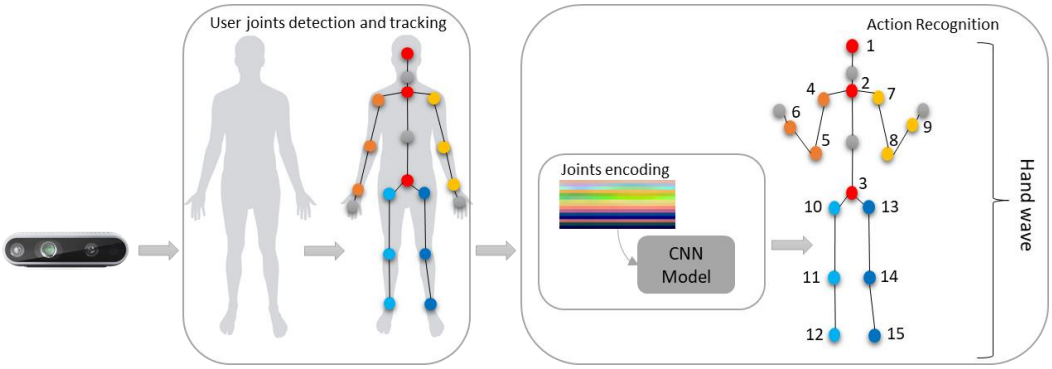


Figure 5.6: The overall system pipeline. It starts by detecting and tracking the user joints through the Nuitrack SDK. Then, the data is served as an input to the learning model to classify the action. The joints used are highlighted in orange (left arm, shoulder, elbow, and wrist joints), yellow (right arm, shoulder, elbow, and wrist joints), light blue (left leg hip, knee, and ankle joints), dark blue (right leg hip, knee, and ankle joints), and red (head, neck, and spine joints).

From the 19 joints that are returned by the SDK, 15 are selected: the right and left arm shoulder, elbow and wrist joints, the right and left leg hip, knee, and ankle joints, and the head, neck, and spine joints. The hand joints were excluded due to some inaccuracy when tracking them.

The selected joints are highlighted in orange (4, 5, and 6), yellow (7, 8, and 9), light and dark blue (10, 11, and 12 and 13, 14, and 15), and red (1, 2, and 3) in Figure 5.6. Finally, the joints data obtained over several frames were used as input for the action recognition model to classify the action.

To explore skeleton-based action recognition with CNN methods, it is necessary to represent a temporal skeleton sequence in an effective way to then feed it into a CNN model. Thus, the idea relies on encoding the spatial and temporal dynamics of skeleton sequences onto a 2D image structure. By using the static postures and their temporal dynamics, they can be shaped into a static color image structure (Pham et al., 2018). This color image representation can be fed into a CNN, enabling it to learn image features and classify them into classes to recognize the original sequence of skeleton data.

The encoding process of a skeleton sequence (S) with N frames into an image space representation takes place in this way: each 3D joint coordinate (x, y, and z) available for a skeleton sequence in a given frame (f) is normalized into the range of 0–255, $k(f)'$, by applying Equation (1):

$$k(f)' = 255 \times \frac{k(f) - \min\{c\}}{\max\{c\} - \min\{c\}} \quad (1)$$

where k is the coordinate (x, y, or z) to be normalized, and $min\{c\}$ and $max\{c\}$ are the minimum and maximum values of all coordinates in the sequence, respectively. The encoding process can be seen in Figure 5.7. In this new image representation, the three components (R, G, B) of a color pixel are the transformed skeleton joints coordinates (x, y, z)—x = R; y = G; z = B, i.e., the raw skeleton data of sequences are converted in 3D tensors, which is used as the input features for the learning model.

All skeleton sequences in image space has a 32 × 15 × 3 size, where 32 is the number of frames and the width of the image, 15 is the number of joints representing the height of the image, and 3 is the number of channels. The sequence size of 32 frames was chosen to analyze one to two seconds of movements to estimate an action.

The human body has four limbs and one trunk. This defines how actions are performed — a simpler action can use one limb, where more complex actions involve the coordination of two or more limbs and the torso. Following what is proposed in (Pham et al., 2018), each skeleton frame is divided into five groups—two arms (P_1, left arm, and P_2, right arm), torso (P_3), and two legs (P_4, left leg, and P_5, right leg)—which allows to keep the local motion characteristics, generating more discriminative features in image-based representations (Figure 5.8).

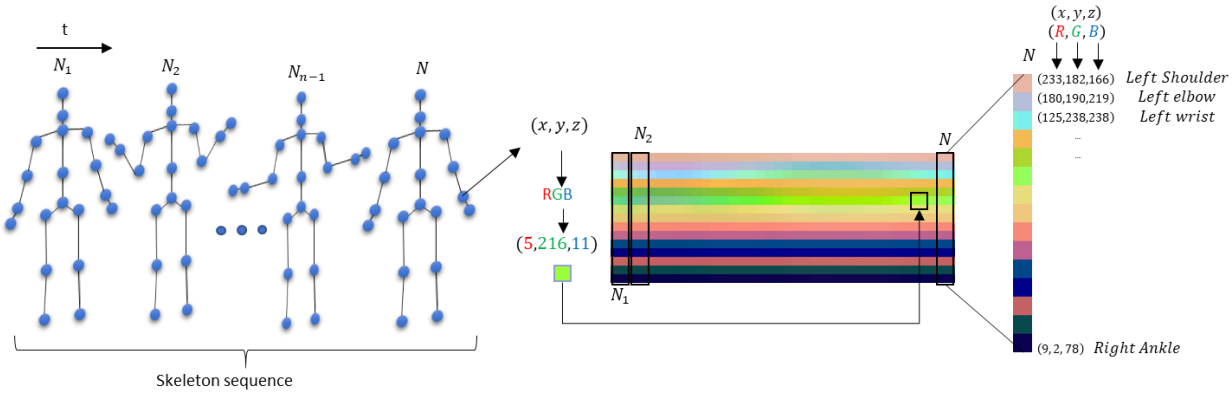


Figure 5.7: The joints are encoded by color in the image space, where N is the number of frames. For example, the encoding process for the wrist joint is highlighted in the middle, where the x, y, and z coordinates are normalized by using Equation (1) into the range of 0–255. Additionally, on the right, the encoding process for the N skeleton frame is highlighted in more detail. The final image dimensions are 15 joints (vertical axis) × 32 frames (horizontal axis) × 3 channels (comprising the converted coordinates to RGB).

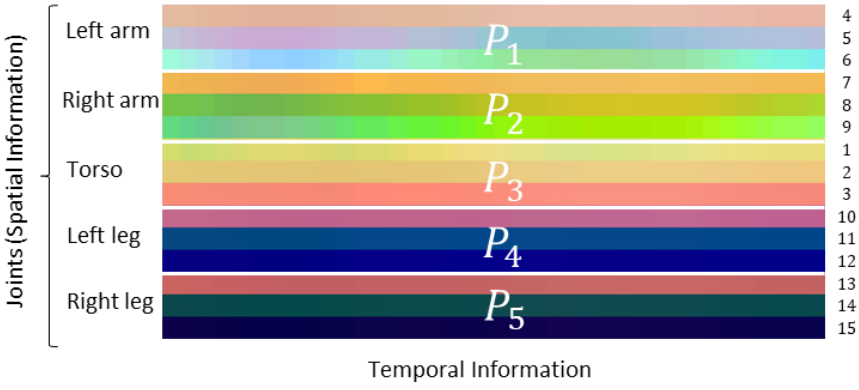


Figure 5.8: Representation of the pixels in image space according to the human body's physical structure. Each coloured strip represents the spatial position of a numbered joint (on the right) over the frames.

Figure 5.9 shows samples of skeleton-based representations of some actions from the TSBD dataset.

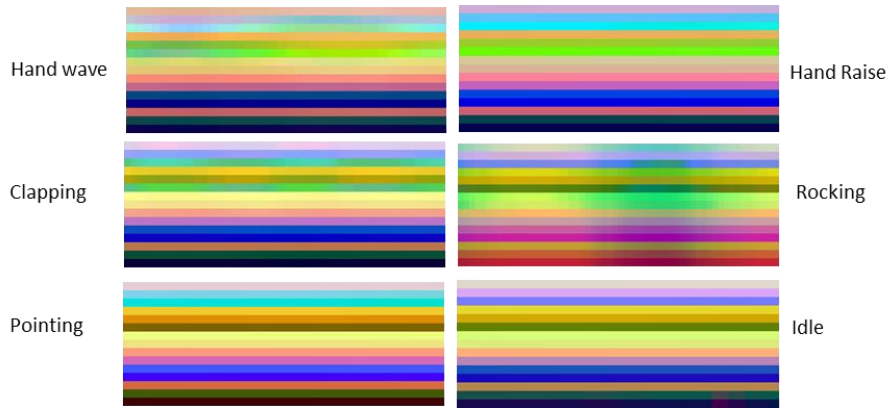


Figure 5.9: Some samples of skeleton-based image representations from the TSBD dataset used in this work.

Deep Learning Network Architecture

The DL network architecture chosen, depicted in Figure 6, consisted of six convolutional layers, each with a 3 × 3 kernel, as well as padding $p = 1$ and stride S = 1. The architecture used was based on the work proposed by (Ludl et al., 2019). Although a large spatial filter (such as 5 × 5 or 7 × 7) can provide an advantage in terms of their expressiveness and ability to extract features at a larger scale, it comes with a high computation cost. It has been shown (Szegedy et al., 2016) that a 5 × 5 convolution can be more computationally efficient, represented by two stacked 3 × 3 filters. Therefore, the network architecture implemented has three times two stacked convolutional layers in order to extract features at a larger scale, followed by a batch normalization layer(Ioffe & Szegedy, 2015), and the rectified linear unit (ReLU) was used as the activation function in the convolutional layers. A max-pooling layer with kernel size 2 × 2 was used after the second and fourth convolutional to reduce the spatial resolution by a factor of two. Then, after the last convolutional layer, a Global Average Pooling (GAP) layer was used. Since Fully Connected (FC) layers are susceptible to overfitting, which can compromise the generalization ability of the network, it was suggested to use global averaging of pooled layers to counteract this effect (Lin et al., 2014). They have the advantage of having no parameter to optimize; thus, overfitting is avoided at this layer. Moreover, global average pooling sums out the spatial information. So, it is more robust regarding spatial translations of the input.

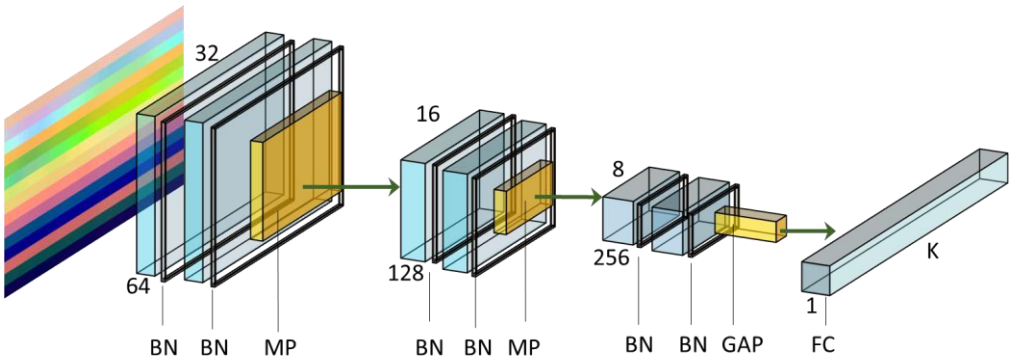


Figure 5.10: Network architecture used in the present work. BN: Batch Normalization; MP: Max Pooling; GAP: Global Averaging Pooling; FC: Fully Connected layer. The values 64, 128, and 256 are the number of filters used in each convolutional layer. The values 32, 16, and 8 are the resulted tensor width after the MP layer. As an example, the image representation (on the left) serves as input to the model. Adapted from (Ludl et al., 2019).

Finally, an FC layer was placed at the end to perform the final action classification. By being a small architecture, the training was fast due to few parameters with a short inference time, resulting in an overall small model while maintaining high accuracy. The network was trained on KERAS using a TensorFlow backend.

Experimental Results and Discussion

To infer if the joints sequence might impact the classifier performance, several skeleton configurations were evaluated based on the classifier accuracy achieved for the two datasets (KARD dataset and the TSBD dataset developed with typical and stereotypical actions).

Two main experiments were conducted:

- E1 compares the accuracy of the skeleton configuration B_1 (control group) with other five skeleton combinations (B_2, B_3, B_4, B_5, and B_6), where the sequence of the group (P_1 to P_5) was changed;
- E2 considered the joints of four skeleton combinations (C_2, C_3, C_4, and C_5), each with 15 joints randomly sequenced and again compared with the skeleton configuration B_1.

The joints were grouped according to the representation displayed in Table 5.1, where the groups are shown with the numbered joints following the same configuration presented in Figure 5.6 and Figure 5.8. For E1, the skeleton configurations used are shown in Table 5.1. For E2, the sequence of the joints was randomly obtained for each skeleton configuration as shown in Table 5.2. The goal was to see if the sequence of the joints in the image representation impacted the model performance.

Table 5.1: Grouped joints in four limbs. (P_1, P_2, P_4, and P_5) and one trunk (P_3) following the convention presented in Figures 1 and 3.

P_1	P_2	P_3	P_4	P_5
LA [1] (4, 5, and 6)	RA [2] (7, 8, and 9)	Torso (1, 2, and 3)	LL [3] (10, 11, and 12)	RL [4] (13, 14, and 15)

[1] LA: Left arm; [2] RA: Right arm; [3] LL: Left leg; [4] RL: Right leg.

Table 5.2: Skeleton configurations of groups of joints used where B_1 is the control group. Here the order of the joint groups is changed for each skeleton configuration.

SC [1]	Joint Groups Sequence				
B_1	P_1	P_2	P_3	P_4	P_5
B_2	P_1	P_3	P_2	P_4	P_5
B_3	P_1	P_4	P_3	P_2	P_5
B_4	P_5	P_2	P_4	P_3	P_1
B_5	P_3	P_1	P_2	P_4	P_5
B_6	P_5	P_4	P_3	P_2	P_1

[1] Skeleton configuration.

Table 5.3: Skeleton configurations used where B_1 is the control group. Here the sequence of the joints was randomly changed for each skeleton configuration.

SC [1]	Joint Number														
B_1	4	5	6	7	8	9	1	2	3	10	11	12	13	14	15
C_2	5	13	11	8	3	6	2	9	15	10	7	12	1	4	14
C_3	2	9	14	12	5	11	8	6	3	7	13	4	15	1	10
C_4	15	7	6	13	4	12	3	1	2	11	8	14	5	10	9
C_5	11	10	1	12	15	2	5	6	14	3	9	4	13	8	7

[1] Skeleton configuration.

In order to verify if the skeleton configuration could impact the accuracy of the model, a statistical analysis was performed based on the ANOVA (F-statistics). Then a pair-wise analysis was conducted using independent Student's t – test (t-statistics), when necessary, to understand the direction of the difference between the control group (B_1) and the other skeleton configurations.

The results concerning the classifier performance are also addressed. Cross-subject evaluation was performed in both stages of the present work.

The model was trained for 200 epochs, with a batch size of 64 and a learning rate of 0.05. A weight decay (L2 regularization) of 1×10^{-4} was used to counteract overfitting. Stochastic Gradient Descent (SGD) was used as an optimizer. The learning rate starts at 0.01, and it was reduced by a factor of ten every 50 epochs. The network was trained over 20 runs to ensure the reproducibility of the results and exclude random effects due to the training process (random initialization of layer weights, random shuffling of

data, and batch creation). Therefore, results are reported in terms of mean value with standard deviation (mean value ± s.d.%).

KARD Results

The KARD dataset was randomly divided into two datasets: the data of seven subjects were selected for training and three subjects for the validation data. The experiments E1 and E2 were conducted. Figure 5.11 shows the skeleton-based image representations for B_2, B_4, C_2, and C_4 for the class 'two-hand wave'.

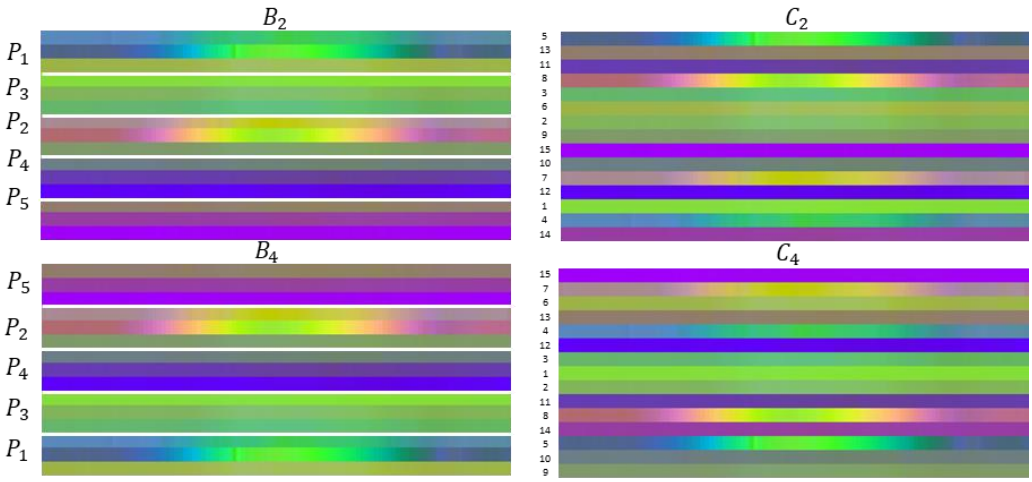


Figure 5.11: Some samples of skeleton-based image representations from the KARD dataset. On the left side, it is shown the images for the skeleton configurations B_2 and B_4. On the right side, the images for the skeleton configurations C_2 and C_4 are displayed.

For all skeleton configurations evaluated in E1, the differences are statistically significant in terms of accuracy ($F(5) = 418.41$, $p < 0.001$), where B_1 had the highest accuracy (94.0% ± 0.4%) (based on a pair-wise evaluation between B_1 and the other five skeleton configurations). Additionally, for skeleton configurations evaluated in E2 a similar conclusion was obtained ($F(4) = 432.67$, $p < 0.001$), where B_1 presented the highest accuracy (94.0% ± 0.4%).

Since the skeleton configuration B_1 presented the highest mean accuracy (94.0% ± 0.4%), Figure 5.12 shows the training accuracy over the 20 runs, and Figure 5.13 shows the error plot bar of the mean accuracy per class as well as the standard deviation for each of the 18 classes over 20 runs.

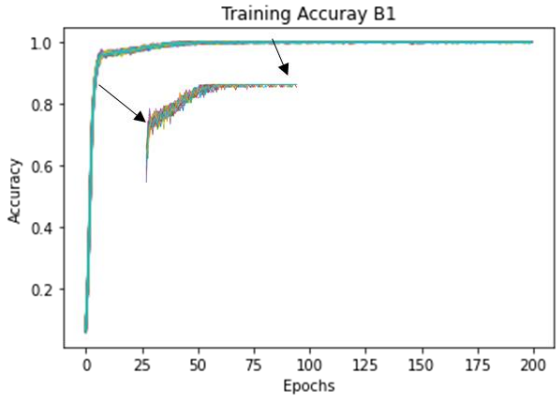


Figure 5.12: Plot of the training accuracy over the 20 runs. The accuracy for the first epochs (5 to 100) is high-lighted.

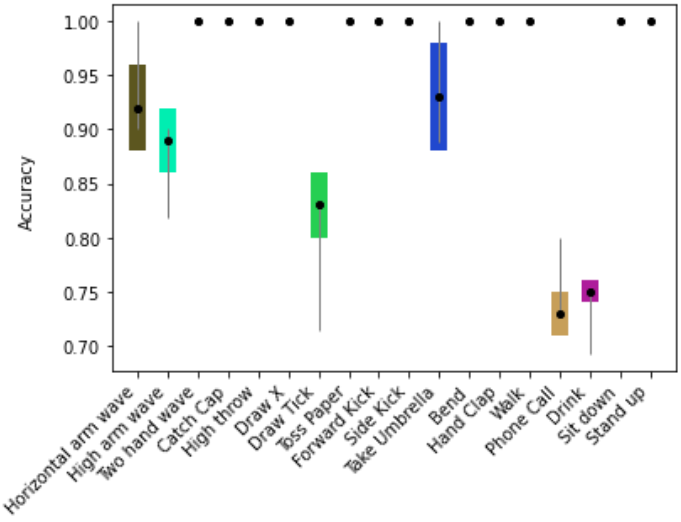


Figure 5.13: The error plot bar of the mean accuracy per class as well as the standard deviation for the model trained with the skeleton configuration B_1.

In general, the classes achieved mean accuracies of 100.0% (Figure 5.13) on the test data. Only 6 out of the 18 classes achieved a lower mean accuracy, namely: Horizontal arm wave (92.0% ± 4.0%), High arm wave (88.3% ± 3.2%), Draw tick (82.1% ± 3.6%), Take umbrella (92.8% ± 5.3%), Phone call (73.8% ± 2.6%), and Drink (74.1% ± 2.1%).

The Matthews Correlation Coefficient (MCC) was also used to assess the classifier performance. The model achieved an MCC of 93.7% ± 0.0%.

TSBD Dataset Results

The original dataset was randomly and balanced divided into three datasets: 20 subjects were selected for training, 5 subjects were selected for validation, and 7 subjects for the test data. The experiments E1 and E2 were performed. Figure 5.14 shows skeleton-based image representations for B_3, B_5, C_3, and

C_5 for the class 'HAND_WAVE'.

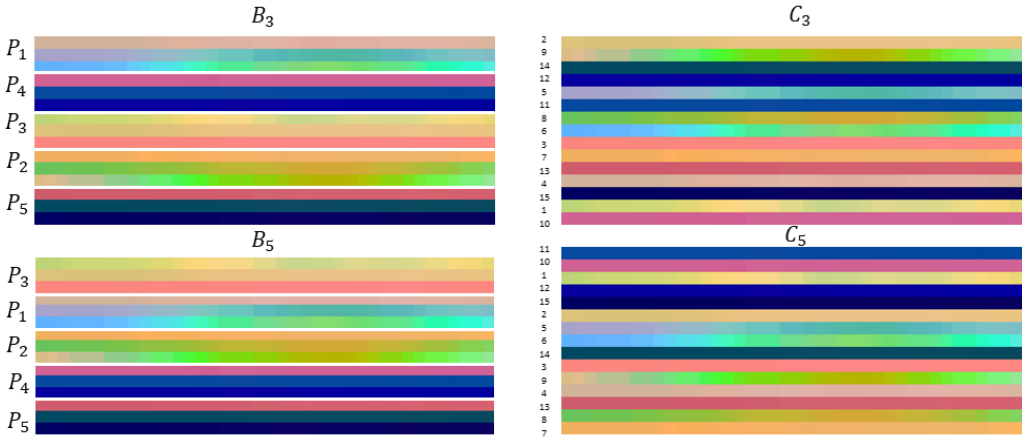


Figure 5.14: Some samples of skeleton-based image representations from the TSBD dataset. On the left side, it is shown the images for the skeleton configurations B_3 and B_5. On the right side, the images for the skeleton configurations C_3 and C_5 are displayed.

For all skeleton configurations evaluated in E1, the differences are statistically significant in terms of accuracy (F(5) = 17.34, $p < 0.001$), being the B_1 with the highest mean accuracy (93.8% ± 0.6%). However, for the pair-wise skeleton configuration B_1 and B_6 the difference is not statistically significant (t(38) = 0.38, $p = 0.35$) considering the validation dataset. Regarding the test dataset, similar results were obtained, that is, statistical differences in terms of accuracy were found (F(5) = 95.09, $p < 0.001$), with B_1 with the highest mean accuracy (92.4% ± 0.0%).

Additionally, for skeleton configurations evaluated in E2 statistical differences were found in terms of accuracy in both validation and test datasets (F(4) = 20.10 and F(4) = 107.95, respectively, $p < 0.001$), where B_1 presented the highest accuracy (93.8% ± 0.6% for the validation dataset and 92.4% ± 0.0% for the test dataset), based on a pair-wise evaluation between the B_1 and the other skeleton configurations. Indeed, the inter-class accuracy when comparing the control group (B_1) with one of the skeleton configurations that presents one of the lowest mean accuracies (C_2); in general, all classes mean accuracy drops (Table 5.4), and the class 'COVER_EARS' presents the most noticeable drop in terms of mean accuracy (from 93.2% ± 0.6% in B_1 to 83.2% ± 2.3% in C_2).

Table 5.4: Inter-class mean accuracy comparison between the control group (B_1) and C_2 with the present work test dataset.

SC [1]	IDLE	SD [2]	CLAP	HW [3]	HR [4]	PT [5]	CE [6]	TURN	RK [7]
B_1	85.9% ± 1.1%	97.1% ± 1.1%	85.1% ± 1.7%	97.2% ± 0.0%	100% ± 0.0%	100% ± 0.0%	93.2% ± 0.6%	83.2% ± 2.2%	92.0% ± 2.3%
C_2	80.5% ± 1.7%	95.9% ± 1.7%	84.3% ± 0.2%	94.6% ± 0.0%	100% ± 0.0%	97.4% ± 0.6%	83.2% ± 2.3%	86.7% ± 2.2%	88.8% ± 2.1%

[1] Skeleton Configuration; [2] STANDING; [3] HAND_WAVE; [4] HAND_RAISE; [5] POINTING; [6] COVER_EARS; [7] ROCKING

Figure 5.15 and Figure 5.16, on the left, show some samples of skeleton-based image representations for the classes 'COVER_EARS' (Figure 5.15) and 'ROCKING' (Figure 5.16) following the skeleton sequences B_1 and C_2, respectively. On the right, class activation maps are shown for the models trained with the same skeleton configurations. The class activation map technique uses the Global Average Pooling (GAP) layer in CNNs to indicate, for a particular category, the discriminative image regions used by the CNN to identify that category (Zhou et al., 2016). It could be used to interpret the prediction decision made by CNN. It is possible to observe that according to the sequence of the joints, the CNN is triggered by different semantic regions of the image. The model trained with the skeleton sequence B_1 correctly predicted the ground-truth class 'COVE_EARS' with the top prediction score of 0.43. Conversely, the model trained with the skeleton sequence C_2 incorrectly predict the sample as 'Clapping' with the top score of 0.49. Considering the samples for the 'ROCKING' class, both models correctly predicted the ground-truth class. However, the model trained with samples with the joints sequence C_2 presents a lower prediction score of 0.72.

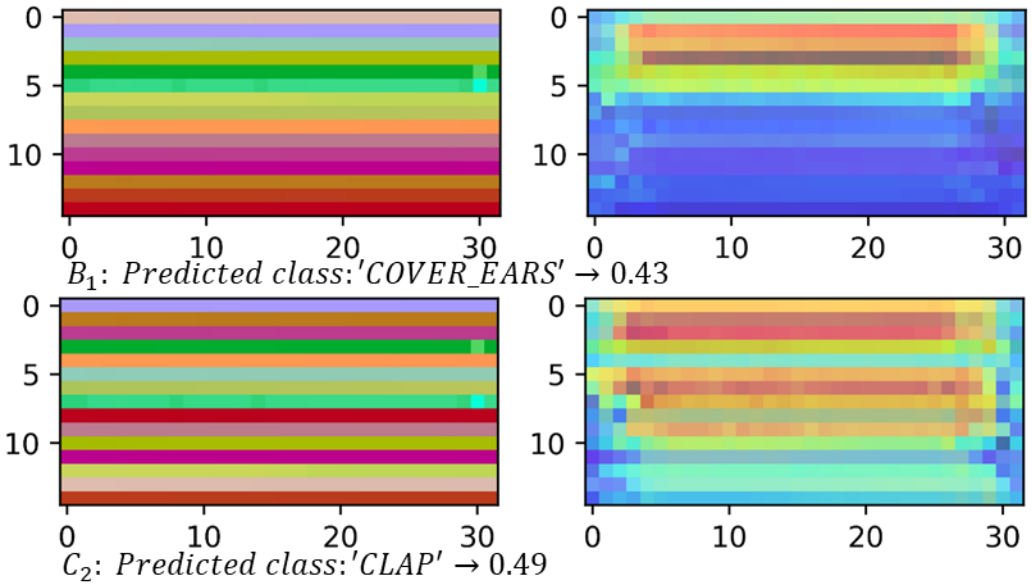


Figure 5.15: Skeleton-based image representations for the ground-truth class 'COVER_EARS' from the TSBD dataset. The predicted class and its score are shown below each sample on the left. On the right, it is possible to see the activation maps for each sample. It can be observed that the highlighted regions vary depending on the skeleton sequence, e.g., for B_1 the upper part of the region in the image is highlighted. On C_2 the regions highlighted are more spread.

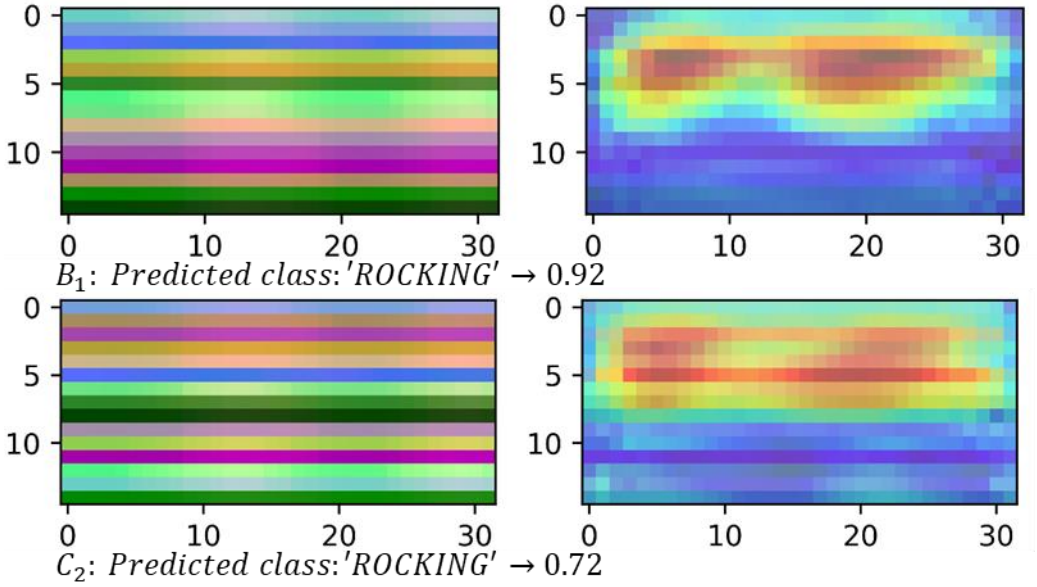


Figure 5.16: Skeleton-based image representations for the ground-truth class 'ROCKING' from the TSBD dataset. The predicted class and their score are shown below each sample on the left. On the right, it is possible to see the activation maps for each sample. It can be noticed that the highlighted areas are, in general, in the same region on the image for both B_1 and C_2.

Since the skeleton configuration B_1 presents the highest mean test accuracy (92.4% ± 0.0%), the following results used this configuration. Figure 5.17 shows the training accuracy plot of the model over the 20

runs. It is possible to see that the training converged rapidly and, in general, it remained stable over the 200 epochs.

As previously reported, the model achieved a validation accuracy of 93.8 ± 0.6% with 20% of the samples of the original dataset and test accuracy of 92.4 ± 0.0% with the test data of seven subjects.

The model achieved an MCC of 91.5% ± 0.4% on the test samples.

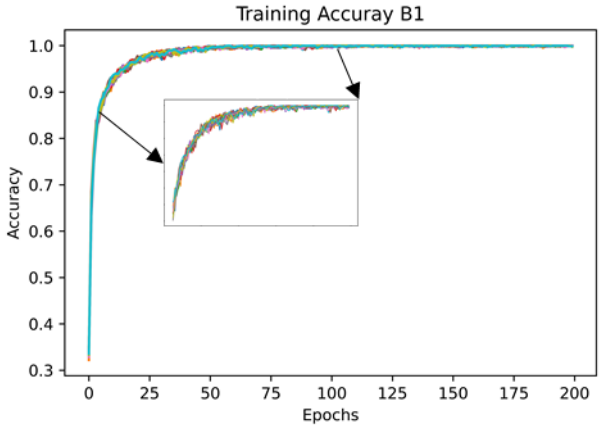


Figure 5.17: Plot of the training accuracy over the 20 runs. The accuracy for the first epochs (5 to 100) is highlighted.

Figure 5.18 shows the error bar plot of the mean accuracy per class as well as the standard deviation for each of the 9 classes over the 20 runs. It is possible to see that most classes had an average accuracy of over 90.0%.

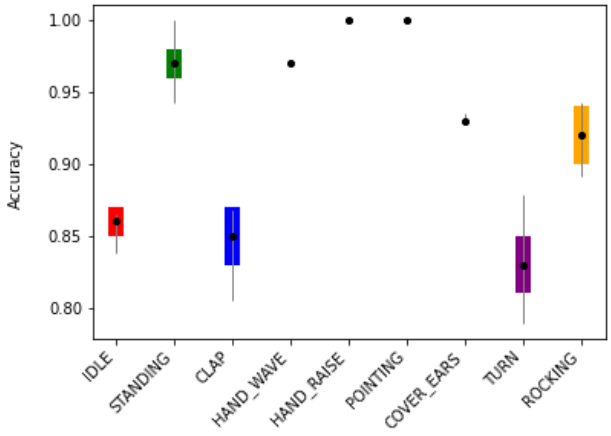


Figure 5.18: The error bar plot of the mean accuracy per class as well as the standard deviation for each of the 9 classes over the 20 runs.

In general, the classes achieved mean accuracies over 90.0% (Figure 5.18 and Table 5.4) on the test data. The classes 'IDLE', 'CLAP, and 'TURN' achieved, on average, the lowest accuracies (85.9% ± 1.1%, 85.1% ± 1.7%, and 83.2% ± 2.2%, respectively).

Figure 5.19 and Figure 5.20 show the generated image representations with the spatial and temporal

information that serves as input for the model, as well as frames of the actions 'CLAP' and 'TURN' performed by the subject in seated and upright positions, respectively. These representations are generated and then classified by the model in real-time. It is worth pointing out that these actions were performed by an adult subject.

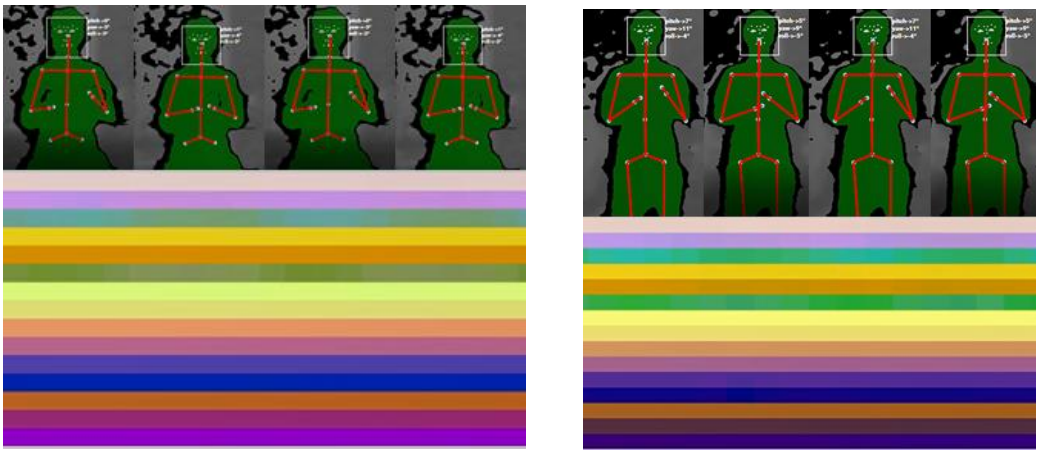


Figure 5.19: The generated image representations with the spatial and temporal information that serves as input for the model (**below**), as well as frames of the actions 'CLAP' (**above**) for the seated (**on the left**) and upright (**on the right**) positions. In both scenarios, the model successfully classified the action as 'CLAP.

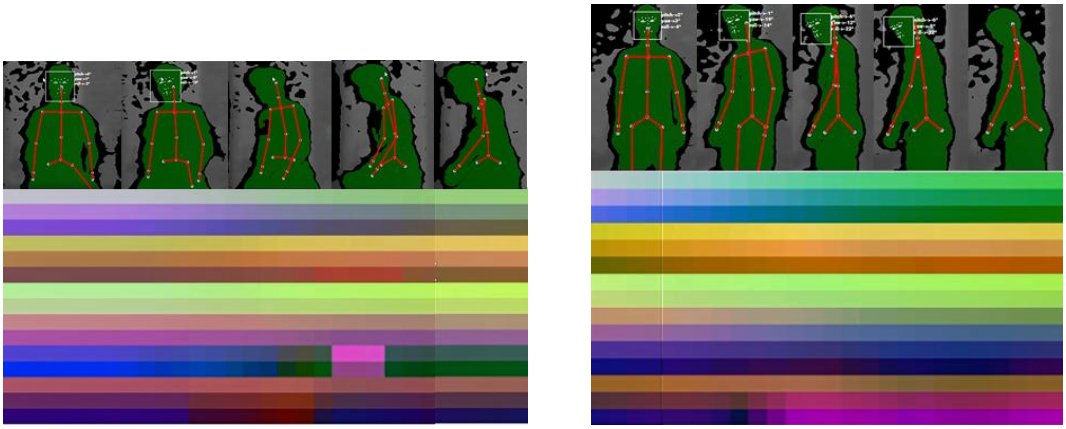


Figure 5.20: The generated image representations with the spatial and temporal information that serves as input for the model (**below**), as well as frames of the actions 'TURN' (**above**) for the seated (**on the left**) and upright (**on the right**) positions. In both scenarios, the model successfully classified the action as 'TURN'.

The computational time was calculated considering the average time for the Nuitrack SDK to retrieve the user joints coordinates plus the time to encode the joints coordinates into a sequence of frames and the classification time. The tests ran on a computer equipped with an Intel(R) Core (TM) i7-8650 quad-core CPU with 16 GB of RAM. The inference was done using OpenCV integrated with the Deep Neural Network module that allows importing saved TensorFlow models (OpenCV, 2020). Based on the performed tests,

the total maximum run time of the system is 0.032 s. Thus, the entire pipeline runs on average at about 31 FPS, which ensures the real-time capability of the system.

Discussion

Based on the two experiments (E1 and E2) conducted with the KARD and the TSBD datasets, it was found statistically significant differences in terms of accuracy between the control group (B_1) and the remaining skeleton configurations. Indeed, the model achieved a mean accuracy of 94.0% ± 0.4% by using the B_1 skeleton combination with the KARD dataset. Furthermore, 12 out of the 18 classes achieved a mean accuracy of 100% ± 0.0% on a benchmark dataset (as seen in Figure 5.13).

Based on the combinations considered, there was one exception where the differences were not found to be statistically significant for the pair-wise skeleton configuration B_1 and B_6 with the TSBD validation dataset of the present work. Concerning the test dataset, it was found statistically significant differences, with B_1 presenting the highest mean accuracy. Therefore, it was possible to infer that the sequence of the joints could impact the model's performance. This could be further noticed when comparing the inter-class accuracy with the present work dataset between the control group (B_1) and one of the skeleton configurations with the lowest mean accuracy (C_2), where in general, all classes mean accuracy dropped. This was more pronounced in the 'COVER_EARS' class, where it dropped from 93.2% ± 0.6% in B_1 to 83.2% ± 2.3% in C_2.

Additionally, it is possible to observe from the activation maps (Figure 5.15 and Figure 5.16) that different regions in the images are activated by some visual pattern. For the 'COVER_EARS' class, this visual pattern focused more on the upper member joint position, since the upper part of the image region for the skeleton sequence B_1 was highlighted in its activation map. Conversely, for the skeleton configuration C_2 the highlighted regions were spread across the image, and the sample was incorrectly classified as 'CLAP'. This might be due to the joints sequence being random, thus grouping the joints into limbs and trunk may allow keeping the local motion characteristics (Pham et al., 2018), particularly for some categories. Additionally, due to the nature of CNNs, it first captures local relationships, so when the joints were hierarchically ordered, the results tended to be better. This was, in general, mostly observed in the group's joints configuration and better preserved in B_1 configuration. So, by maintaining the local and global joints hierarchy, the model (B_1) performed better in comparison to the other joint's sequences. Considering the example for the 'ROCKING' class, in both cases, the predicted class was the ground-truth class, suggesting that the joints sequence did not impact as much on the model's final prediction.

However, it was possible to notice that the predicted class score was lower for the skeleton configuration C_2.

Since the skeleton configuration B_1 presented the highest mean accuracy, further tests were conducted to assess the model performance. By analyzing the remaining results, it was possible to observe that most of the classes had mean accuracies over 90% (Figure 5.18). However, 'IDLE', 'CLAP', and 'TURN' achieved mean accuracies of 85.9% ± 1.1%, 85.1% ± 1.7%, and 83.2% ± 2.2%, respectively. This may be due to some samples being similar in motion or, in the case of 'IDLE', some frames of the other classes could contain idle moments when the user did not move.

As shown in Figure 5.19 and Figure 5.20, the model could successfully classify the actions, considering they were executed by an adult subject from two different positions, seated and upstanding. It is worth pointing out that there was no data of adult subjects in the TSBD dataset. Furthermore, all the samples were recorded, as depicted in Figure 5.5, in a seated position. This could be mainly due to the normalization process that is scale and translation invariant (Pham et al., 2018). Thus, the model was able to successfully classify the actions for these additional scenarios, effectively learning the skeleton motion characteristics.

So far, only a few works have tried to detect stereotyped behaviors of children with ASD. One of them (Goncalves et al., 2012a) has the goal of detecting the hand wave (flapping) gesture with the Kinect sensor, achieving an accuracy of 51%. Another work (Jazouli et al., 2019) tried to detect other stereotypes behaviors besides the hand wave. In order to have the same basis of comparison, Table 5.5 compares both works with the current model in terms of accuracy for the similar studied behaviors 'HAND_WAVE' and 'ROCKING'.

The present method achieved a similar performance for 'ROCKING', while it outperformed the other works for the detection of stereotyped behavior 'HAND_WAVE'.

Concerning the performance of the pipeline, it runs on average at about 31 FPS on a quad-core CPU, being fast enough for most real-time applications.

Table 5.5: Comparison in terms of accuracy of other works in the literature that focus on the detection of stereotypical behaviour.

Work	'HAND_WAVE'	'ROCKING'
(Goncalves et al., 2012a)	51.0%	-
(Jazouli et al., 2019)	91.5%	92.2%
Current Model	97.2% ± 0.0%	92.0% ± 2.3%

5.1.2 Summary and Conclusions

The proposed approach learns directly from the original skeleton data in an end-to-end manner.

The selected non-verbal behaviors for building the dataset consisted of typical and stereotypical patterns. In the first stage, two experiments (E1, sequence of groups, and E2, sequence of joints) with the KARD and the TSBD datasets were carried out in order to infer if the sequence of the joints in the skeleton image-based representation might impact the model accuracy. It was found statistically significant differences in terms of accuracy between the control group ($B_1 - P_1, P_2, P_3, P_4, \text{and } P_5$) and the remaining skeleton configurations. Thus, indicating that the sequence of the joints can impact the model's performance. Since the skeleton configuration B_1 presented the highest mean accuracy on the test dataset, further tests were conducted to assess the model performance. The proposed model achieved an overall validation accuracy of 93.8% ± 0.6% (with data from five subjects) and a test accuracy of 92.4% ± 0.0% (with data from seven subjects). Additionally, the present approach delivers state-of-the-art performance compared to other methods on the detection of stereotyped behaviors (Table 5.5). More specifically, it achieved similar or even better performance when classifying stereotypical behaviors—'HAND_WAVE' and 'ROCKING'.

The whole pipeline is able to work in real-time, running on average at 31 FPS entirely on an Intel(R) Core (TM) i7-8650 quad-core CPU in contrast to most state-of-the-art approaches that use dedicated GPU hardware when implementing deep learning approaches.

5.2 The Reasoning Layer

The reasoning layer is responsible for deciding the interaction flow of the session by taking as input the processed data from the perception layer. The interaction flow can be influenced by the detected behaviours as well as the child's performance, namely the time he/she might take to answer the robots prompts.

The experimental setup has three main elements – the robot (ZECA), the child, and the experimenter – as depicted in 3.2. Since the goal is to adapt the robot's behaviour to the child actions in a supervised manner it is important to define the key actions of the child that ZECA should intervene, the number of times ZECA could prompt the child and the span of time for ZECAs interventions. Furthermore, as the experimenter/researcher is included in the experimental setup, he/she will also intervene when pertinent. Thus, to define these points it was conducted a number of semi-structured interviews (SSI), namely 4, with specialists as well as with a parent. Participants were selected according to the following criteria of intentionality: professionals with different backgrounds in terms of training area (occupational therapists and speech pathologists); who had a minimum experience of five years in supporting children with ASD; parent of a children with ASD; individuals with technological background in robotics namely engineers who have interacted with the target group; and who had an interest in and availability in participating in the SSI interviews. An informed, clarified, and free consent request, which contained the ethical procedures to be followed in the SSI session (see Appendix A.5), besides an acceptance statement about audio and video recording were signed by each participant prior their session.

The SSIs were conducted individually with each participant with an average duration of 60 min. The goals of the SSI were the following:

(1) Define the robot actions accordingly to the child's behaviour;

(2) Define the experimenter/researcher actions accordingly to the child's behaviour

(3) Define a limit of the robot's behaviour adaptation accordingly to the child's actions.

5.2.1 Structure of the Interview

The structure of the SSI is the following:

- Opening:
  - Delivery and signature of the informed consent;
  - Brief introduction about the *Robótica–Autismo* project;
- Presentation of the study:

  - Explanation of the goals of the SSI session;
  - Presentation of the experimental setup and its various components – namely the robot (ZECA), the OPT (PlayBrick), and the sensors – with depth explanation on how each component works and interacts with each other;
- Guide of Questions:
  - Personal identification, namely age, profession, years of experience with the target group (children with ASD) as well as the degree of severity, the number of children with ASD he/she have worked with and if the therapy sessions are conducted individually or in group.
  - Ask if the participant has any knowledge of a system with similar capabilities.
  - Based on their experience over support sessions with children with ASD, ask:
    (Q1) The most common typical and stereotypical behaviours shown by the target group;
    (Q2) To justify based on their experience how he/she usually intervene on such behaviours;
    (Q3) To choose from the mentioned behaviours the ones that in their view and taking in consideration the system capabilities ZECA could intervene;
    (Q4) To define the number of times and ways the robot should intervene as well as the time between interventions;
    (Q5) To define when the experimenter should intervene considering the behaviours/situations listed by the participant, thus defining the limit of the robot intervention to certain actions of the child;
    (Q6) To define the time that the system should give the child to answer and if and how many times it should be extended.
    (Q7) The total duration in time of the session taking into consideration the activities – Recognize and Storytelling;
    (Q8) To classify the experimental configuration, i.e., if he/she considered it as dyadic with mediation or triadic.
- Closure:
  - Acknowledgment to the participant;
  - Further contact;
  - Report on how the information is going to be used and next steps of the project.

5.2.2 Results

In Table 5.6 the main participants' characteristics (personal identification) are summarized.

Table 5.6: Participants characterization.

Participant	Category	Age	Profession	YoE[1]	NoC[2]	DoS[3] | G/I[4]	Recordings (V, A)[5]	KoS[6]
P1	Professional	28	Occupational Therapist	6	+30	ALL | Both	Yes	No
P2	Professional	35	Speech Therapist	6	+30	ALL | Both	Yes	No
P3	Professional	34	Software Engineer	6	+30	Mild | Both	Yes	Yes
P4	Parent	44	Nurse	8	1	Mild | -	No	No

[1]Years of experience; [2]Number of children; [3]Degree of Severity; [4]Interaction in group or individually; [5]Recordings of video and/or audio; [6]Knowledge of similar systems.

All participants with exception of one as seen in Table 5.6 authorized the recording (video and/or audio) of the session. The participant P4 data were obtained through annotations done by the interviewer of the participant answers to the questions during the session. Thus, data was transcribed verbatim and crossed-referenced with notes taken throughout the SSI session.

Transcripts of the authorized recordings were analysed line by line, broken into discrete parts and grouped/coded together by similarity of content and/or meaning according to the goals of the study, by using a qualitative data analysis software (webQDA). The webQDA is a qualitative and web-based data analysis software intended for all researchers and professionals conducting qualitative research (*WebQDA – Qualitative Data Analysis Software*, 2022). It allows the analysis of text, image, video, audio sources, among others, collaboratively, synchronously or asynchronously.

To ensure data integrity multiple methods of data collection were used – video/audio and annotations – and more than one person was involved in the analysis of the data.

Figure 5.21 shows the assembled tree code used to group the participants answers into codes. This tree code was constructed based on the participants answers to the SSI questions. A word cloud of the 50 most frequent words mentioned across all the SSIs, Figure 5.22, were also designed, summarizing the main concerns and topics based on participants ideas.



Figure 5.21: Tree code based on the questions and participants answers.



Figure 5.22: 50 most frequent words said in the SSIs (in Portuguese).

Regarding the question about if any of the participants had any knowledge of a system with similar capabilities, in general, they said no with the exception of participant P3 who mentioned that has knowledge of systems with robots – *"...with the robot yes, with the other things no ..."* (P3).

Relating to the remained questions, the corresponding answers were the following transcribed bellow with the data of the participants P1, P2, and P3. Concerning the contributions of P4, these are identified in the questions Q6, Q7, and Q8.

(Q1) The most common typical and stereotypical behaviours shown by the target group:

- The behaviours could be grouped in three categories – typical, stereotypical, and aggressive – as mentioned by the participants (P1, P2, and P3): "*By the ability to point is not it? that pointing is also non-verbal behaviour...*" (P3); "*We had agitation, a lot of agitation, motor and behavioural.*"(P2); "*Sometimes aggressive behaviour, without a doubt...*"(P2); *"...they have a lot of regulatory behaviours that stereotypies don't just go through flapping ... they*

go through running around the room ..., when they get too excited, this is called sensory seeking." (P1).

Thus, the mentioned typical behaviours were the following: pointing, standing, hands-up, apathy, cover ears, and leaning. The most stated stereotypical behaviours were rocking (balancing), and hand-flapping (hand shaking). Additionally, aggressive behaviours were also mentioned by the participants, which can be auto-aggressive (towards him/herself) or hetero-aggressive (towards others).

(Q2) To justify based on their experience how he/she usually intervene on such behaviours:

- The participants said they usually resort to touch – "*the touch, bringing the child to the activity*"(P2) or by verbal commands "*I usually go for the verbal part so I don't get too invasive... huh? Let's play?*" (P1) in case the child is distracted or listless. Taking into account stereotypical behaviours, the participants mentioned that it is usually a self-regulation behaviour, suggesting that they would not intervene – *"...rocking, for example, gets into the same situation, for example, as covering the ear if it's while for self-regulation, isn't it?"*(P2); *"... they are regulating themselves taking into account that they are sometimes slower or faster, or try to organize their thinking so that they can issue a response..."*(P1) – further adding – "*It's as if, physiologically, the person needs that, so if they need that and that even keeps them focused, yeah, it's not, it's not destroying anything.*" (P2); "...*it's a thing to self-regulate to allow for to organize my thinking. I would not intervene*"(P1). Conversely, in case of aggressive behaviours, it should be stopped – *"...must block aggressive behaviour..."*(P3).

(Q3) To choose from the mentioned behaviours the ones that in their view and taking in consideration the system capabilities, ZECA could intervene:

- All participants agreed that ZECA could intervene in the cases of distraction and/or when the child is taking time to answer – *"Okay, to call attention..."*, *"...the child does not respond..."* (P3); *"... and the robot speaks... to ... to try to direct attention"* (P2); *"...draw attention..."*(P1).

  However, even in these and other behaviours both the robot and the experimenter could intervene – *"...but there are other behaviours that maybe, it could even be a robot and the experimenter. There is going to be a level that the robot will try and then it will have to be the experimenter..."* (P2) – further stressing this – *"The child will always have moments of out-of-focus ..."*; *"And that's where the experimenter comes in, as I also think the*

experimenter can, I don't know, help ZECA to regulate when it should or should not, because otherwise we will spend entire sessions catching the child's attention."(P2). Additionally, it was also suggested that ZECA should not react to every behaviour – *"Ham, doesn't seem like it has to react to every gesture, does it? It can get weird."* (P3).

Concerning stereotypical behaviours, it was stated that ZECA should have no reaction – *"It's part and I think there should be no reaction"*, *"...I consider that the robot should not react."* (P3). Participant P2 further stress this idea by affirming that the experimenter should have this sensitivity – *"Well, I think the experimenter here has a very important role. To be able to see if the behaviour is being self-regulating or being dysregulated, right?"(P2) "Therefore, I think that the experimenter here has to have this sensitivity."* (P2).

(Q4) To define the number of times and ways the robot should intervene as well as the time between interventions:

- Following the behaviours mentioned by the participants where ZECA could intervene, for calling attention the participants agreed that ZECA should do it one up to three at most – *"Okay, so call attention one time."* (P3); *"... robot talk to her/him try to direct the attention once..."* (P2); *"Draw attention, I would go to the limit up to 3 times..."*(P1). In case the child gets up, the participants suggest that the robot could try to call the child for the activity once – *"Zeca could it once"* (P3).
- Most of participants mentioned the use of verbal communication as a way for the robot intervene. In case the child is taking long time to answer ZECA could reinforce the question - *"How did I feel?"* (P3). To call the attention – *"João, shall we continue playing?"*, *"...(It did not work) try "hey, let's play?""*(P1). In case the child gets up, ZECA could combine verbal and non-verbal communication – "It can say: *"Come to me", right?"*(P2); *"...or close down in some way."*(P1) – show a sadness reaction. Additionally, if the child is taking more time to answer than the usual, ZECA could give more time – *"...give time for the answer"* (P3). Additionally, in the Storytelling game it could be significant to repeat the story. In general, participants stress this idea by stating that it should be prompted to the child the option of hearing the story again and in case his/her response is affirmative, it should be repeated – *"I wouldn't tell the story again, it is super simple... Unless, the child asks to repeat it, it can happen in sometimes, they may want to hear it again."*(P3); *"If we have the answer of yes, we repeat the story..."* (P2).

(Q5) To define when the experimenter should intervene considering the behaviours/situations listed by the participant, thus defining the limit of the robot intervention to certain actions of the child:

- Considering aggressive behaviours, these should be blocked – *"...must block aggressive behaviour..."*(P3). For the stereotypical behaviours, as mentioned in Q3, the experimenter should have the sensitivity – *"Because then, for a person who has never seen stereotypic behaviour, he will want to stop right away, because he thinks that it is to stop, don't you?"* (P2) *"These are regulatory behaviours and there, without a doubt, the experimenter has a lot of impact."*(P2). Participant P1 further stress to not intervene if the stereotypical behaviour is self-regulating – *"This is one thing to self-regulate to allow me to organize my thinking. I would not intervene."*(P1). Moreover, P1 also adds that neither the robot or the experimenter should intervene – *"Neither the robot and the experimenter...It's part and I think there should be no reaction", "...I consider that the robot should not react."* (P3).
- P3 also suggests that the experimenter should intervene in cases where the child has some technical problem with the material – *"I would say that the experimenter should intervene if there is a technical problem and if the child asks you to share something, come from the child."* (P3). Additionally, P3 further adds – *"I think if there is a non-verbal behaviour of the child towards the experimenter, he should respond, right? Otherwise, we weren't stimulating social interaction, but it shouldn't respond to say, that's right. For example, say: "Do you think that is the correct answer? Right, then submit.""* (P3).
- For calling the attention, getting up, and in case the child is not or taking more time to answer the robot prompts, the participants agreed that the experimenter should intervene – *"Maybe... because sometimes the child can be distracted with something, right? Could have nothing to do with not paying attention, his/her focus may change..."* (P3); *"If there are frequent moments of inattention... the researcher has to enter..."*, *"Therefore, here many times the experimenter may also have the need..."* (P2). Considering the getting up behaviour, P1 stated – *"if he/she gets up and runs, I advise the experimenter to intervene first."* (P1). P2 also supports this idea – *"He/she got up and dispersed...then the experimenter enters... the touch, bringing the child to the activity."* (P2). Additionally, participants mentioned that the researcher should intervene by aiding the child – *"I will defend whenever yes, because it is a complement."*, *"...by for example explaining why the robot was surprised (in the story)..."*, *"...or ask: "Do you want help to understand the story?""*(P1). P2 also shares this opinion, stating that the researcher could intervene by

reducing the options available – *"I'd take the Brick, so let's see: "Is he happy or are you sad?" An alternative, right? In order to have an answer from the child."*, *"... reduce the number of responses"*, *"...the researcher has to reduce the options..."* (P2).

(Q6) To define the time that the system should give the child to answer and if and how many times it should be extended:

- Concerning the response time, P1 and P2 agreed that the overall time should be at most 1 minute – *"Okay, I think a minute...important because it always depends on the child's processing"* (P2); *"So I would say until I don't know a minute or two at most."*; *"I think I'll be honest, I think after 1 minute..."* (P1). During this time ZECA could intervene one or two times for example every 30 seconds – *"I start counting to 30 and then, from there, I get into action, because I think that later, more than that, they get lost with their own thoughts..."* (P1); *"The child does not respond in 30 seconds... then the robot asks again..."*, *"...if after this minute the child does not respond and I put the experimenter"* (P2). P3 also suggests that ZECA could intervene 2 times but with a shorter period between the interventions – *"Maybe also 2 times", "I would say 7 to 10 seconds."* (P3). Participant P4 could not suggest a time limit - *"I don't know. Time is a difficult thing to establish because every child is different."* (P4). However, P4 also stated that ZECA could try 2 or 3 times – *"...maybe repeat/try 2 or 3 times."* (P4)

(Q7) The total duration in time of the session taking into consideration the activities – Recognize and Storytelling:

- P1 and P2 agreed that the session time overall should be at most 30 minutes that includes the introduction plus the activity time considering the recognize and storytelling activities – *"So 25 minutes/half hour at most."* (P1); *"We should have in mind ... an initial conversation with the child, ... structuring time..."*, *"I think it must be counting 30 minutes."* (P2). Furthermore, the focus time in the activity P1 and P2 suggested 10/15 minutes – *"...10 minutes I think they could do it..."* (P1); *"In focus, activity would be 15 minutes."* (P2). Participant P3 suggested the time for each task should be 5 minutes, thus in total 10 minutes – *"I felt like 5 minutes, 5 minutes doing the same task was the balance"* (P3). Additionally, P4 suggested a longer time – *"50 minutes (with introduction time plus activity time)"* (P4).

(Q8) To classify the experimental configuration, i.e., if he/she considered it as dyadic with mediation or triadic:

- P1 classified it as dyadic with mediation – *"Dyadic with mediation because the robot is aimed at the child."* (P1). However, the rest of the participants (P2, P3, and P4) classified it as triadic – *"So I would say triadic"* (P3); *"Triadic."* (P4). Participant P2 further stress this idea – *"Each time I feel like triadic."*, *"I don't see him (experimenter) sitting in the corner, in the corner of the room..."*, *"I see him next to the child."* , *"He is the mediator with the most social competence and I see it in a triadic way, without a doubt, because he will have to be... He will have to be a highly qualified researcher with a very active participation in the investigation."*(P2).

Following the ideas given by the participants during the SSIs, the robot adaptation behaviour was programmed as depicted in Figure 5.23. The flowchart shows the robot's adaptation procedure in different scenarios taking into account the recommendations of the participants of the SSIs. It is worth to mention that the robot's prompts are only displayed after the system notifies the experimenter and it is greenlighted by the experimenter as this is a supervised architecture.

When the process starts, the experimenter chooses the activity to be played – Recognize or Storytelling. In both cases, the session flows as depicted in Figure 5.23 section 1. During the session, if the system detects at any time that the child got up, the system informs the experimenter and if the experimenter agrees, the system enters into pause mode (the activity is paused). Then the procedure follows the steps depicted in the section 3 (with the titled box "In any case:") of the Figure 5.23.

During the "waiting for answer" stage, the system follows the actions illustrated in the section 2. The "child distracted" routine (section 5) only runs during this stage as the robot face prompts during the Recognize and the stories prompts during Storytelling activities have a brief duration in comparison to the time the child can take to answer the robot prompt. Concerning these distraction patterns, participants mentioned that a child will always have moments of out-of-focus, and in these times is where the experimenter can more easily intervene (by touch for example) in comparison to the robot. Otherwise, if executed by the robot, ZECA could spend entire sessions trying to catch the child's attention which could break the interaction. Therefore, this implementation follows the recommendation that was suggested by the SSIs participants as to not constantly call the child attention. Finally, concerning typical and stereotypical behaviours the programmed procedure follows the actions described in section 4. Additionally, Figure 5.23 shows where the experimenter would intervene.

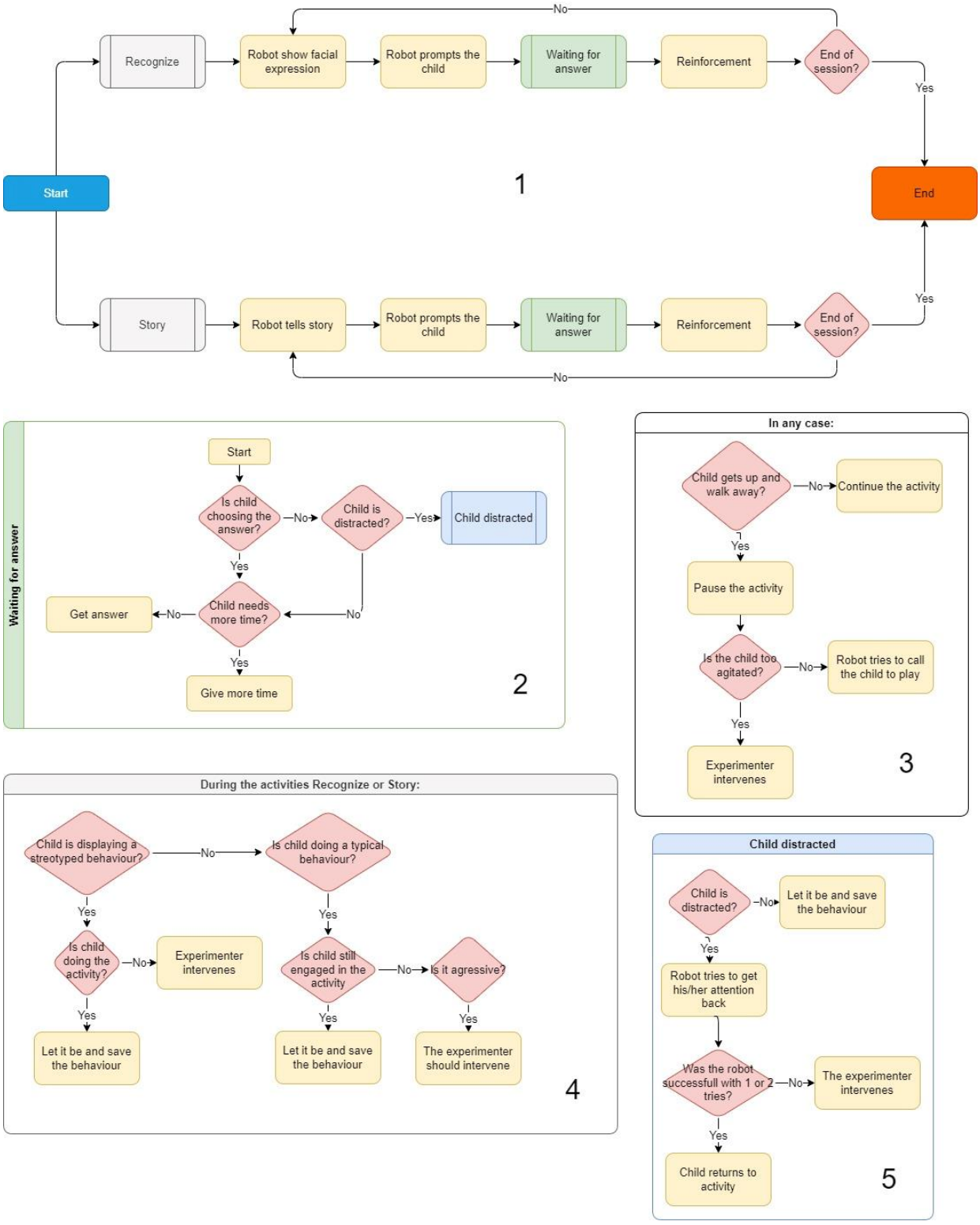


Figure 5.23: Robot adaptation procedure flow.

5.2.3 Summary and Conclusions

In order to decide when and how the robot ZECA could intervene during a support session, a collection of semi-structured interviews was conducted with professionals with different backgrounds and a parent. With their feedbacks it was proposed in Figure 5.23 the ZECA adaptation procedure, where it shows the possible scenarios where ZECA could intervene.

It is also possible to conclude that the experimenter needs to have the sensibility to supervise when ZECA should intervene or not even in the behaviours where the participants agreed that ZECA could intervene. Those key behaviours could be moments of distraction, when the child is taking a longer time to answer, if the child gets up, among others. ZECA could intervene by calling the child's attention (through verbal commands), give more time for the child to answer or aid the child to use the OPT (using verbal prompts), display a sad non-verbal behaviour (e.g. sad facial expression) in case the child gets up, among others. Thus, it is possible to analyse from the participants answers, the experimenter should have the control of the session, deciding if ZECA should or not intervene in all scenarios.

5.3 The Actuation Layer

The actuation layer is responsible for providing the feedback to the child, through the robot and OPT. The dynamics of this layer is influenced by the outputs from the reasoning layer. The reinforcement feedback provided to the child can be configured in both the robot and the OPT. The OPT reinforcement feedback can be done via haptic feedback, by displaying write/wrong images as well as displaying colours (such as red/green) by lighten up the bar LEDs depending on the correctness of the answer given by the child, Figure 5.24.

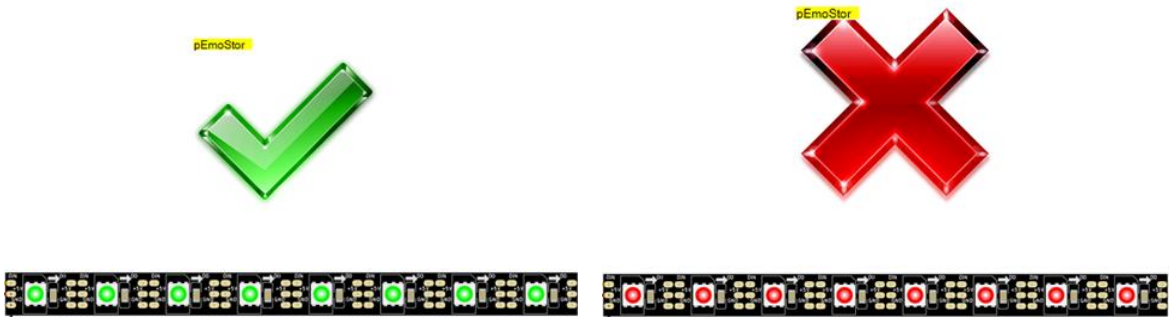


Figure 5.24: Reinforcement feedback images prompted by the OPT - on the left for a correct answer, on the right for a wrong answer. The LEDs showcase the colours green (on the bottom left) or red (on the bottom right) according to the correctness of the answer.

ZECA can display verbal, non-verbal (gestures), and sound prompts as a reinforcement to the child's answer depending on the correctness of the answer. The reinforcements are simultaneously displayed by ZECA and the OPT.

In addition, since it is a supervised architecture, it was developed a Graphical User Interface (GUI) where it is possible to visualise the data from the different devices from the perception layer (Figure 5.25), establish the connection with the robot, select the activity to be played, see the session flow (Figure 5.26), visualise the child information (Figure 5.27), among other things. Since it is a supervised architecture, at any time, the researcher/therapist can pause/resume and start/stop the activity on the GUI. The GUI prompts the detected behaviours and recommend possible actions that will be carried out automatically, unless the researcher/therapist intervenes. This intervention can be done through the GUI (Figure 5.26) by pressing the green "Yes" button.

It is possible to record the session data such as child performance, head orientation, facial cues, and the non-verbal behaviours detected. Through the GUI it is established the TCP/IP connection with the robot (ZECA) and the with the OPT. Furthermore, if desired the researcher/therapist can extend the session time.

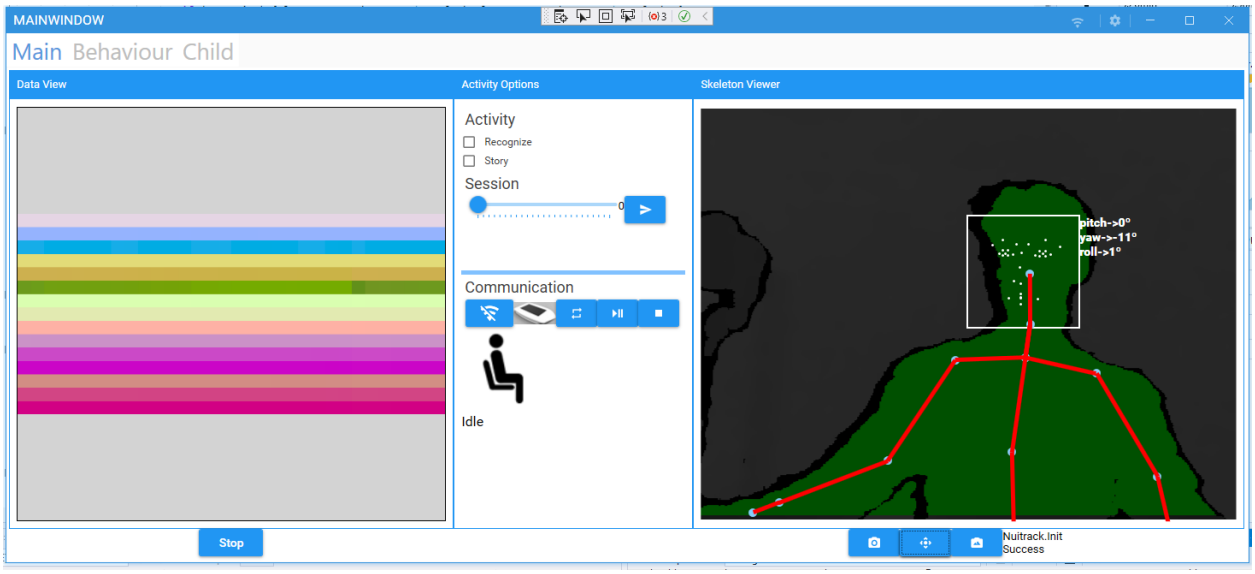


Figure 5.25: GUI main tab where it is possible to visualize the skeleton data and gestures as well as configure the session time, choose the activity and establish the communication with ZECA and OPT.

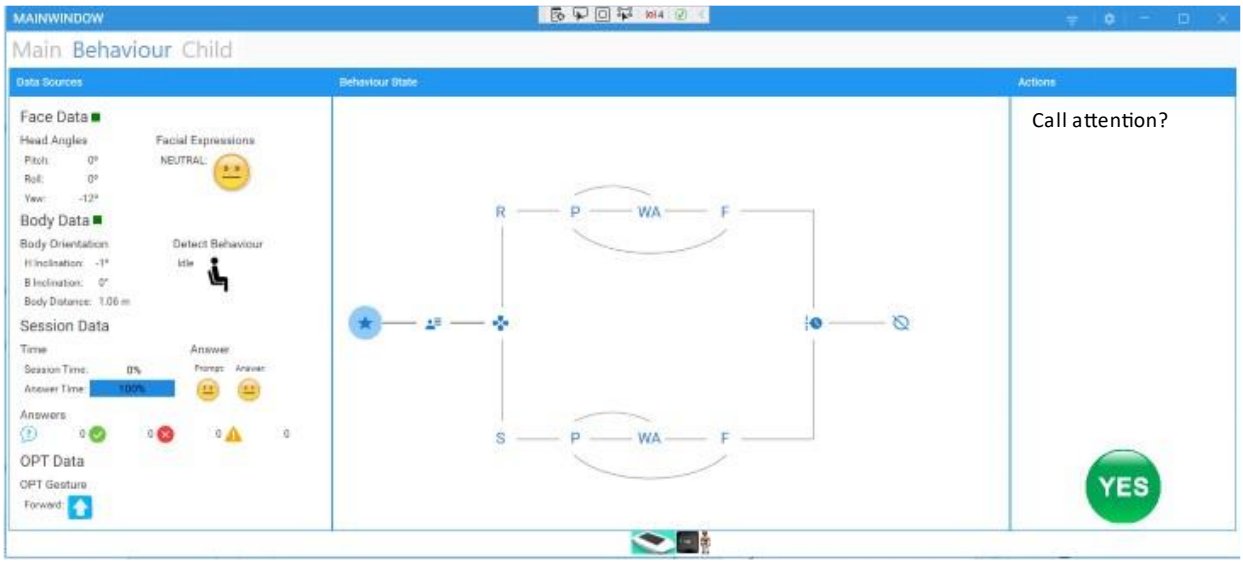


Figure 5.26: GUI behaviour tab displaying face, body, and session data. It is possible to visualize the session flow.

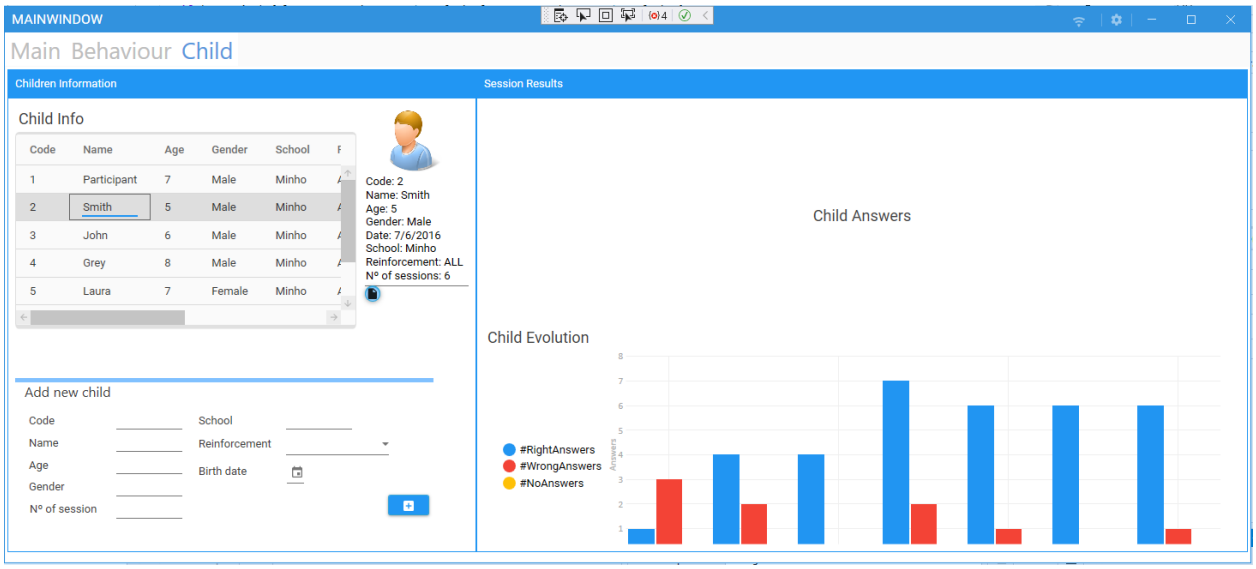


Figure 5.27: GUI child tab where it is possible to add the child data and visualize the child performance over the sessions.

Additionally, to the GUI, the system also provides feedback about the session to the experimenter/researcher through the use of a remote device with as screen as depicted in Figure 5.28. Since having an additional screen could interfere in the interaction dynamics, by for example being an additional device that could distract the child, the remote device offers an alternative way where the research/ experimenter can easily glance and decided if the action should be carried out or not. The system informs the recommended action in both the GUI and the device at the same time. The experimenter/researcher can confirm it or not by pressing the "Yes" or "No" buttons, respectively. In addition, if necessary, the experimenter can pause/resume the activity and repeat the robot prompt. The device has an haptic motor, so the researcher is notified through vibrations when the system needs some

input. Therefore, the researcher can be focused with the child during the activity.

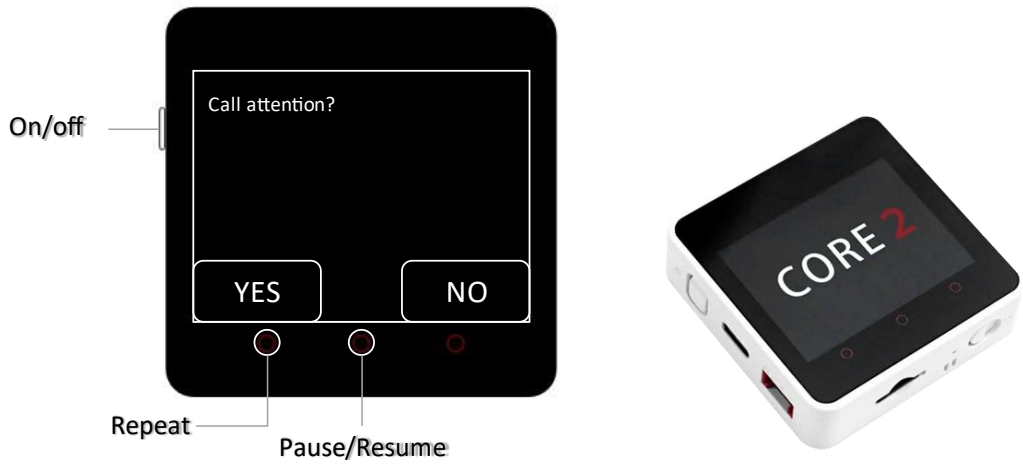


Figure 5.28: Remote device used by the researcher/experimenter to confirm or not the system recommended actions as well as pause/resume the activity and repeat the robot prompt.

5.3.1 Summary and conclusions

The supervised behavioural framework has three main layers – Perception, Reasoning, and Actuation. The perception layer is responsible for sensing and analysis of the data received from the sensors. The interaction flow is defined in the reasoning layer. Finally, ZECA will display the different behaviours and the OPT will give different feedbacks accordingly to the output of the Reasoning layer. Concerning the reinforcement (positive or negative), they are displayed simultaneously by ZECA and the OPT and they can be configured according to the child preferences. Since it is a supervised approach, the system GUI will prompt the detected behaviours and depending on the interaction state, it will recommend possible actions that will be carried out automatically, unless the researcher/therapist intervenes. Additionally, the recommended actions are also sent to a remote device. Thus, the researcher/experimenter does not have the necessity to constantly look at the monitor to see the GUI.

5.4 Pilot Study

A small pilot study was conducted at a school with 2 children with ASD to evaluate the system limitations and determine if it facilitates natural and comfortable interaction during support sessions. The participants, aged 7 to 9 years (M = 8.0 ; SD = 1.0), were diagnosed with high-functioning ASD and had parental consent to participate. Each child participated in 4 sessions, using the triadic setup (child, robot,

researcher) outlined in section 3.2. None of the children had prior experience with the system (robot and OPT).

The game scenarios used in this pilot study were the Recognize and Storytelling activities that are described in section 3.2.3. The distribution of the game scenarios over the sessions is the following:

- Session 1: Recognize;
- Session 2: Recognize + Storytelling;
- Session 3: Recognize + Storytelling;
- Session 4: Storytelling;

The videos produced along the sessions were analysed using the Observer XT 11 specialized software by Noldus (*Noldus | Innovative Solutions for Behavioral Research*, 2015) in order to quantify predetermined behaviours performed by children. These predetermined behaviours concern the typical and stereotypical actions that the behavioural model (section 5.1.1) was trained on, and are detailed bellow:

- IDLE: the child is in an idle position normally interacting with the robot and/or the OPT;
- Standing: the child gets from a seated position to a standing position;
- Hand/s Raise: one or both hands are raised by the child;
- Clapping Hands: the child joins both hands producing a sound;
- Pointing: the child points at something to attract the attention of the experimenter;
- Cover ears: the child cover one or both ears;
- Rocking: the child rocks his/her body back and forward;
- Hand wave: the child waves/shakes one or both hands;
- Turn: the child turns his/her body to one side or the other.

The main goal of this pilot study was to verify the implemented framework dynamics in a school setting with the target group and detect the system constraints.

5.4.1 Results

The results focus on game performance and behavioural analysis, in particular the detection of the behaviours that the model was trained on. Data concerning system adaptation prompts was also recorded. It is worth to point out that one of the goals of the pilot study was to test the complete developed system with the framework introduced in the present chapter in a school setting. All participants answered the robot prompts. The percentage of successful, unsuccessful, and unanswered prompts performed by the children along the sessions were considered, not to evaluate the children's knowledge but to try to perceive if they understood the procedure of the games. The number of right, wrong, and not answered prompts are shown in Figure 5.29 for each child. All participants answered the robot prompts and in general there was a positive trend, regarding the number of successful answers along the sessions in both game scenarios.

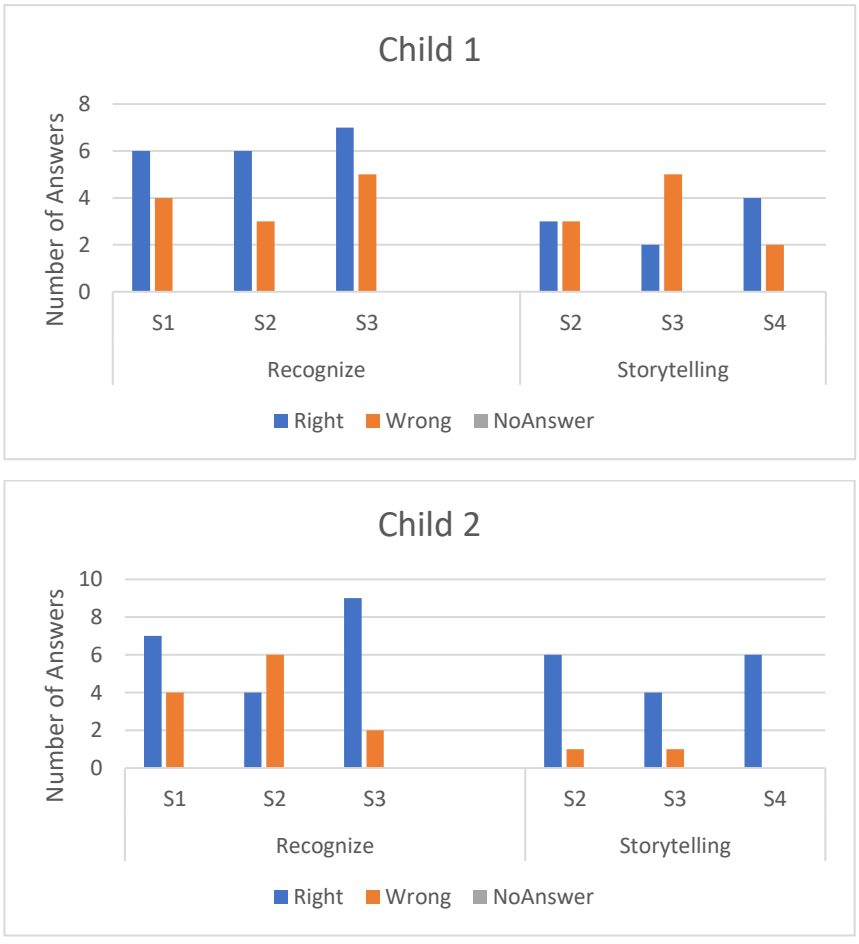


Figure 5.29: Children's performance in terms of right, wrong, and no answers along the sessions for both game scenarios – child 1 at the top; child 2 at the bottom.

Regarding the detection of the children's behaviours (typical and stereotypical) over the sessions, data analysed from the videos were compared to the data recorded by the GUI (section 5.3). Since the system uses the Nuitrack middleware software (section 5.1), the detection of behaviour is only run if the Nuitrack detects a person in the scene. Table 5.7 shows the sessions and the game scenarios where the system detected the children's behaviours. The detected behaviours (M ± SD) by the system for child 1 were the following: IDLE (199.9 ± 51.3), CLAP (4.93 ± 1.7), TURN (6.0 ± 6.2), ROCKING (3.0 ± 1.6), HAND_WAVE (2.0 ± 1.1), POINTING (3.0 ± 1.2), and COVER_EAR (2.5 ± 0.8). During the analysis of the videos, the same behaviours were coded with the exception of the CLAP behaviour as it was not exhibited by the child. Concerning the child 2, the detected behaviours were the following: IDLE (235.7 ± 78.4), CLAP (4.9 ± 3.2), TURN (1.8 ± 0.2), ROCKING (3.7 ± 3.1), POINTING (2.4 ± 0.1), and COVER_EAR (2.5 ± 0.8). During the analysis of the videos, the same behaviours were coded with the exception of the CLAP and ROCKING behaviours as they were not exhibited by the child.

Table 5.8 shows the frequency of the analysed behaviour with Noldus versus the detected behaviours by the model per session and child. Figure 5.30 and Figure 5.31 compares the average duration of behaviours detected by the system to the video analysis done with Noldus for Child 1 and Child 2, respectively.

Table 5.7: Sessions with the performed game scenarios where the system detected the children's behaviours (marked as ✓). Session 1 and 4 only the Recognize (R) and Storytelling (S) game scenarios were performed, respectively.

Child \ Sessions	1(R)	2(R)	2(S)	3(R)	3(S)	4(S)
1	✓	✓	✓	✓	✓	✓
2	✓		✓		✓	✓

It is possible to observe that for most behaviours the detection duration time is congruent to the duration of the behaviour analysed via Noldus. Moreover, child 1 exhibited more different behaviours than child 2, as seen by the video analysis and detected behaviours by the trained model, Figure 5.30, Figure 5.31, and Table 5.8.

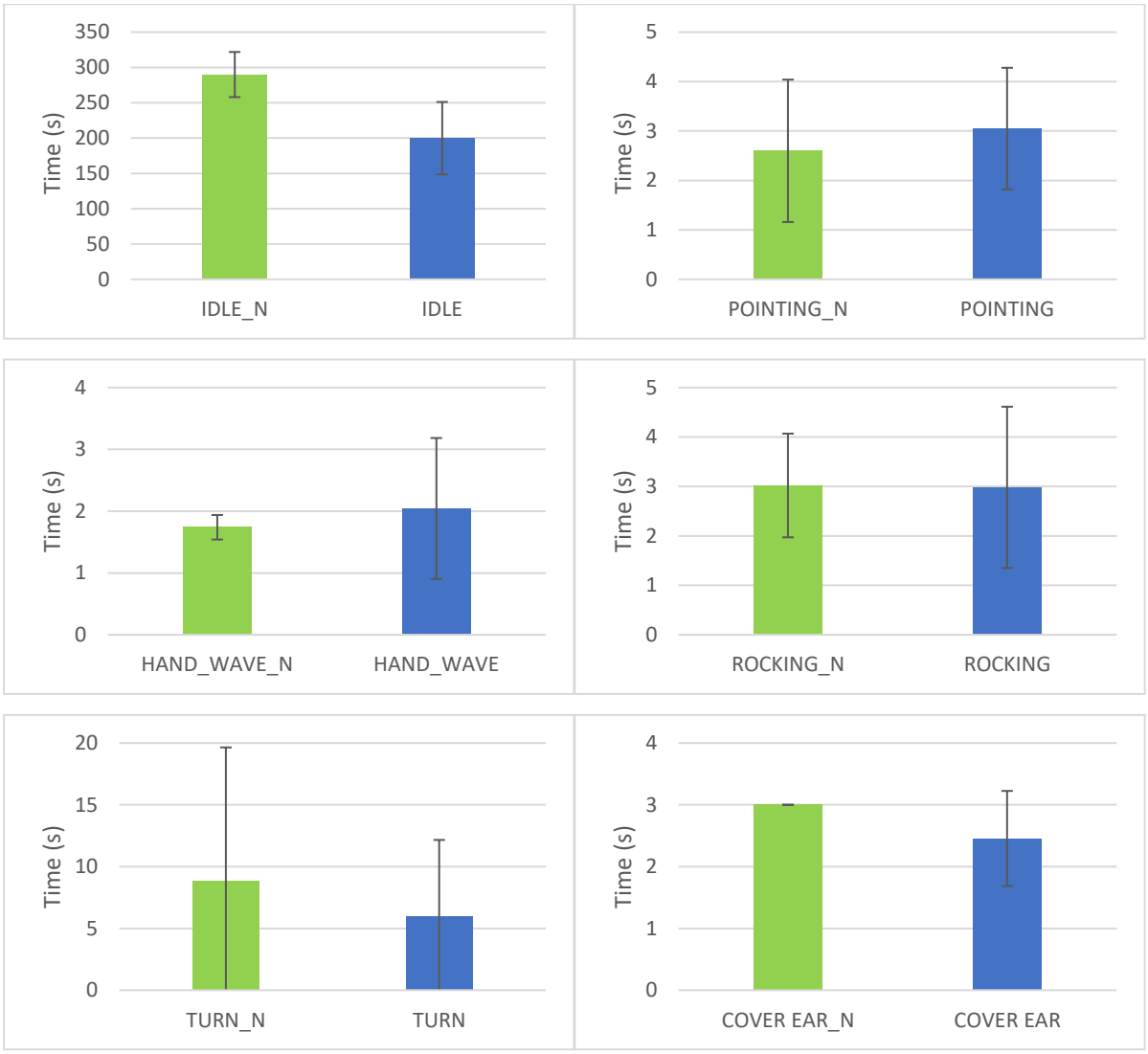


Figure 5.30: Comparison of the average duration of the behaviours obtained from video analysis (in ■ and _N) with the average duration of the detected behaviour by the trained model (in ■) for child 1.

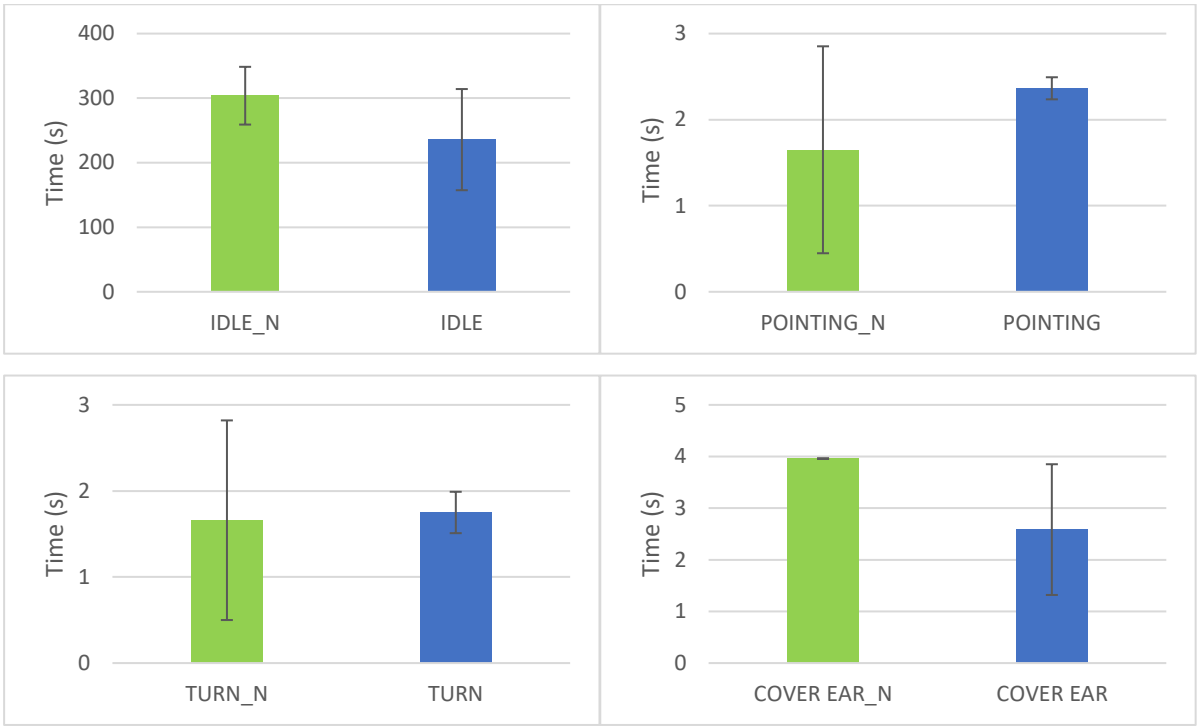


Figure 5.31: Comparison of the average duration of the behaviours obtained from video analysis (in ■ and _N) with the average duration of the detected behaviour by the trained model (in ■) for child 2.

Table 5.8: Frequency of the analysed behaviour with Noldus (■) vs. the detected behaviours (■) by the model per session and child.

		Child 1						Child 2					
Behaviour	N/M	1	2R	2S	3R	3S	4	1	2R	2S	3R	3S	4
Hand_Wave	■				1		1						
	■				1		2						
Rocking	■		2		2	2				—	—		
	■		2		3	3				3	1		
Turn	■	1			1			1			1		
	■	1	1		2			1			1		
Cover_Ear	■	—				1					1		—
	■	1				1					1		1
Pointing	■			1	1	1					2		1
	■	1		1	1	1					3		2

*N – Noldus; M – the trained model; R – Recognize; S – Storytelling;

5.4.2 Discussion

In general, the typical behaviours that were detected did not triggered any action by the system due to their short duration with exception of the TURN behaviour for child 1. Since the child remained interactive with the robot and OPT, the action suggested by the system was not executed. Regarding the stereotypical behaviours detected, none resulted in actions by the system since as recommended by the participants in the interviews conducted in section 5.2.1. Concerning other system prompts, namely adding more time and call the child attention to interact with the OPT, they were triggered but none were commanded to be executed by the experimenter as the child consequently interacted with the OPT and answered the robot prompt within the time frame of the answer.

There were moments of out of focus with both children along the sessions. However, in these moments the experimenter intervened by tapping the child on the arm, following one of the suggestions of the participants from the interview. This was enough to quickly bring the child attention back to the activity.

Concerning the CLAP behaviour, it was wrongly detected for both children. In case of child 1, the clap behaviour was detected most times by the model when the child was closely flicking his/her fingers mid-air as this behaviour was noted during the video analysis. For child 2, it was noted during the video analysis, that the CLAP behaviour was detect when the child moved the OPT closer to his/her face, raising both hands, which could mean that the trained model is only taking into account the positions and distance of the hands in relation to the body and not considering the slight movement of the hands.

Regarding the frequency of the detected behaviours they were in general in par with the analysed behaviours via Noldus. The model, in some cases, tended to over detect some behaviours. Conversely, the behaviours average detection duration is, in general, consistent with the duration of the behaviour analysed through Noldus.

The Nuitrack middleware, with its real-time capabilities, is utilized as the initial step in the pipeline for person detection and skeleton data extraction. However, the system sometimes failed to detect the person in the scene, preventing further behaviour detection. Additionally, even in the sessions where the system detected the children's behaviour, some of the times the system lost track of the person. This may have been caused by lighting changes, the user's movement within the scene, or occlusion. In spite of this, in the sessions where the system detected the children's behaviour, it successfully detected the majority of the behaviours that the children exhibited.

Concerning the children's performance, it is possible to observe a positive trend in both game scenarios, indicating that the children understood the robot prompts and successfully interacted with OPT. They

were also attentive to the OPT feedback, lights, haptic as well as the images for correct and incorrect answers displayed on the screen.

This pilot study allowed to perceive the system dynamics in a school setting with children with ASD and further support the introduction of autonomy mechanisms in a supervised manner.

References

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477553

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

CDC. (202 C.E.). *Signs and Symptoms of Autism Spectrum Disorders | CDC*. https://www.cdc.gov/ncbddd/autism/signs.html

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H. L., Coeckelbergh, M., Costescu, C., David, D., de Beir, A., Fang, Y., Ju, Z., Kennedy, J., Liu, H., Mazel, A., Pandey, A., Richardson, K., Senft, E., Thill, S., … Ziemke, T. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn*, *8*(1), 18–38. https://doi.org/10.1515/pjbr-2017-0002

Gaglio, S., Re, G. lo, & Morana, M. (2015). Human Activity Recognition Process Using 3-D Posture Data. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2014.2377111

Goncalves, N., Rodrigues, J. L., Costa, S., & Soares, F. (2012). Automatic detection of stereotyped hand flapping movements: Two different approaches. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2012.6343784

Hidalgo, G., Cao, Z., Simon, T., Wei, S.-E., Raaj, Y., Joo, H., & Sheikh, Y. (2022). *CMU-Perceptual-Computing-Lab/openpose - System Requirements*. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/installation/0_index.md#operating-systems-requirements-and-dependencies

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*.

Jazouli, M., Majda, A., Merad, D., Aalouane, R., & Zarghili, A. (2019). Automatic detection of stereotyped movements in autistic children using the Kinect sensor. *International Journal of Biomedical Engineering and Technology*. https://doi.org/10.1504/IJBET.2019.097621

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

Ludl, D., Gulde, T., & Curio, C. (2019). Simple yet efficient real-time pose-based action recognition. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*. https://doi.org/10.1109/ITSC.2019.8917128

Nuitrack. (2019). *Nuitrack Full Body Skeletal Tracking Software - Kinect replacement for Android, Windows, Linux, iOS, Intel RealSense, Orbbec*. https://nuitrack.com/

Nuitrack Face Tracking. (2019). *nuitrack-sdk/Unity_Face_Tracking.md at master · 3DiVi/nuitrack-sdk · GitHub*. https://github.com/3DiVi/nuitrack-sdk/blob/master/doc/Unity_Face_Tracking.md

Nuitrack Frame Rate. (2018). *Skeleton tracking specs ? 60 - 90 fps? - Nuitrack SDK - Nuitrack*. https://community.nuitrack.com/t/skeleton-tracking-specs-60-90-fps/916

OOM with CUDNN installed and GPU requirements met · Issue #1929. (2021). https://github.com/CMU-Perceptual-Computing-Lab/openpose/issues/1929

OpenCV. (2020). *OpenCV: Deep Neural Networks (dnn module)*. https://docs.opencv.org/master/d2/d58/tutorial_table_of_content_dnn.html

OpenFace CPU. (2016). *High CPU Usage · Issue #82 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/82

OpenFace Pupil and Gaze Estimations. (2021). *Inaccurate pupil (and gaze) estimations · Issue #958 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/958

OpenFace Pupil Estimation. (2018). *Inaccurate Iris/Pupil Localization · Issue #359 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/359

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2018). Exploiting deep residual networks for human action recognition from skeletal data. *Computer Vision and Image Understanding*. https://doi.org/10.1016/j.cviu.2018.03.003

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism?

A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308

webQDA – Qualitative Data Analysis Software. (2022). https://www.webqda.net/

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 2921–2929. https://doi.org/10.1109/CVPR.2016.319

6 EXPLORING THE USE OF A SUPERVISED ARCHITECTURE FOR PROMOTION OF SOCIO-EMOTIONAL SKILLS IN CHILDREN WITH ASD

Summary

This chapter presents a study of the behaviour and performance of children with ASD when developing socio-emotional skills in a hybrid supervised HRI triadic scenario. The study was based on the research previously presented.

6 Exploring the use of a supervised architecture for promotion of socio-emotional skills in children with ASD

6.1 Research Questions

6.2 Methods

6.3 Results

6.4 Discussion of the results

6.5 Summary of hypothesis and conclusions

Using the research outlined in prior chapters, the present chapter centres on the study of a supervised hybrid approach for developing emotional skills and promoting social interaction in children with ASD. Children with ASD can find it complex to understand the other's mental state (Baron-Cohen et al., 1985), in comparison to typically developing children that acquire this skill at 4 years of age (Baron-Cohen, 1991). Authors in the literature have proposed technological tools as a way to teach this skill to children with ASD (Diehl et al., 2012; Lund, 2009; Lund et al., 2005; Noor et al., 2012; Pennisi et al., 2016; Saleh et al., 2021; Scassellati et al., 2012), mainly through the use of Objects based on Playware Technology (OPT) and assistive robots. Systems that can synthesize affects, for example virtual characters (avatars), and robotic platforms can be applied in autism intervention (S. Costa, 2014; Pennisi et al., 2016). The majority of the systems proposed in the literature are controlled using the Wizard-of-Oz (WOZ) setup, meaning that in reality the robot does not adapt its behaviour to the children's actions (Pennisi et al., 2016). Thus, there is a need to introduce some autonomy, enabling the robot to perform some autonomous behaviours, adapting its behaviour to the children's actions. Additionally, none of the referred systems explore the use of a hybrid approach combining assistive robots and an OPT. Therefore, the study presented in this chapter aims to evaluate the use of the proposed supervised hybrid approach, already described in chapter 4, as a tool to mediate triadic interactions and to teach recognition and labelling emotions. It focused on verbal and non-verbal communications as well as performance measures to evaluate social communication in a triadic relationship.

This goal achievement was measured using observational data from video analysis regarding pre-established behaviours, data recorded from the robot, and quantitative ratings from a comparison between a pre-test and a post-test.

The procedures presented in this chapter were approved by the Ethics Committee of the University of Minho, Portugal and the Portuguese National Committee for Data Protection. The official documents can be found in the Appendix A.7. All the interveners (schools and clinics) who participated in the study agreed on a protocol with the University of Minho. Parents of the children signed an informed consent in through which they became aware of the goals and the methods applied in the research (Appendix A.4). The children's teachers/clinicians were informed and consulted about the game scenarios to be performed and gave suggestions intended to improve them.

6.1 Research Questions

In the present study, the following research questions were considered:

RQ1: What are the differences in the children's behaviour when OPT is added to a HRI support session?

RQ2: What are the differences in the children's behaviour, or performance, when the robot modifies its behaviour according to the children's non-verbal actions or performance?

RQ3: In what extent does the inclusion of social awareness behaviours and adaptation in a hybrid approach based on a robotic and OPT system benefit social interaction of children with another person (experimenter)?

To address these questions, data concerning verbal and non-verbal communications as well as the children's performance in the game scenarios were recorded and analysed.

Regarding performance, comparisons were made considering the number of successful, unsuccessful, and unanswered prompts between each activity. Moreover, it was intended to verify an improvement when comparing the first to the last support session in terms of number of successful answers.

Concerning verbal and non-verbal behavioural data, comparisons were made within each game scenario, comparing for example the behaviour (e.g. eye gaze) in the first vs. last session, as well as, comparing the behaviours between game scenarios. The eye gaze towards the experimenter was also quantified in both game scenarios and along the sessions. Additionally, the joint attention time (JAT) was measured by the eye gaze time at the experimenter, together with the time the children looked ate the OPT and the robot along the sessions. JAT is defined as the capability children have to share attention with others about an object or an event, gazing alternatively at the object and at the peer (Johnson et al., 2007).

Finally, data from a common task (described in detail in section 6.2.2) performed by the children in a pre- and post-test were analysed in terms of the number of correct matches. In this activity, each photo showing portraying an emotion was randomly shown to the child. Then, the child has to choose the correct match from the PECS cards.

Concerning RQ1, data collected from video analysis namely the duration of the children's gaze towards the robot and OPT were compared in both game scenarios along the sessions. Regarding RQ2, data from video analysis namely, eye gaze and joint attention time, were compared in both game scenarios along the sessions. Comparisons were also made for the same behaviours between game scenarios. Similar comparisons were made considering the children performance data in the game scenarios. Finally, about RQ3, comparisons were conducted concerning joint attention time (eye gaze towards experimenter),

verbal behaviours data as well as performance data in both game scenarios.

The following hypotheses are proposed:

H1: Children will be more engaged in the HRI with OPT.

H2: Children will be more engaged and with better performance in the HRI when the robot adapts its behaviour to the children's non-verbal actions or performance.

H3: Children will stay engaged with another person for longer when the robot predicts the children's actions based on the children's eye gaze/head movement, non-verbal and verbal actions, and performance in the activity.

6.2 Methods

The methods used to answer the research questions were presented in Chapters 3 and 4, specifically the ethical issues (section 3.3), the OPT – PlayBrick (section 4.3), the robot (section 3.2.2), the software architecture (chapter 5), and the room setup (section 3.2.4).

The videos produced along the sessions were analysed using the Observer XT 11 specialized software by Noldus (*Noldus | Innovative Solutions for Behavioral Research*, 2015) in order to quantify predetermined behaviours performed by children. The selection of the behaviours was done according to its relevance (regarding engagement and interactive behaviour), and feasibility to identify them in the recorded data. Regarding non-verbal communication, the literature was consulted to choose the expected behaviours to be observed in children with ASD (Mundy et al., 1986; Soares et al., 2019). Therefore, Table 6.1 shows the list of the coded behaviours. In this list, state events stand for behaviours that take a period of time and therefore have a duration.

Table 6.1: Selected behaviours coded/used during the video analysis procedure. Verbal and non-verbal behaviours as well as typical and stereotypical were selected.

Coded Behaviour	List of behaviours and its significance
Eye Gazing Behaviour (State event)	– Robot: head orientation of the child looking towards the robot; – Experimenter: head orientation of the child looking towards the experimenter, except experimenter's eyes; – OPT (PlayBrick): head orientation of the child looking towards the OPT device; – Eye contact: head orientation of the child looking towards the experimenter, specifically his eyes.
Verbal Communication (Sate Event)	– Echolalia: words or sentences repeated after the experimenter or the robot; – Vocalisations: oral sounds made by the child without meaning or context; – Speech: words and sentences said by the child.
Non – Verbal behaviours (State Events): typical and stereotypical behaviours are considered.	– IDLE: the child is in an idle position normally interacting with the robot and/or the OPT; – Standing: the child gets from a seated position to a standing position; – Hand/s Raise: one or both hands are raised by the child; – Clapping Hands: the child joins both hands producing a sound; – Pointing: the child points at something to attract the attention of the experimenter; – Cover ears: the child cover one or both ears; – Rocking: the child rocks his/her body back and forward; – Hand wave: the child waves/shakes one or both hands; – Turn: the child turns his/her body to one side or the other.

In order to ensure inter-reliability, an independent coder was trained to code the behaviours listed in Table 6.1 and the following rules were drafted:

- When looking at an interaction, the coder should first classify the function to then, second identify the particular behaviour code;
- If a behaviour is not well-defined, it may not be codable. It is better not to code a behaviour than categorize it without sufficient information;

- Do not code any behaviour that is obscured (e.g. by the experimenter blocking the camera's view of the child);
- When the child exhibits behaviours that were not specified in the list, they are not coded;
- A behaviour ends if the child stops exhibiting that behaviour or shows another behaviour, directly related (for example, looking at ZECA/ looking at the experimenter);
- For the Eye Gaze, turning away ends the behaviour. Turning back immediately and looking again counts as new behaviour;
- Considering the eye gaze towards the experimenter, if the child makes eye contact this behaviour is coded in parallel with the eye gaze towards the experimenter;
- An activity starts after the sentence "Press START to begin", and it ends after the robot says "See you soon".

The onset and offset times of behaviours and coding of events were used to record the behavioural sequences. In order to facilitate this, it is useful for these codes to be mutually exclusives, with no need to record the offset times. For the present study, the adequate form of data representation is time-event sequences (Bakeman & Gottman, 1997), consisting simply of codes for the events, ordered as they occurred being possible to determine, for instance, how often specific behavioural codes co-occurs or if a certain coded behaviour tends to follow or precede others in systematic ways.

6.2.1 Participants

A total of 6 children participated in the study. All children met the following inclusion criteria: aged 4 to 10 years old, diagnosed with high functioning ASD by a professional clinician, and with acceptance of parents to participate in the study. All children that participated in the study are verbal but their attempts to initiate interactions and make friends are difficult or typically unsuccessful (Level 1 of severity levels, (American Psychiatric Association, 2013b)).

The experimenter did not have direct access to the children's medical files, but he had access to information provided by teachers or therapists who guaranteed the children's diagnosis. The experimenter did not know any of the children prior to the experiments. All children receive weekly therapy from speech, occupational therapists and psychologists. During the time window of this experiment, it was requested to both teachers and therapists to not perform activities focusing on emotion recognition. The mean age (M) and corresponding standard deviation (SD) of the children are: 5.0 ± 0.6. The majority of the children in the sample is male, M = 83.3% (1 in 6 is female).

According to the children's therapists, the preferred types of reinforcement used in support sessions with the children participating in the study were (by percentage): Verbal = 83.3%; Movement = 0.0%; Verbal + Movement = 16.7%; Verbal + Movement + Sound = 0.0%. Additionally, in general, the therapists mentioned that the children would be comfortable with the types of reinforcements that the system can provide.

6.2.2 Procedures

A total of six sessions were performed with each child, being the first and second session performed in the same day. A familiarisation phase took place in the first session, where the experimenter had the opportunity to interact with the children and introduce the robot (ZECA) to the child. The sixth session was performed two weeks after the fifth session. The task performed in the pre- and post-tests consisted in matching 5 PECS cards representing the five basic emotions (anger, happiness, sadness, scared, and surprise) to 5 photographs of a man (pre-test) or a woman (post-test) showing the same facial expressions, Figure 6.1. The number of corrected matches was measured and if the task was completed. The task is performed without the support of the proposed system.

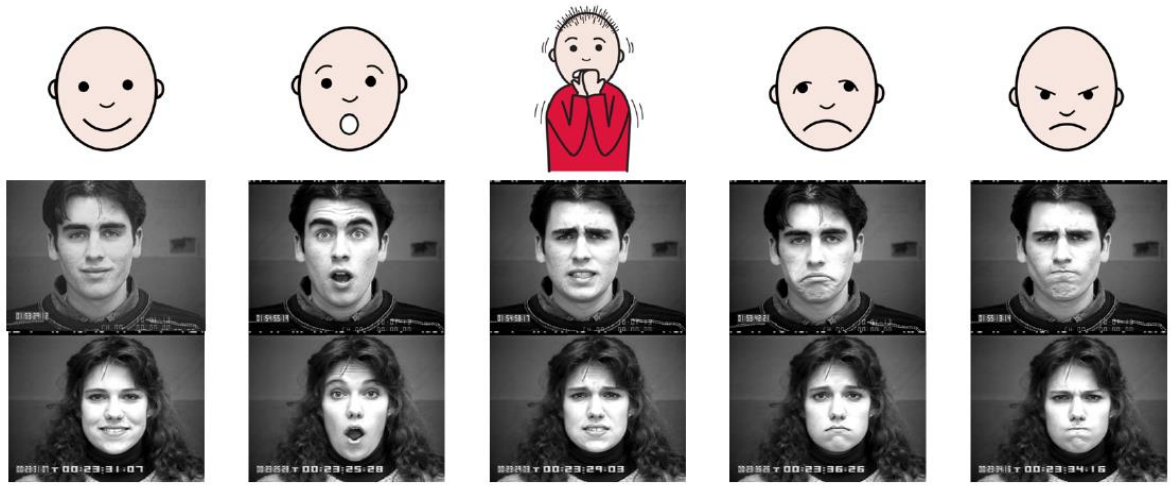


Figure 6.1: Images used in pre- and post-tests sessions. At the top it shows the 5 PECS cards which were matched to the figures with the man and the woman, respectively. From left to right, the portrayed emotions are: happiness, surprise, fear, sadness, and anger.

The game scenarios used in this study were the ones tested along the chapters 3 and 4, described in detail in section 3.2.3 – Recognize (to identify and label facial expressions matching emotions) and Storytelling (to evaluate the affective state of a character at the end of a story). Both have in common the development of emotion recognition skills.

The distribution of the game scenarios over the sessions follows the insights from previous research (Soares et al., 2019). It takes in consideration the difficulty level, being the Recognize activity the basic task and so it is the first to be conducted with the children. The Storytelling activity is considered harder for the children, since it consists in the identification of character's affective state in a story (Soares et al., 2019). The game scenarios were presented in equal number of times – 3 for the Recognize and 3 for Storytelling game scenarios. Each session with the children took between 5 to 10 minutes, according to the number of game scenarios that the children had to perform.

The experimenter interacted with the children at least once per week, during five weeks. When any children had to miss his/her session, it was re-scheduled. The children missing more than one session were excluded from the experiments. The game scenarios were performed as presented in the following list:

- Session 1: Pre-test;
- Session 2: Recognize;
- Session 3: Recognize + Storytelling;
- Session 4: Recognize + Storytelling;
- Session 5: Storytelling;
- Session 6: Post-test.

To ensure inter-rater reliability, 15% of the videos were re-coded by a second independent coder, resulting in a Cohen's kappa k = 0.65. This is acceptable, as having a Cohen's kappa value higher than 0.60 suggests a good agreement between the ratters (Bakeman & Gottman, 1997).

6.3 Results

The results focus on behavioural analysis (verbal and non-verbal), game performance, and pre- and post-test. Due to sample size (n = 6) and since the obtained data do not follow a normal distribution, non-parametric tests were used to statistically analyse the acquired information. Regarding behavioural analysis, the videos were coded following the method described in section 6.2 and the indicators listed in Table 6.1. Due to technical difficulties for session 4, it was analysed one less video in comparison to the other sessions. Thus, the comparisons made that includes data from session 4 regarding the video analysis were made with one less observation. The comparison of sessions represents dependent data since they were performed by the same children (e.g.: first session vs. last session) using Wilcoxon test to infer on the mean ranks of the two related populations. This test is reported by using the Z statistic,

after Normal approximation.

The data was collected, processed and the average values were presented. To simplify the text, further references to the 'mean' term in the variable name was omitted and only the identifier of the analysed variable was retained (e.g.: instead of *mean joint attention time*, only the variable is mentioned as *joint attention time*).

6.3.1 Analysis of Eye Gaze

This section presents the results concerning the analysis of the eye gaze, meaning where the children focused their attention during the support sessions. The eye gaze behaviour was used as a measure to calculate the joint attention time, which has a direct relationship with engagement.

Figure 6.2 shows the eye gaze behaviour performed by children along all the sessions. It is possible to observe that over all the sessions, the children focused more on the Robot + Experimenter + OPT than elsewhere as, in average, their attention level was above 80%. Comparing the first (Session 2), where only the Recognize game scenario was performed, with the last sessions (Session 5) where only the Storytelling game scenario was conducted, it was verified that the children gaze towards the robot was higher in Session 2 (38%) in comparison to 21.8% in Session 5. Conversely, by doing the same comparison in terms of the eye gaze now towards the OPT, it is possible to observe that it was higher in the last session 61.8% (Storytelling game scenario in comparison to the first 40.5% (Recognize scenario), an increase of approximately 20%. Additionally, the eye gaze towards the robot decreased along the sessions in contrast to the eye gaze towards the OPT, which increased along the sessions. Concerning the eye gaze towards the experimenter this maintained relatively the same along the sessions, on average 3.5%.

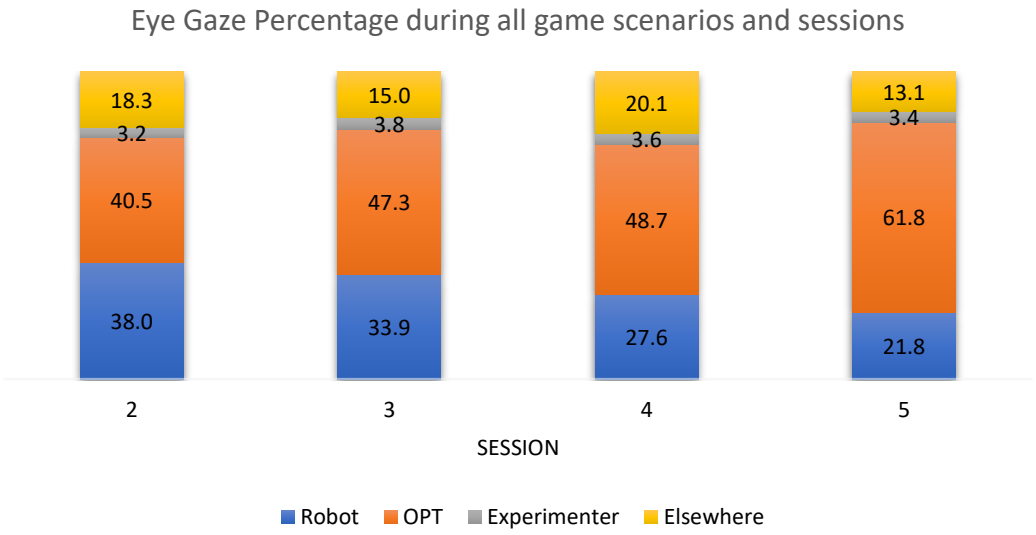


Figure 6.2: Average percentage of eye gaze time along the sessions. The percentage time of children gaze towards the OPT increased along the sessions. Overall, children stayed focused on the task since the percentage of gaze to elsewhere is small in comparison to the other percentages of gaze and it decreased over the sessions (with exception of session 4).

Using the data presented in Figure 6.2 and particularising it for each game scenario, the joint attention time was calculated. The joint attention time is the sum of the time children looked at the robot, plus the OPT, and the experimenter.

A Wilcoxon test was used to compare the percentage of joint attention time (considering the total time of the session) in the first and last session of each game scenario, Figure 6.3. Regarding both game scenarios, no significant differences were found when comparing the percentage of joint attention time – Recognize ($p = 0.34$) and Storytelling ($p = 0.66$). In the Recognize game scenario, the joint attention time decrease from 81.7 to 74.6%. Regarding the Storytelling game scenario the joint attention time remained relatively stable – 87.2 (first session) and 86.9% (last session). Conversely, by comparing the duration of joint attention time between the Recognize and Storytelling game scenarios in sessions 3 and 4 (were both games were conducted), it was found significant difference, $p < 0.01$. By observing Figure 6.3, it is possible to see that the percentage of joint attention time stayed relatively high, having a slight drop in the last session in the Recognize game scenario.

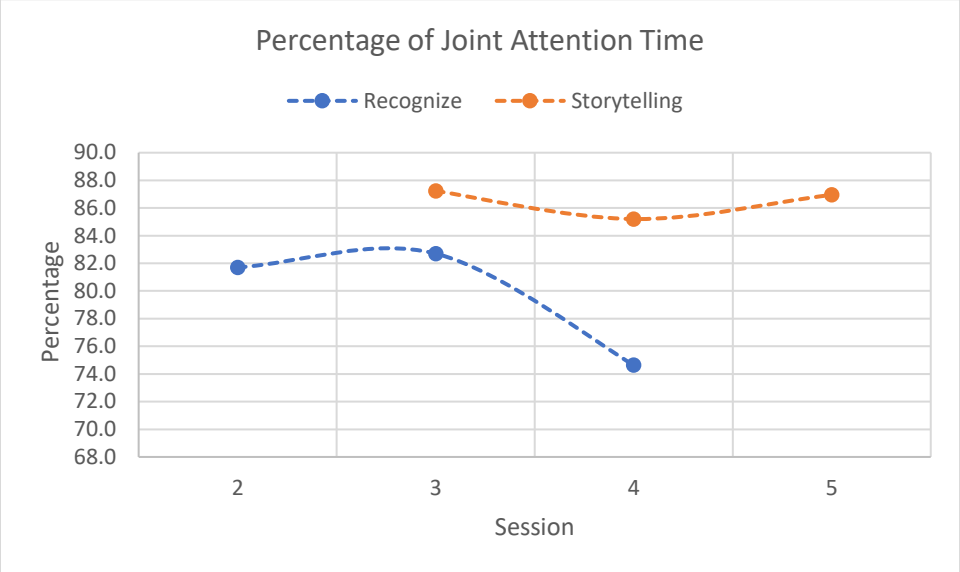


Figure 6.3: Mean percentage of the joint attention time along the sessions for both game scenarios. The joint attention time in the Storytelling stayed relatively higher in comparison to the Recognize game scenario.

Figure 6.4 and Figure 6.5, show the time in seconds of eye gaze towards the robot in comparison to the OPT for the Recognize (Figure 6.4) and Storytelling game scenarios (Figure 6.5). A Wilcoxon test was performed to compare the eye gaze time, in seconds, towards the robot versus the OPT in each game scenario. Regarding the Recognize game scenario, no significant differences were found, $p = 0.15$. However, regarding the Storytelling game scenario, significant difference was found when doing the same comparison, $p < 0.01$. Indeed, it is possible to see this difference in time when observing the Figure 6.5 for the Storytelling game scenario.

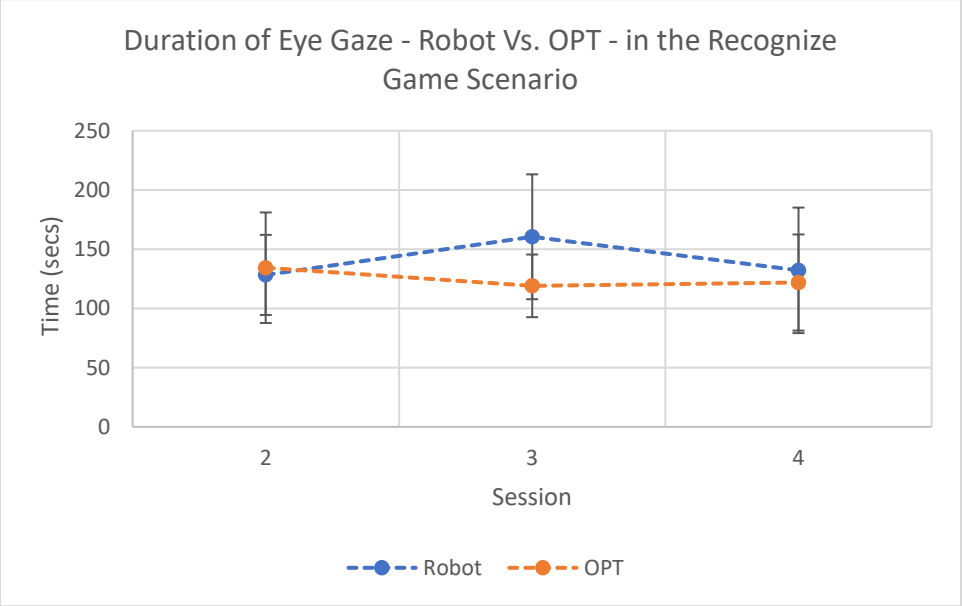


Figure 6.4: Comparison of the average duration of eye gaze towards the robot and OPT along the sessions in the Recognize game scenario.

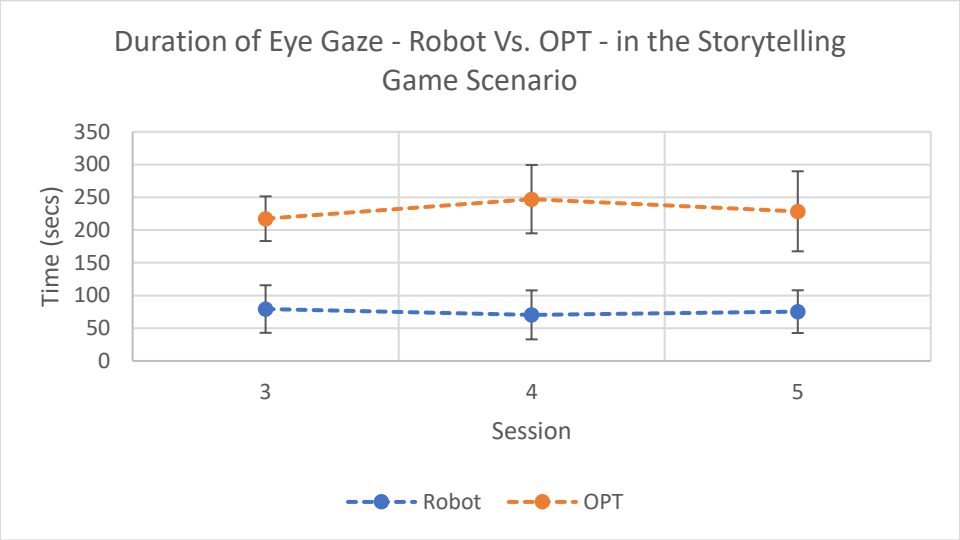


Figure 6.5: Comparison of the average duration of eye gaze towards the robot and OPT along the sessions in the Storytelling game scenario.

Figure 6.6 shows the percentage of the children eye gaze towards the experimenter along the sessions for both game scenarios. The percentage of eye gaze towards the experimenter is similar along the sessions in both game scenarios, being slightly higher in the Storytelling game scenario. A Wilcoxon test was used to compare the percentage of the children eye gaze towards the experimenter in the first and last session of each game scenario and no significant differences were found for both game scenarios, $p > 0.05$. Similarly, a comparison was made concerning the children eye gaze towards the experimenter between the Recognize and Storytelling game scenarios for the sessions 3 and 4 (were both games were

conducted), and no significant difference was found, p = 0.08.

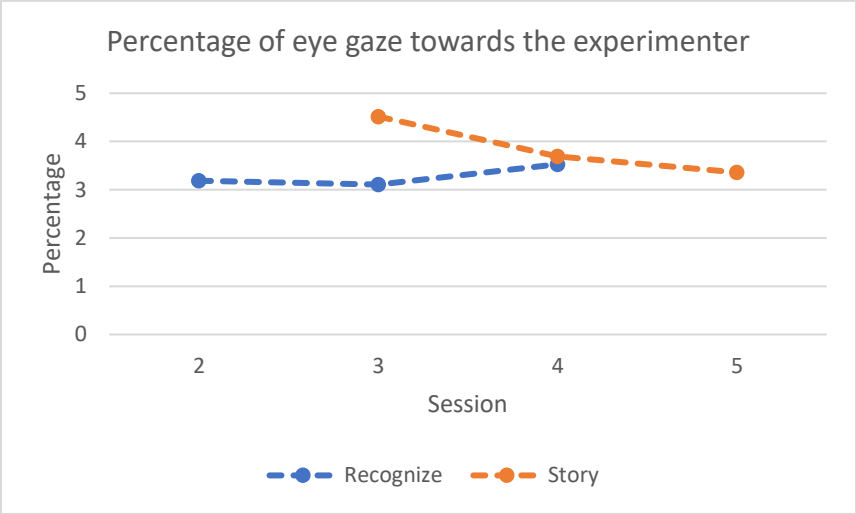


Figure 6.6: Mean percentage of eye gaze towards the experimenter along the sessions in both game scenarios. it stayed relatively constant in both game scenarios.

6.3.2 Verbal communication

Concerning the comparison of the duration of verbal behaviour of the children along the sessions, three behaviours were considered, namely echolalia, vocalization, and speech. However, all children forms of verbalization were expressed as speech. When comparing the duration of verbal behaviour between the first and last session in each game scenario, no significant difference was found using the Wilcoxon test (p = 0.5 and p = 0.2 for the Recognize and Storytelling game scenarios, respectively). However, by comparing the duration of verbal behaviour between the Recognize and Storytelling game scenarios for the sessions 3 and 4 (where both games were conducted), it was found significant statistical difference, $p < 0.05$. Figure 6.7 shows the duration of the verbal behaviour in both game scenarios (Recognize and Storytelling) along the sessions. Indeed, it is possible to observe that the duration of verbal behaviour along the sessions in the Recognize game scenario overall remains similar. Conversely, the duration of the verbal behaviour decreases along the sessions in the Storytelling game scenario but it remains higher in comparison to the Recognize game scenario.

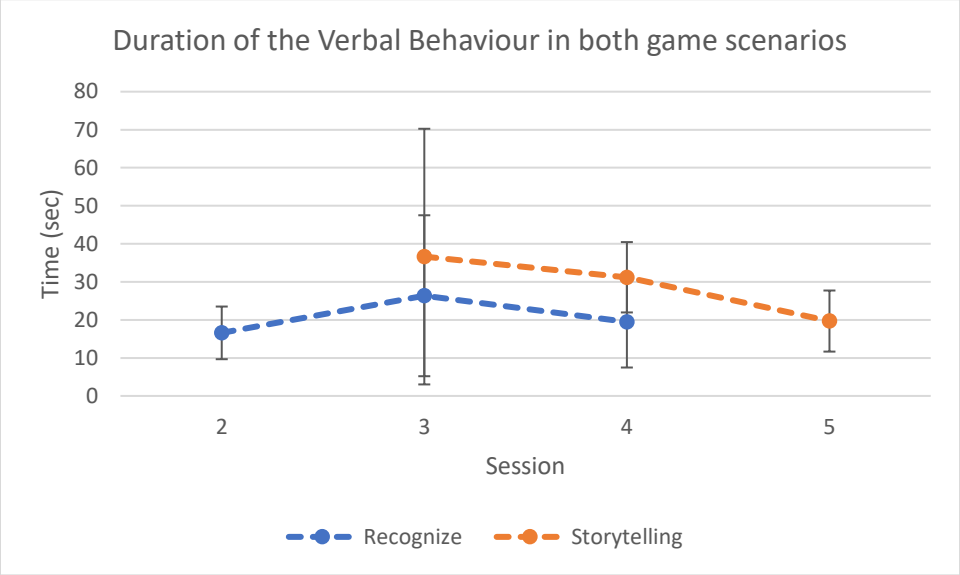


Figure 6.7: Mean duration of the Verbal Behaviour in both game scenarios along the sessions. The mean duration of verbal behaviour was higher in the Storytelling game scenario.

6.3.3 Performance in the game scenarios

This section addresses the performance in both game scenarios and the results concerning the system behaviour adaptation. The percentage of successful, unsuccessful, and unanswered prompts performed by the children along the sessions were considered, Figure 6.8 and Figure 6.9. Wilcoxon statistical tests were conducted in order to compare the percentage of successful answers in the first and last session in each game scenario (e.g. comparison of the performance in session 2 and 4 in the Recognize game scenario).

When comparing the first and last session in the Recognise game scenario, no significant differences were found ($p = 0.3$). The same was verified in the Storytelling game scenario, $p = 0.08$. On average, the performance in the Recognize game scenario increased 5.5% (from 73.6 to 79.1%, Figure 6.8) while in the Storytelling game scenario it increased by 11.1% (from 79.4% to 90.5%, Figure 6.9).

Besides the comparison of the first and last session in terms of successful answers (SA), it is important to verify if the number of SA overcame the children's number of unsuccessful and unanswered prompts (UUP) along the sessions in each game scenario. There was no significant difference when comparing the number of SA with the number of UUP in the first session of the Recognize game scenario ($p = 0.06$). However, significant differences were found in the last session of this game scenario comparing the number of SA with the number of UUP ($p < 0.05$). In the Storytelling game scenario, significant differences were found in the first ($p < 0.05$) and the last ($p < 0.05$) sessions.

By observing the Figure 6.10 it is possible to notice that there was a positive evolution, regarding the number of successful answers along the sessions, for both game scenarios.

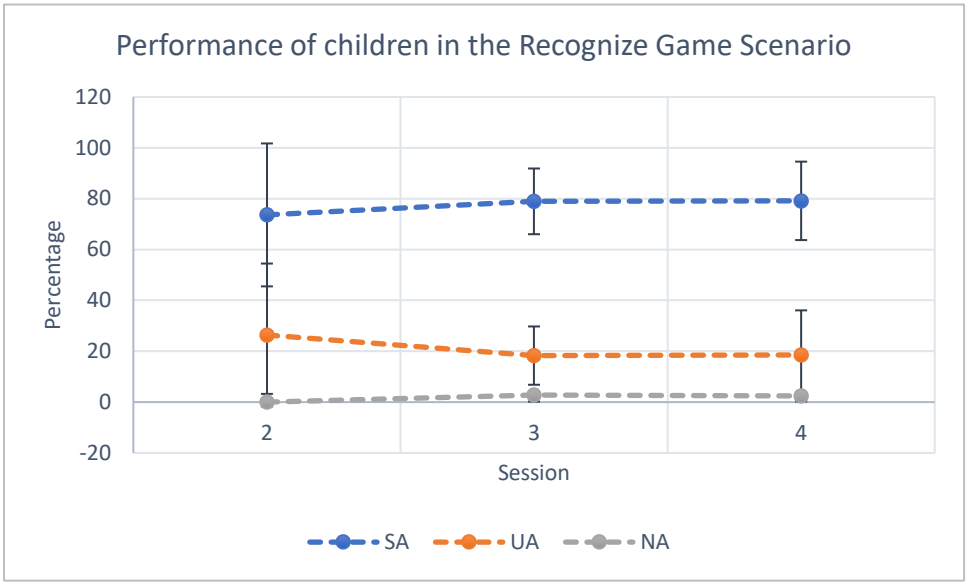


Figure 6.8: Children's performance in the Recognize game scenario along the sessions. The percentage of successful answers increased along the sessions.

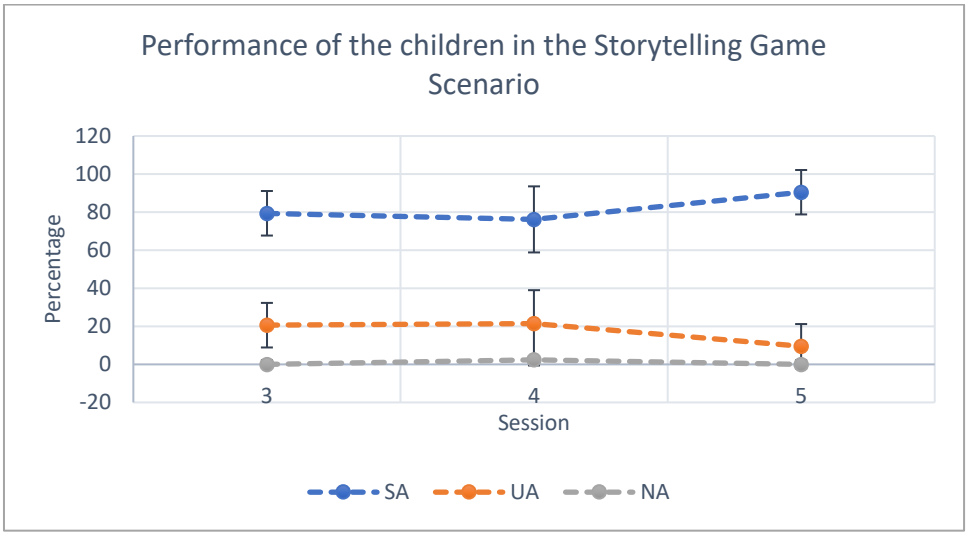


Figure 6.9: Children's performance in the Storytelling game scenario along the sessions. The percentage of successful answers increased along the sessions.

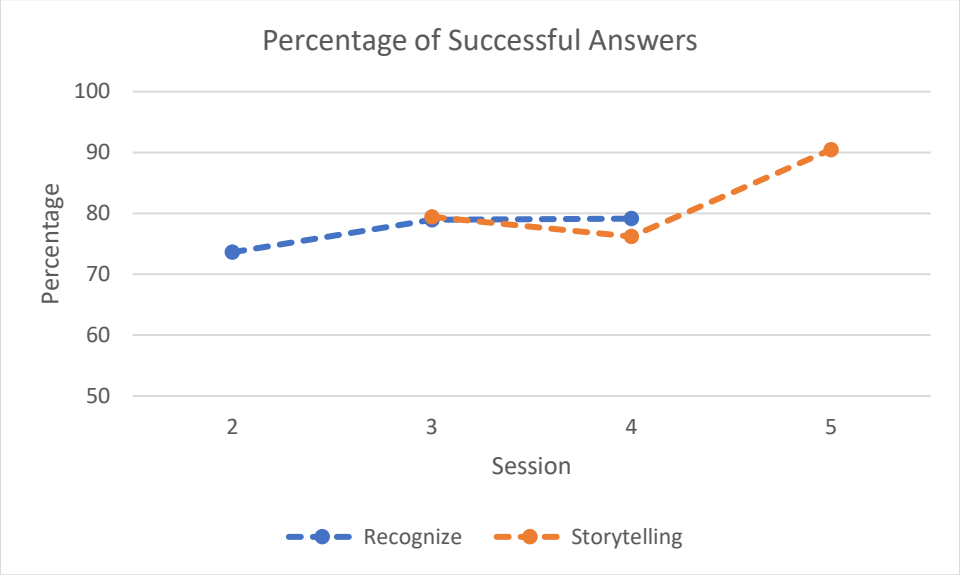


Figure 6.10: Percentage of successful answers along all the sessions, per game scenario. The percentage of successful answers increased in both game scenarios when comparing the first and last sessions.

Since the system consists of a supervised architecture that can adapt the system behaviour to the child's actions (performance + non-verbal behaviours), the number of system behaviour adaptation prompts was analysed along the sessions. Considering all the programmed behaviour adaptation (described in section 5.2), the most frequent ones that happened along the sessions were to add more time to answer the robot prompt and to call the child's attention to interact with the OPT. Table 6.2 shows the frequency of system prompts concerning behaviour adaptation along the sessions as well as the number of prompts that resulted in correct answers.

It is possible to observe that in the sessions 3 and 4 there was higher frequency of system prompts, in particular in session 3 and in the Storytelling game scenario, where this activity is first introduced to the children. Regarding the total number of system prompts (n = 10), 7 prompts resulted in the children correctly answer the robot prompts.

Figure 6.11, shows the frequency of system prompts concerning behaviour adaptation along the sessions for both game scenarios. It is possible to observe that the frequency of system prompts decreased along the sessions for both game scenarios. For the OPT call adaptation system prompt, data from the IMU embedded in the OPT device is used to detect the level of engagement of the children (as described in section 5.1), in particular, when the system is waiting for a response from the child.

Table 6.2: The number of system prompts concerning behaviour adaptation (AT = Add more time; OPT call = call the child's attention to interact with the OPT) along the sessions for both game scenarios (R = Recognize; S = Storytelling) as well as the number of prompts that resulted in correct answers.

Session	Scenarios	System Prompts	Frequency	Correct
2	R	OPT Call	-	-
		AT	2	1
3	R	OPT Call	-	-
		AT	1	0
	S	OPT Call	1	1
		AT	3	3
4	R	OPT Call	-	1
		AT	-	-
	S	OPT Call	-	-
		AT	2	1
5	S	OPT Call	-	
		AT	1	1

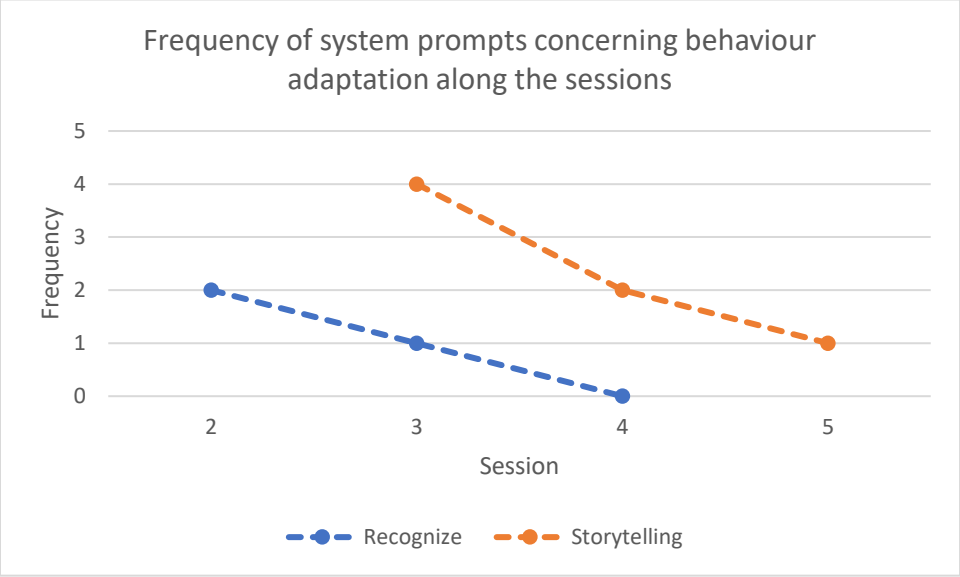


Figure 6.11: Frequency of system prompts concerning behaviour adaptation along the sessions for both game scenarios.

Concerning the detection of the children behaviours (typical and stereotypical) over the sessions, data analysed from the videos were compared to the data recorded by the Graphical User Interface (see section 5.3). Since the system uses the Nuitrack middleware software (section 5.1.1), the detection of behaviour is only run if the Nuitrack detects a person in the scene. Table 6.3 shows the sessions and the game

scenarios where the system detected the children behaviours. The detected behaviours (M ± SD) by the system were the following: IDLE (158.8 ± 91.4), CLAP (43.0±55.7), TURN (6.0 ± 3.2), ROCKING (4.2 ± 1.6), and HAND_RAISE (1.7 ± 0.0). During the analysis of the videos, the same behaviours were coded with the exception of the CLAP behaviour as it was not exhibited by the children. Additionally, the POINTING behaviour was coded during the video analysis but it was not detected by the system. Figure 6.12 shows the average time of the detected behaviours by the system in comparison to the analysis of the videos with Noldus.

Table 6.3 : Sessions with the performed game scenarios where the system detected the children behaviours (marked as ✓). Session 1 and 4 only the Recognize (R) and Storytelling (S) game scenarios were performed, respectively.

Child \ Sessions	1(R)	2(R)	2(S)	3(R)	3(S)	4(S)
1	✓	✓				
2				✓		
3				✓	✓	✓
4						
5						✓
6				✓	✓	✓

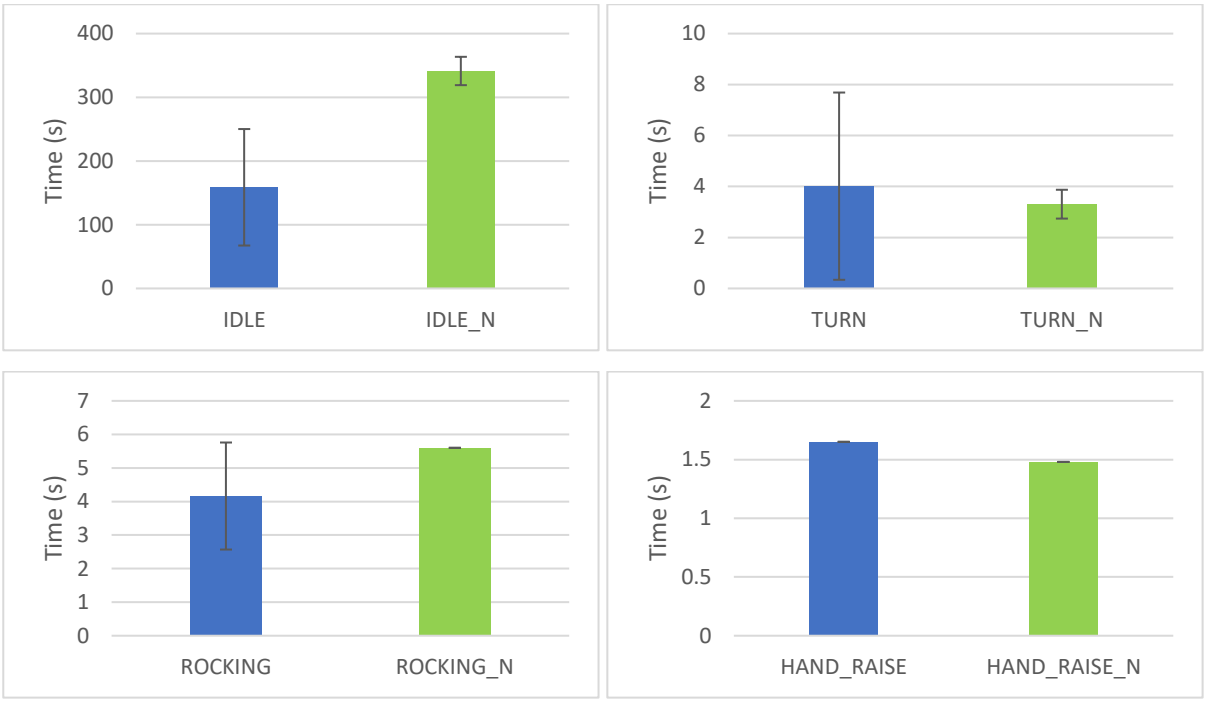


Figure 6.12: Comparison of the average duration of the detected behaviours (in ■) with the duration of the same behaviours obtained from video analysis (in ■ and _N), in seconds.

Concerning the analysis of the videos, the most common behaviour detected was IDLE.

The typical behaviours that were detected did not triggered any action by the system due to the their short duration. Regarding the stereotypical behaviours detected, none resulted in actions by the system since as recommended by the participants in the interviews conducted in section 5.2.2, the system and the experimenter should not intervene when stereotypical behaviours are exhibited by the children unless they are aggressive.

6.3.4 Pre- and post- test analysis

As mentioned in section 6.2.2, the children that participated in the present study did the performance task before the experiments and two weeks after the experimental procedure with the two game scenarios – Recognize and Storytelling. Figure 6.13 presents the average number of correct matches in the pre- and post-test. It is worth to point out that all children concluded the performance task. A Wilcoxon test was used to compare the corrected matches in the pre-test and post-test sessions and no significant statistical differences were found ($p = 0.05$).

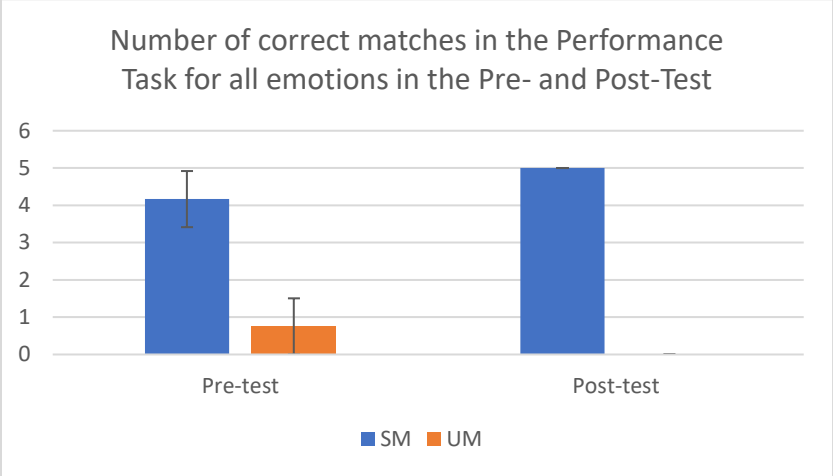


Figure 6.13: Number of successful (SM) and unsuccessful matches (UM) in the Performance Task for all emotions in the Pre- and Post-Test. There was no significant differences in the number of correct matches when comparing the pre- and post-test sessions.

6.4 Discussion of the results

By analysing the results in regards of eye gaze behaviour, it was verified that over all the sessions, the children focused more on the Robot + Experimenter + OPT than elsewhere as in average their attention level was above 80%. No significant differences were found when comparing the joint attention time in both game scenarios, Table 6.4. However, the joint attention time increased by approximately 5.2% from the first session to the last, Figure 6.2. Additionally, when comparing in each game scenario the duration of joint attention time in sessions 3 and 4 (where both games were conducted), it was found statistical significant difference with higher joint attention time during the Storytelling activity, meaning that the children stayed more focused during this game scenario. Moreover, when conducting a comparison in regards to the duration of the eye gaze of the child towards the robot versus the OPT over the sessions in both game scenarios, it was found significant differences in the Storytelling game scenario. Indeed, the children gazed more towards to the OPT in comparison to the robot in this game scenario. This outcome was expected as similar results were also found in the pilot study conducted in section 4.3.2 The same was not found in the Recognize game scenario.

An unexpected result concerns the decrease of the joint attention time in the Recognize game scenario, when comparing the first and last session. However, the same outcome was found in a previous study (Soares et al., 2019) where the same game scenario was performed with ZECA but without the OPT, Table 6.4. The percentage of joint attention time decreased drastically from 83.6 to 34.2%, a drop of 49.4%. In the present study the joint attention time decreased from 81.7 to 74.6% (in the Recognise game

scenario), a slight drop of 7.1% in comparison to the study conducted in (Soares et al., 2019). This may be due to the addition of the OPT, since the study done in (Soares et al., 2019) performed the same game scenarios but the child had access to cards representing the facial expressions and printed visual cards showing the story scenarios (in case of the Storytelling activity) instead of using an OPT device. Thus, the use of the OPT (PlayBrick) device may have influenced the interest of the children in the both game scenarios, since the eye gaze towards the OPT increased over the sessions and despite of a small drop in joint attention time in the Recognize game scenario (when comparing first vs. last session), the joint attention time remained stable and high, particularly in the Storytelling game scenario (86.9% in the last session), which could mean that the images of the story scenarios used as visual cues attracted the attention of the children.

Table 6.4: Statistical comparison of the children's joint attention time.

Behaviour	Work	Sessions	Scenario	p	Differences
Joint Attention Time	Present study	first. Vs last.	Recognize	.34	First: 81.7%
					Last: 74.6%
			Storytelling	.68	First: 87.2%
					Last: 86.9%
	(Soares et al., 2019)	first. Vs last.	Recognize	.001*	First: 83.6%
					Last: 34.2%
			Storytelling	.001*	First: 36.3%
					Last: 89.5%

*statistical significative difference

Concerning the children's verbal behaviour, no statistical significant differences were found when comparing the first and last session in each game scenario. However, in sessions 3 and 4 (where both games were conducted), it was found significant differences. Indeed, the verbalization behaviour duration was higher in the Storytelling game scenario, especially higher in sessions 3 and 4. Similarly, by analysing the number of system behaviour adaptation prompts, these were higher in sessions 3 and 4 and particularly high for the Storytelling game scenario. By adding more time or calling the child attention to interact with the OPT may have influenced the children verbalisation behaviour, overall and in particular for the sessions 3 and 4, since the duration of this behaviour and the number of system adaptation prompts are higher during these sessions where both game scenarios were performed.

Regarding the children's performance in the game scenarios, there were no significant differences when comparing the first and last session, Table 6.5. Nevertheless, there was an increase by 5.5% in the Recognize game scenario and by 11.1% in the Storytelling game scenario.

In general, the children kept improving along the sessions and statistical significant differences were found when comparing the number of successful answers versus unsuccessful in the last session of both game scenarios, Table 6.6.

Comparing with (Soares et al., 2019), similar results were found (Table 6.6). An interesting observation is that the number of successful answers in both game scenarios from the present work is higher in the first and last session in comparison to the work from (Soares et al., 2019), Table 6.5. This might be due to the use of the OPT device and the capability of the system to adapt its behaviour to the children's actions, since these are the additional add-ons to the triadic interaction when comparing to (Soares et al., 2019). Indeed, by adding more time and/or call the child's attention led to more successful answers. Concerning the number of system adaptation prompts, these were higher in the third and fourth sessions where both game scenarios were performed. It is possible to observe that the number of system prompts that led to successful answers were higher (7) than the ones that led to unsuccessful answers (3). This might suggest that by adding more time or by calling the child's attention to interact with the OPT influenced positively the children's performance. Additionally, the number of these prompts decreased for both game scenarios along the sessions.

The proposed system to automatically detect the children's behaviours (typical and stereotypical) did not run in all sessions. The present approach uses the Nuitrack middleware (due to its real-time capabilities) as the first step in the pipeline in order to detect one person in the scene and extract skeleton data. Thus, most of the times the Nuitrack system was not able to detect the person within the scene and consequently the next steps of the behaviour detection pipeline could not be run. Additionally, even in the sessions where the system detected the children's behaviour, some of the times the system lost track of the person. This might happen due to light conditions or the way the user moved within the scene or by occlusion. In spite of this, in the sessions where the system detected the children's behaviour, it successfully detected the majority of the behaviours that the children exhibited. The exception was the CLAP behaviour which was wrongly detected. It was noted during the video analysis, that the CLAP behaviour was detect when the child moved the OPT closer to his/her face, raising both hands, which could mean that the trained model is only taking into account the positions and distance of the hands in relation to the body and not considering the slight movement of the hands.

No typical or stereotypical behaviours detected triggered any action by the system. As suggested by the participants interviewed in section 5.2.1, the system or experimenter should only intervene if the children present aggressive behaviours.

The results of the pre- and post-tests indicate that children performance improved, as there was an

increase in the number of correct matches in the post-test in comparison to the pre-test session.

Table 6.5: Statistical comparison of the children's percentage of successful answers.

Measure	Work	Sessions	Scenario	p	Differences
Percentage of Successful answers	Present study	first. Vs last.	Recognize	.30	First: 73.6%
					Last: 79.1%
			Storytelling	.08	First: 79.4%
					Last: 90.5%
	(Soares et al., 2019)	first. Vs last.	Recognize	<.01*	First: 50.5%
					Last: 73.4%
			Storytelling	<.001*	First: 62.7%
					Last: 82.3%

*statistical significative difference

Table 6.6: Statistical comparison of the average percentage of successful answers (SA) vs. the average percentage of unsuccessful plus unanswered prompts (UUP).

Measure	Work	Sessions	Scenario	p	Differences
SA Vs. UUP	Present study	First.	Recognize	.06	SA: 73.6%
					UUP: 26.4%
			Storytelling	.013*	SA: 79.4%
					UUP: 20.6%
		Last.	Recognize	.013*	SA: 79.1%
					UUP: 20.9%
			Storytelling	.013*	SA: 90.5%
					UUP: 9.5%
	(Soares et al., 2019)	First.	Recognize	>.05	SA: 52.7%
					UUP: 47.3%
			Storytelling	>.05	SA: 62.4%
					UUP: 37.6%
		Last.	Recognize	>.05	SA: 74.4%
					UUP: 25.6%
			Storytelling	<.003	SA: 82.0%
					UUP:18.0%

*statistical significative difference

6.5 Summary of hypothesis and conclusions

The present study targets the analysis of joint attention time and children's performance focusing on emotion recognition skills. The goal was to verify if the supervised behaviour architecture developed in the previous chapters had any measurable influence in the game scenarios which encourage the labelling and identification of emotions. Regarding the research questions presented in the beginning of the chapter, here highlighted in bold, the following implications were found:

RQ1: What are the differences in the children's behaviour when OPT is added to a HRI support session?

The average duration in seconds of the eye gaze towards to the OPT exceeded largely the eye gaze towards to the robot, in particular for the Storytelling game scenario. Meaning that the children indeed relayed on the OPT device, especially for the Storytelling game scenario, as in this games visual cues showing the story scenario are presented. Additionally, the joint attention time over the sessions in both game scenarios was consistently higher in comparison with the previous work in (Soares et al., 2019). Which might indicate that the addition of the OPT to the robot interaction successfully captivate the interest of the children in the both game scenarios. The number of successful answers were higher to the sum of the unsuccessful answers with the unanswered prompts and significant statistical differences were found when comparing this for the last session of the Recognize game scenario and for both first and last sessions for the Storytelling game scenario. Moreover, when comparing the number of successful answers in both game scenarios from the present work with (Soares et al., 2019), these numbers are higher in the first and last session in the present study. Since the activities and the robot are the same, the higher percentage of success in terms of answers in this study might indicate that the OPT was a facilitator in the interaction, and adding it to the interaction with the robot, both helped the children to, for example, understand the perspective of the character in the story. Additionally, the revised images of stories scenarios developed with the recommendations of the focus group may have helped the children to better understand the story.

However, the results comparing the pre-test to the post-test were not conclusive since there was not found significant statistical difference in the number of correct matches ($p = 0.05$), which might indicate that the children acquired the knowledge but had the difficulty to generalize it.

RQ2: What are the differences in the children's behaviour, or performance, when the robot modifies its behaviour according to the children's non-verbal actions or performance?

The verbalization duration in terms of speech was higher in the sessions where both games happened and significant statistical differences were found when comparing the average verbalization duration in these sessions during the Recognize game scenario with the Storytelling game scenario. There was a higher rate of system adaptation prompts in both first sessions of the game scenarios, with the adding more time the most frequent system suggestion. This higher rate of system adaptation events was more frequent in the Storytelling game scenario, especially in the first and second sessions where this game happened. This might indicate that the children verbalized more when the system prompt with more time for them to answer.

There was a higher number of system prompts that resulted in correct answers, which might suggest that adding more time or calling the child attention to the interaction influenced positively the children's performance. Moreover, the results regarding the performance of the children in both games scenarios were higher in the present study in comparison to (Soares et al., 2019), in both first and last sessions. Additionally, as the system intervention prompts decreased along the sessions, the rate of successful answers increased in both game scenarios.

RQ3: In what extent does the inclusion of social awareness behaviours and adaptation in a hybrid approach based on a robotic and OPT system benefit social interaction of children with another person (experimenter)?

None of the detected behaviours (typical and stereotypical) prompt the system to intervene. However, there were system intervention prompts in terms of adding more time and to call the children to interact with the OPT. In the sessions where the system intervention prompts happened the children tended to verbalized more, and the form of verbalization was speech. In fact, significant statistical differences were found when comparing the average duration of verbalization in the Recognize game scenario versus the Storytelling game scenario in the sessions where both games happened (sessions 3 and 4). Looking at the percentage of eye gaze towards the experimenter, no significant statistical differences were found. Nevertheless, in the sessions where the children verbalized more, the average percentage of the eye gaze towards the experimenter was higher. Additionally, it is worth to point out that there was an increase of joint attention time from the second to the third session where both game scenarios happen and more adaptations prompts where executed. This might indicate that the inclusion of some adaptation capabilities to the system may somewhat benefit the social bonding of the children with another person,

in this case the experimenter, as it was observed an increase of verbalization and eye gaze percentage towards the experimenter.

In summary, the present study aimed to promote social interaction and emotion recognition skills in children with ASD through the use of a triadic hybrid interaction (with a robot and an OPT), where children performance as well as verbal and non-verbal behaviours were analysed.

The presence of external factors (detailed in section 7.2) and the constraints like the inherent characteristics of a population of children with ASD with the addition of the inclusion criteria for the study influenced the final size of the sample. The conducted study is rather a quantitative study than an extensive proof of concept experiment. Additionally, the present sample reflects the statistics regarding the prevalence of ASD, that is ASD is 4 times more prevalent in males than females (CDC, 2022a). Even with the present sample size, the research highlighted interesting results regarding the performance and display of verbal and non-verbal behaviours by children with ASD in a triadic interaction where the OPT is used as an add-on and the system has supervised autonomous intervention capabilities. The most common non-verbal behaviour shown by the children was the IDLE behaviour, which might indicate that they normally and actively interacted in the activity. They also often verbalized (in terms of speech), especially in the sessions where system adaption occurred, being adding more time the most common. Thus, the introduction of some system adaptations might have encouraged verbal communication. Therefore, it can be paramount the introduction some degree of autonomy, enabling the robot to perform some autonomous behaviours whilst keeping track of the interaction data in a supervised manner.

The analysis of joint attention time allowed the identification of the children's main focus attention. It was observed that the eye gaze towards the OPT increased along the sessions. Furthermore, this increase coincides with the introduction of the Storytelling game scenario (started in the third session). Thus, the children used/relayed more on the OPT during the activities, particularly more in the Storytelling game scenario as it showed visual cues and, as supported by the literature (Dettmer et al., 2000), children with ASD appear to mainly use vision as their main input. Additionally, the use of the OPT eliminates the necessity of having several visual cards thus making the interaction seamlessly, as the interaction is dynamic. Moreover, the OPT can offer a unique way to quantify the child behaviour without being too invasive (as for example the use of wearable sensors). In the present study, the IMU data from the OPT is used to feed the system in order for the robot to "know" the level of engagement of the child, calling the child's attention if needed.

The children's performance in the two game scenarios provide evidence regarding the use of the present

hybrid approach (robot + OPT) as a valuable tool to encourage the acquisition of emotion recognition skills and promote social interaction with children with ASD. Even though it was not possible to verify if this new skill was generalized to other contexts, the Storytelling game scenario results provided strong insights in terms of:

- the children focus in the activity, especially the eye gaze towards the OPT when comparing it to the Recognize game scenario;
- the performance when comparing the successful answers to the sum of unsuccessful answers and the unanswered prompts;
- the number of system adaptation prompts that led to correct answers by the children;
- the change of behaviours by the children, namely the increase of verbalization and eye gaze towards the experimenter, especially in the sessions where more system interventions occurred.

References

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders, DSM5. In *Arlington* (5th. Ed.). APA. https://doi.org/10.1176/appi.books.9780890425596.744053

Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. In *Natural theories of mind: Evolution, development and simulation of everyday mindreading.* (pp. 233–251). Basil Blackwell.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

CDC. (2022). *Autism and Developmental Disabilities Monitoring (ADDM) Network | CDC*. https://www.cdc.gov/ncbddd/autism/addm.html

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Dettmer, S., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2000). The Use of Visual Supports to Facilitate Transitions of Students with Autism. *Focus on Autism and Other Developmental Disabilities*, *15*(3), 163–169. https://doi.org/10.1177/108835760001500307

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. In *Research in Autism Spectrum Disorders* (Vol.

6, Issue 1, pp. 249–262). Elsevier Ltd. https://doi.org/10.1016/j.rasd.2011.05.006

Johnson, C. P., Myers, S. M., Lipkin, P. H., Cartwright, J. D., Desch, L. W., Duby, J. C., Elias, E. R., Levey, E. B., Liptak, G. S., Murphy, N. A., Tilton, A. H., Lollar, D., Macias, M., McPherson, M., Olson, D. G., Strickland, B., Skipper, S. M., Ackermann, J., del Monte, M., … Yeargin-Allsopp, M. (2007). Identification and evaluation of children with autism spectrum disorders. In *Pediatrics* (Vol. 120, Issue 5, pp. 1183–1215). https://doi.org/10.1542/peds.2007-2361

Lund, H. H. (2009). Modular playware as a playful diagnosis tool for autistic children. *2009 IEEE International Conference on Rehabilitation Robotics, ICORR 2009*, 899–904. https://doi.org/10.1109/ICORR.2009.5209606

Lund, H. H., Klitbo, T., & Jessen, C. (2005). Playware technology for physically activating play. *Artificial Life and Robotics*, *9*(4), 165–174. https://doi.org/10.1007/s10015-005-0350-z

Mundy, P., Ungerer, J., & Sherman, T. (1986). DEFINING THE SOCIAL DEFICITS OF AUTISM. *Journal of Child Psychology and Psychiatry*, *27*(5), 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Noldus | Innovative solutions for behavioral research. (2015). http://www.noldus.com/

Noor, H. A. M., Shahbodin, F., & Che Pee, N. (2012). Serious Game for Autism Children: Review of Literature. *World Academy of Science, Engineering and Technology*, *64*(4), 647–652. https://doi.org/10.5176/2251-1679_CGAT30

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. In *Autism Research* (Vol. 9, Issue 2, pp. 165–183). https://doi.org/10.1002/aur.1527

Saleh, M. A., Hanapiah, F. A., & Hashim, H. (2021). Robot applications for autism: a comprehensive review. In *Disability and Rehabilitation: Assistive Technology* (Vol. 16, Issue 6, pp. 580–602). Taylor and Francis Ltd. https://doi.org/10.1080/17483107.2019.1685016

Scassellati, B., Henny Admoni, & Matarić, M. (2012). Robots for Use in Autism Research. *Annual Review of Biomedical Engineering*, *14*(1), 275–294. https://doi.org/10.1146/annurev-bioeng-071811-150036

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

7 CONCLUSIONS AND FUTURE WORK

Summary

This chapter presents the conclusions of the research developed in the framework of this thesis and outlines some possible future work avenues.

7 Conclusions and Future Work

7.1 Conclusions

7.2 Challenges and Rewards of Multidisciplinary Research

7.3 Operational Constraints

7.4 Future Work

7.1 Conclusions

The research developed in the framework of this thesis concerns the design and implementation of a supervised hybrid approach for developing emotional skills with children with ASD. The focus of this research involves the promotion of social interaction behaviours between a child with ASD and other person, using technology as a tool and object of shared attention. Children with ASD present social communication deficits in terms of verbal and non-verbal communication, eye gaze, among others. Therefore, developing tools in order to attract their attention are fundamental to develop a wide range of social skills. In the present research, skills such as eye gaze, verbal and non-verbal communication, and recognition of emotions were explored in triadic interactions with the robot and OPT as common objects of shared attention between the child with ASD and the experimenter. In order to fulfil this goal a supervised behavioural framework was developed. This framework has three key components: Perception, Reasoning, and Actuation layers.

The research goals are detailed in the following items with their corresponding conclusions:

G1: To evaluate the use of objects based on playware technology in a Human-Robot interaction (HRI) scenario:

An initial prototype (PlayCube) was developed in order to perceive the feasibility of an extra element in the triadic interaction. The tests conducted indicated that the children reacted positively to the OPT (PlayCube) and that they understood the mechanics of the OPT as well as the game scenario.

Following this trend, from the process of conducting a four-step study (focus group, think at loud, online questionnaire, and pilot study) that included qualitative as well as quantitative procedures it was gathered feedback from several professionals who work directly with children with ASD. With this feedback, the story scenarios were updated taking into account the colour pallet, the number of stimuli, and the spatial location of the elements of each scenario. Furthermore, a new OPT device was developed having in mind the guidelines and conclusions from the studies conducted with these specialists from different backgrounds. Their inputs were paramount especially in defining the form and size as well as the type of feedback that the OPT should have. The initial pilot study allowed to perceive that the children used/relayed on the OPT during the activities as there was a significant difference in the mean gaze time towards the OPT. Therefore, the overall process consisted in consulting-designing-testing-updating and until reaching the desired solution.

G2: To modify the robot behaviour according to eye/head movements patterns and non-verbal actions from children with ASD:

In the perception layer of the developed framework it was built the elements to extract the user/child head pose by using directly the output information of the Nuitrack SDK, which returns the pitch, roll, and yaw angles. Regarding the detection of non-verbal behaviours, initially it was selected based on previous work the most common stereotyped behaviours such as hand-flapping (or hand wave), rocking, covering the ears, among others. In addition, other behaviours (typical actions) were considered, namely, idle, standing, clap, hand raise, pointing, and turn. With these data a dataset was created and a machine learning model was trained on such behaviours. The trained model consisted in a convolutional neural network with six convolutional layers. The proposed model achieved an overall validation accuracy of 93.8% ± 0.6% and a test accuracy of 92.4% ± 0.0%. In summary, it was developed a gesture detection pipeline that starts by extracting the user joints position data (by using Nuitrack SDK) to then pre-process it and encoded the data into an image space representation. Then, this image representation serves as input to the trained machine learning model which then classifies the action. The whole pipeline is able to work in real-time, running on average at 31 FPS entirely on an Intel(R) Core (TM) i7-8650 quad-core CPU and without the use of a dedicated GPU. This allows to ensure the real-time factor and the easiness of conveyance of the system since it can run on a mini-PC or laptop.

G3: To equip a humanoid robot with a behaviour model by employing machine learning techniques which predicts the behaviour of a child with ASD based on previous non-verbal behaviours, eye/head movements and actions on objects based on playware technology:

In order to define the interaction flow of the session, which happens in the reasoning layer and is exhibited by the system in the actuation layer, a collection of semi-structured interviews was conducted with specialists with different backgrounds and a parent. The interaction flow can be influenced by the detected behaviours as well as the child's performance, namely the time he/she might take to answer the robots prompts. The results from the interviews allowed to define which actions of the child the robot should intervene, thus defining the limits of the robot behaviour adaptation according to the child's actions. Some of these actions included moments of distraction, when the child is taking a longer time to answer, if the child gets up, among others. The robot could intervene by calling the child's attention (through verbal commands), give more time for the child to answer or aid the child to use the OPT (using verbal prompts), among others. Additionally, participants also mentioned that a child will always have moments of out-of-

focus, and in these periods is where the experimenter can easily intervene (by touch for example). Therefore, the experimenter should have the control of the session, deciding if the robot should or not intervene in all scenarios. A fully autonomous approach is not desired, as it lacks the often important input of the supervising adult (Zaraki et al., 2018). Thus, it was taken a semi-autonomous supervised approach where the experimenter would have the sensibility to supervise when the systems prompts concerning adaption should be executed, as proposed in chapter 5.

G4: To evaluate the social interaction between the child with ASD, the robot and the experimenter, using the developed system:
The interaction was performed in a triadic setup in a semi-autonomous supervised manner with the robot and OPT being objects of shared attention. Children significantly gazed more at the OPT, in particular, during the Storytelling game scenario, indicating that the OPT was successful and a facilitator in attracting the child's attention. Moreover, the joint attention time remained high, above 80%, over all sessions. Thus, suggesting that the OPT could be used a complement to the traditional HRI with children with ASD. System adaptation prompts contributed for the success of the children during the activities, as there was a higher number of these prompts that led to more successful answers. Although no significant statistical differences were found, there was a higher percentage gaze towards to the experimenter in sessions where more system adaptations were executed. In fact, it may also have contributed for the longer duration in verbal behaviour (in terms of speech) as significant statistical differences were found.
This might indicate that the inclusion of some adaptation capabilities to the system may somewhat benefit the social interaction of the children with another person, in this case the experimenter, as it was observed an increase of verbalization and eye gaze percentage towards the experimenter.

Joint attention time is a difficult task for children with ASD in almost all their growth phases, influencing non-verbal behaviours, for example eye gaze, as mentioned in section 2.1. Additionally, children with ASD present limitations with respect to understanding the perspective of others and interacting with their peers. The type of support method presented in section 2.1.2 was adapted to include a robot and an object with playware technology, forming a triadic interaction between the child with ASD and an adult. The environment was arranged to decrease distracting factors and to support in individual contexts. Visual supports and answer cards were embedded within the OPT programming seamlessly appearing in the screen when necessary, decreasing the number of extra objects in the interaction.
The results showed thar using a hybrid supervised autonomous approach with the robot and OPT being

objects of shared attention, may benefit the social interaction of the children with another person while encouraging the acquisition of emotion recognition skills by children with ASD.

Hence, the research questions (RQ) presented in section 1.2 were answered, and the proposed hypothesis were tested. Not all expectations were satisfied. Nevertheless, the answers to the RQs may indicate future paths to follow:

RQ1: What are the differences in the children's behaviour when OPT is added to a HRI support session?

The joint attention time increased over the sessions (with exception of the second one) and in parallel the eye gaze time towards the OPT increased, having the children spent significantly more time gazing at the OPT in comparison to the robot, especially in the Storytelling game scenario. In contrast to similar work in the literature (Soares et al., 2019), in the present work the joint attention time stayed high in both game scenarios. In fact it was hypothesised that children would be more engaged with the introduction of the OPT in a HRI. Thus, it might indicate that the addition of the OPT to the robot interaction successfully captivate the interest and attention of the children. The use of the OPT eliminates the necessity of having several visual cards thus making the interaction seamlessly, promoting a more dynamic and friendly interaction.

RQ2: What are the differences in the children's behaviour, or performance, when the robot modifies its behaviour according to the children's non-verbal actions or performance?

Children´s performance has a positive trend in both game scenarios, being higher in both game scenarios when comparing to (Soares et al., 2019). In addition, there was a higher number of system adaptations prompts that resulted in correct answers, which might indicate that adding more time or calling the child attention to the interaction influenced positively the children's performance. Children verbalised significantly more in particular in the sessions where more system adaptation prompts were executed. It is worth to point out that the system intervention prompts decreased along the sessions while the rate of successful answers increased in both game scenarios.

RQ3: In what extent does the inclusion of social awareness behaviours and adaptation in a hybrid approach based on a robotic and OPT system benefit social interaction of children with another person (experimenter)?

None of the detected social awareness behaviours (typical and stereotypical, such as covering the ears

and rocking) prompt the system to directly intervene. Nonetheless, in the sessions where the system intervention prompts happened (in terms of adding more time or calling the child attention to interact with the OPT) the children tended to verbalize more. It was expected a change in the children eye gaze towards the experimenter. Thus, this expectation was partially met, since there was not found significant statistical difference in terms of the children eye gaze percentage towards the experimenter. However, the lack of eye contact is a characteristic that defines children in the autism spectrum. Nevertheless, in the sessions where the children verbalized more, the average percentage of the eye gaze towards the experimenter was higher. This might indicate that the inclusion of some adaptation capabilities to the system may somehow benefit the social interaction of the children with another person, in this case the experimenter, as it was observed an increase of verbalization and eye gaze percentage towards the experimenter.

In review, the present work contributed to several applicative domains of research:

Robot-Playware Assisted Play: it was proposed a methodological approach of how to design, conduct, and analyse robot-assisted play with playware technology devices with children with ASD. It was developed the guidelines concerning on how to design, implement, and test Objects with Playware Technology.

Assistive Technology: a supervised autonomous approach with the robot and OPT being objects of shared attention was developed. The developed game scenarios were updated and fitted within the OPT. These can be adapted to each child, by the level of autonomy introduced in the system, namely giving more time to answer, call the children attention to the activity, among others. In addition, the feedback provided by the robot and the OPT can be tailored to each child. The development of a deep learning model that allows to detect and record the non-verbal behaviours (typical and stereotypical) during the sessions.
This allowed to tackle some of the main difficulties that are inherent characteristics of this target group, such as: attracting the child's attention, the association of specific situations to emotions, and motivate the interaction with a peer.

Learning and Development: the use of the present supervised hybrid approach demonstrated to be useful in assisting children with ASD to improve emotion recognition and social skills, while helping them to focus their attention.

ASD Research: the conducted studies may lead to a new method for support in ASD. It could be used as a complement for learning and, potentially, as an add-on to traditional approaches, diversifying the supporting practices. Thus, the present approach could offer a way for creating learning opportunities that imply the training of competences, for example via repetition, implying greater levels of attention and involvement in the task while providing adequate reinforcements and adaptation. At the same time, it can promote higher levels of motivation and child participation in the development of emotion recognition skills.

7.2 Challenges and Rewards of Multidisciplinary Research

Multidisciplinary research is essential, but it has its challenges. Combining and investigating different areas of knowledge involves a great deal of commitment and determination. Nevertheless, it is rewarding that the children interacted with the robot and the OPT, especially verbally by sharing their thoughts on how the robot felt in a particular story or communicating how they would feel in a similar situation. Indeed, it was surprising that they interacted verbally with the experimenter, as they struggle in Human-Human Interaction. The adaptation feature in terms of adding more time was important as each child has its own time to answer. This was mostly seen in the first sessions of each game scenario. The forms of feedback offered by the robot, the OPT, and the experimenter motivated the child along the interaction and the children tended to ask to play more with the robot.

Informally, the therapists and parents reported that the children were excited to play with the robot. In particular, one child was asking for ZECA most of the time and in fact in some occasions, at the end of the session some children tended to hug the robot.

These are some examples of how a tool that is different from the usual can impact the support sessions of children in this target group.

7.3 Operational Constraints

Autism Spectrum Disorder (ASD) is characterized by a triad of difficulties in relation to interaction, communication, and reduced interest. Furthermore, a child can be anywhere on the autism spectrum. It requires some social adaptation by the experimenter to adapt to each child, while guaranteeing that the constraints of the interactions are the same for all children.

Since March of 2020, with the declaration of a state of emergency in the context of the COVID-19 pandemic (Decreto do Presidente da República n.º 14-A/2020, 18 of March), the execution of the tasks defined in the project work plan was affected. According to the rules in force at the time, several key tasks had to be adjusted or postponed due to impossibility of execution.

In particular, this project requires direct contact with children, which was prohibited due to the implementation of the rules defined by the General Directorate of Health (DGS in Portuguese) and limitations imposed by the state of emergency: social distance, the restriction of free movement and the closure of schools and support institutions/associations for children with specific needs (Decreto n.º 2-A/2020, of 20 of March; Despacho RT-06/2021 da Universidade do Minho).

This unforeseen situation produced a significant delay in the schedule, especially in carrying out the tests in a school/clinical environment with the target group.

Following this, the Portuguese Foundation for Science and Technology (FCT) opened exceptional COVID-19 impact mitigation grant calls providing up to an extra year for research. As the present work was heavily impacted by the COVID-19 pandemic, an application was submitted to extend the scholarship in order to allow conducting the experiments initially planned. The decision stated that: *"...in this case, it is not evident that there has been serious damage and that, consequently, obtaining the doctor's degree has been compromised..."*. The full statement is disclosed in the Appendix A.8.

With this decision, some adaptations to the initial plan were implemented. This primarily affected the tests that were conducted in schools and/or clinics, impacting how they were conducted and the size of the sample. A quantitative study was conducted, and in order to minimize the number of contacts, the study was performed in one place. COVID-19 tests were done prior conducting the quantitative study and pilot studies. Masks were used by the experimenter in all the sessions which may have impacted in some way the interaction with the child.

It is worth to point out that the end of the state of alert was declared only on September, 30 2022 [https://www.portugal.gov.pt/pt/gc23/comunicacao/noticia?i=fim-do-estado-de-alerta].

In spite of the challenges, the results of this thesis and the feedback from the professionals after the

conclusion of the procedures are very encouraging.

7.4 Future Work

Further developments could be divided into short-term research and long-term research. As short-term research, and more on the technical side, the behavioural model could be expanded in order to detect more typical and stereotypical behaviours. The behaviour pipeline could be updated to better detect and track the user, and other machine learning techniques could be employed to better classify the behaviours. Further adaptations could be introduced to the system and additional tests could be conducted.

A report could be printed out containing a summary of the session with the performance of the child, the detected behaviours, among other metrics.

In addition, several experimental studies could be conducted in order to narrow down the beneficial factors of the developed system, and its impact. For example, further studies could follow the same line of research of the present work and investigate the impact of further system adaptations in the interaction between the robot, child, and experimenter.

A larger study could be conducted so that there is a greater validity of the results and a greater representation of children with ASD and their behavioural variability. A larger sample would also allow for the possibility of generalizing the results obtained in this study.

Regarding long-term research, during the present study, namely in the quantitative study, it was noticed some verbal communication during the interactions with children with ASD. Although in some cases children presented very limited speaking skills, a system could be developed to promote social verbalization with children with ASD. It could be studied the effects of the robot support sessions with and OPT device used to promote verbal communication with children with ASD. The system could be equipped with simple speech to text model or a more complex language model that can interpret and appropriately respond to the child's prompts. New game scenarios could be developed or the same game scenarios could be used with simple adaptations as for example asking the child to answer verbally. This would be the preferred method to answer. However, in order to ensure some answers from the child, the OPT could be used also as an input method for answering. Further adaptations could be implemented taking advantage of the behavioural model to detect child's actions and promote more verbalization.

In order to study the effect of this kind of interaction, it could be measured the change in the children's vocabulary used before and after the experimental phase, in terms of frequency, duration, as well as, if

they adopted new words into their vocabularies.

This research could be the next step, combining the present study approach with the verbalization feature in order to promote social interaction with more focus on communication skills.

References

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot. *Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems*, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

REFERENCES

Abdul-Azim, H. A., & Hemayed, E. E. (2015). Human action recognition using trajectory-based representation. *Egyptian Informatics Journal*, *16*(2), 187–198. https://doi.org/10.1016/j.eij.2015.05.002

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In *Handbook of Practical Program Evaluation: Fourth Edition*. https://doi.org/10.1002/9781119171386.ch19

Affectiva. (2023). *Solutions - Affectiva : Affectiva*. https://www.affectiva.com/what/products/

Afsar, P., Cortez, P., & Santos, H. (2015). Automatic human action recognition from video using Hidden Markov Model. *Proceedings - IEEE 18th International Conference on Computational Science and Engineering, CSE 2015*, 105–109. https://doi.org/10.1109/CSE.2015.41

Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. In *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2014.04.011

Ahmad, M., & Lee, S. W. (2006). HMM-based human action recognition using multiview image sequences. *Proceedings - International Conference on Pattern Recognition*, *1*, 263–266. https://doi.org/10.1109/ICPR.2006.630

Almeida, L., & Freire, T. (2010). *Metodologia da investigação em psicologia e educação* (6ª). Psiquilíbrios.

Amado, J. (2014). *Manual de investigação qualitativa em educação* (2ª, Ed.). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0879-2

Ambady, N., & Rosenthal, R. (1992). *Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis* (pp. 256–274).

American Psychiatric, A. (2013). Dsm-5. In *Diagnostic and statistical manual of mental disorders, DSM-5*. https://doi.org/doi.org/10.1176/appi.books.9780890425596.dsm17

American Psychiatric Association. (2013a). Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). *Diagnostic and Statistical Manual of Mental Disorders 4th Edition TR.*, 280. https://doi.org/10.1176/appi.books.9780890425596.744053

American Psychiatric Association. (2013b). Diagnostic and Statistical Manual of Mental Disorders, DSM5.

In *Arlington* (5th. Ed.). APA. https://doi.org/10.1176/appi.books.9780890425596.744053

Azevedo, J., Silva, V., Soares, F., Pereira, A. P., & Esteves, J. S. (2018). An application to promote emotional skills in children with autism spectrum disorders. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11243 LNCS*. https://doi.org/10.1007/978-3-030-02762-9_30

Bader, R. (2006). *Using Social Stories to Increase Emotion Recognition and Labeling in School-Age Children with Autism*. Alliant International University.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*. http://arxiv.org/abs/1803.01271

Bakeman, R., & Gottman, J. M. (1997). *Observing Interaction: An Introduction to Sequential Analysis* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Baltrusaitis, T. (2014). Automatic facial expression analysis. In *Pattern recognition*. University of Cambridge.

Baltrusaitis, T., Robinson, P., & Morency, L. P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477553

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L. P. (2018). OpenFace 2.0: Facial behavior analysis toolkit. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*. https://doi.org/10.1109/FG.2018.00019

Barajas, A. O., al Osman, H., & Shirmohammadi, S. (2017). A Serious Game for children with Autism Spectrum Disorder as a tool for play therapy. *2017 IEEE 5th International Conference on Serious Games and Applications for Health, SeGAH 2017*. https://doi.org/10.1109/SeGAH.2017.7939266

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology*, *84*, 490–515.

Baron-Cohen, S. (1991). Precursors to a theory of mind: Understanding attention in others. In *Natural theories of mind: Evolution, development and simulation of everyday mindreading.* (pp. 233–251). Basil Blackwell.

Baron-Cohen, S., Golan, O., & Ashwin, E. (2009). Can emotion recognition be taught to children with autism spectrum conditions? In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2009.0191

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ?

Cognition, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2). https://doi.org/10.1017/S0021963001006643

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or Asperger Syndrome. *Visual Cognition*, *4*(3). https://doi.org/10.1080/713756761

Begum, M., Serna, R. W., Kontak, D., Allspaw, J., Kuczynski, J., Yanco, H. A., & Suarez, J. (2015). Measuring the Efficacy of Robots in Autism Therapy. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 335–342. https://doi.org/10.1145/2696454.2696480

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013a). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013b). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Bennett, M., & Goodall, E. (2016). A Meta-analysis of DSM-5 Autism Diagnoses in Relation to DSM-IV and DSM-IV-TR. *Review Journal of Autism and Developmental Disorders*, *3*(2). https://doi.org/10.1007/s40489-016-0070-4

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., & Chetouani, M. (2014). Interactive Technologies for Autistic Children: A Review. *Cognitive Computation*, *6*(4), 722–740. https://doi.org/10.1007/s12559-014-9276-x

Breazeal, C. (2000). Sociable machines: Expressive social exchange between humans and robots. In *Expressive Social Exchange Between Humans and Robots*.

Cannon, W. B. (1927). The James-Lange theory of emotions: a critical examination and an alternative

theory. *The American Journal of Psychology*, *39*, 106–124. https://doi.org/10.2307/1415404

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

CDC. (202 C.E.). *Signs and Symptoms of Autism Spectrum Disorders | CDC*. https://www.cdc.gov/ncbddd/autism/signs.html

CDC. (2022a). *Autism and Developmental Disabilities Monitoring (ADDM) Network | CDC*. https://www.cdc.gov/ncbddd/autism/addm.html

CDC. (2022b). *Diagnostic Criteria | Autism Spectrum Disorder (ASD) | CDC*. https://www.cdc.gov/ncbddd/autism/hcp-dsm.html

Chevalier, P., Li, J. J., Ainger, E., Alcorn, A. M., Babovic, S., Charisi, V., Petrovic, S., Schadenberg, B. R., Pellicano, E., & Evers, V. (2017). Dialogue Design for a Robot-Based Face-Mirroring Game to Engage Autistic Children with Emotional Expressions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10652 LNAI*, 546–555. https://doi.org/10.1007/978-3-319-70022-9_54

Christensen, D. L., Baio, J., van Naarden Braun, K., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L. C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M. S., Zahorodny, W., & Yeargin-Allsopp, M. (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*. https://doi.org/10.15585/mmwr.ss6503a1

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. In *Child Development* (Vol. 75, Issue 2). https://doi.org/10.1111/j.1467-8624.2004.00673.x

Costa, A. P., Charpiot, L., Lera, F. R., Ziafati, P., Nazarikhorram, A., van der Torre, L., & Steffgen, G. (2018). More Attention and Less Repetitive and Stereotyped Behaviors using a Robot with Children with Autism. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2018.8525747

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders* [Doctoral Thesis]. University of Minho.

Costa, S., Soares, F., & Santos, C. (2013). Facial expressions and gestures to convey emotions with a humanoid robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), *8239 LNAI*, 542–551. https://doi.org/10.1007/978-3-319-02675-6_54

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders*, *2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

D., C. P., Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., Souberman, E., & Wertsch, J. v. (1979). L. S. Vygotsky: Mind in Society. The Development of Higher Psychological Processes. *The American Journal of Psychology*, *92*(1). https://doi.org/10.2307/1421493

Daily, S. B., James, M. T., Cherry, D., Porter, J. J., Darnell, S. S., Isaac, J., & Roy, T. (2017). Affective Computing: Historical Foundations, Current Applications, and Future Trends. In *Emotions and Affect in Human Factors and Human-Computer Interaction*. https://doi.org/10.1016/B978-0-12-801851-4.00009-4

Damasio, A. (1994). Descartes' error: Emotion, rationality and the human brain. In *New York: Putnam*.

Darwin, C. (1872). The expression of the emotions in man and animals. *London, UK: John Marry*, 374. https://doi.org/10.1037/h0076058

Dautenhahn, K., & Billard, A. (2002). Games children with autism can play with Robota, a humanoid robotic doll. *Universal Access and Assistive Technology*, 179–190. https://doi.org/10.1007/978-1-4471-3719-1_18

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009a). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3–4), 369–397. https://doi.org/10.1080/11762320903123567

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009b). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3–4), 369–397. https://doi.org/10.1080/11762320903123567

Dautenhahn, K., & Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, *12*(1), 1–35. https://doi.org/10.1075/pc.12.1.03dau

Dautenhahn, K., & Werry, I. (2007). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & CognitionPragmatics and Cognition*. https://doi.org/10.1075/pc.12.1.03dau

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress.

Developmental Psychology, *40*(2). https://doi.org/10.1037/0012-1649.40.2.271

Dettmer, S., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2000). The Use of Visual Supports to Facilitate Transitions of Students with Autism. *Focus on Autism and Other Developmental Disabilities*, *15*(3), 163–169. https://doi.org/10.1177/108835760001500307

Dhingra, N. (2022). LwPosr: Lightweight Efficient Fine Grained Head Pose Estimation. *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022*. https://doi.org/10.1109/WACV51458.2022.00127

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. In *Research in Autism Spectrum Disorders* (Vol. 6, Issue 1, pp. 249–262). Elsevier Ltd. https://doi.org/10.1016/j.rasd.2011.05.006

D'Mello, S., & Calvo, R. (2013). Beyond the basic emotions: what should affective computing compute? *... Abstracts on Human Factors in Computing ...*, 2287–2294. https://doi.org/10.1145/2468356.2468751

Dollar, P., Rabaud, V., Cottrell, G., & Belongie, S. (2005). Behavior recognition via sparse spatio-temporal features. *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 65–72. https://doi.org/10.1109/VSPETS.2005.1570899

Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2015.7298714

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, *19*.

Ekman, P., Friesen, W. V, O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. In *Journal of Personality and Social Psychology* (Vol. 53, Issue 4, pp. 712–717). American Psychological Association. https://doi.org/10.1037/0022-3514.53.4.712

el Kaliouby, R., Picard, R., & Baron-Cohen, S. (2006). Affective computing and autism. *Annals of the New York Academy of Sciences*, *1093*, 228–248. https://doi.org/10.1196/annals.1382.016

Esteban, P. G., Baxter, P., Belpaeme, T., Billing, E., Cai, H., Cao, H. L., Coeckelbergh, M., Costescu, C., David, D., de Beir, A., Fang, Y., Ju, Z., Kennedy, J., Liu, H., Mazel, A., Pandey, A., Richardson, K., Senft, E., Thill, S., ... Ziemke, T. (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn*, *8*(1), 18–38.

https://doi.org/10.1515/pjbr-2017-0002

Ferrari, E., Robins, B., & Dautenhahn, K. (2009). Therapeutic and educational objectives in robot assisted play for children with autism. *IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man)*, 108–114. https://doi.org/10.1109/ROMAN.2009.5326251

Gaglio, S., Re, G. lo, & Morana, M. (2015). Human Activity Recognition Process Using 3-D Posture Data. *IEEE Transactions on Human-Machine Systems*. https://doi.org/10.1109/THMS.2014.2377111

Gehring, J., Auli, M., Grangier, D., & Dauphin, Y. N. (2017). A convolutional encoder model for neural machine translation. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 123–135. https://doi.org/10.18653/v1/P17-1012

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*(2). https://doi.org/10.1007/s10803-005-0057-y

Goncalves, N., Rodrigues, J. L., Costa, S., & Soares, F. (2012a). Automatic detection of stereotyped hand flapping movements: Two different approaches. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*. https://doi.org/10.1109/ROMAN.2012.6343784

Goncalves, N., Rodrigues, J. L., Costa, S., & Soares, F. (2012b). Preliminary study on determining stereotypical motor movements. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1598–1601. https://doi.org/10.1109/EMBC.2012.6346250

Gray, C. A., & Garand, J. D. (1993). Social Stories: Improving Responses of Students with Autism with Accurate Social Information. *Focus on Autism and Other Developmental Disabilities*, *8*(1). https://doi.org/10.1177/108835769300800101

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. In *Autism*. https://doi.org/10.1177/1362361313476767

Güss, C. D. (2018). What is going through your mind? Thinking aloud as a method in cross-cultural psychology. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/fpsyg.2018.01292

Hampshire, P. K., & Hourcade, J. J. (2016). Teaching Play Skills to Children with Autism Using Visually Structured Tasks. *TEACHING Exceptional Children*, *46*(3), 26–31. https://doi.org/10.1177/004005991404600303

Hanley-Hochdorfer, K., Bray, M. A., Kehle, T. J., & Elinoff, M. J. (2010). Social stories to increase verbal

initiation in children with autism and asperger's disorder. *School Psychology Review*, *39*(3). https://doi.org/10.1080/02796015.2010.12087767

Happé, F., Briskman, J., Frith, U., Happé, F., & Frith, U. (2001). Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: II. Real-life skills and preferences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(3), 309–316. https://doi.org/10.1017/S0021963001006916

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. In *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-005-0039-0

Hashimoto, T., Kobayashi, H., & Kato, N. (2011). Educational system with the android robot SAYA and field trial. *IEEE International Conference on Fuzzy Systems*, *8*(3), 766–771. https://doi.org/10.1109/FUZZY.2011.6007430

Hidalgo, G., Cao, Z., Simon, T., Wei, S.-E., Raaj, Y., Joo, H., & Sheikh, Y. (2022). *CMU-Perceptual-Computing-Lab/openpose - System Requirements*. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/installation/0_index.md#operating-systems-requirements-and-dependencies

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion: a further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(February), 671–680. https://doi.org/10.1111/j.1469-7610.1986.tb01836.x

HTC. (2023). *VIVE - VR Headsets, Games, and Metaverse Life | United States*. https://www.vive.com/us/

Hudlicka, E. (2008). Affective Computing for Game Design. *Scientist*, 5–12.

Hudock, R. L., Kashima-Ellingson, Y., & Bellini, S. (2011). Increasing Responses to Verbal Greetings in Children With Autism Spectrum Disorders: A Comparison of the Effectiveness of Social Story and Generic Story Interventions. In *School Psychology Forum* (Vol. 5, Issue 3).

Huijnen, C. A. G. J., Lexis, M. A. S., Jansens, R., & de Witte, L. P. (2016). Mapping Robots to Therapy and Educational Objectives for Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-016-2740-6

Intel. (2022). *Intel NUC 10 Performance kit NUC10i7FNKN Product Specifications*. https://ark.intel.com/content/www/us/en/ark/products/214614/intel-nuc-10-performance-kit-nuc10i7fnkn.html

Intel RealSense. (2020). *Depth Camera D435 – Intel® RealSense™ Depth and Tracking Cameras*. https://www.intelrealsense.com/depth-camera-d435/

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*.

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205. http://www.jstor.org/stable/2246769

Jazouli, M., Majda, A., Merad, D., Aalouane, R., & Zarghili, A. (2019). Automatic detection of stereotyped movements in autistic children using the Kinect sensor. *International Journal of Biomedical Engineering and Technology*. https://doi.org/10.1504/IJBET.2019.097621

Johnson, C. P., Myers, S. M., Lipkin, P. H., Cartwright, J. D., Desch, L. W., Duby, J. C., Elias, E. R., Levey, E. B., Liptak, G. S., Murphy, N. A., Tilton, A. H., Lollar, D., Macias, M., McPherson, M., Olson, D. G., Strickland, B., Skipper, S. M., Ackermann, J., del Monte, M., … Yeargin-Allsopp, M. (2007). Identification and evaluation of children with autism spectrum disorders. In *Pediatrics* (Vol. 120, Issue 5, pp. 1183–1215). https://doi.org/10.1542/peds.2007-2361

Kahn, P. H., Gary, H. E., & Shen, S. (2013). Children's Social Relationships With Current and Near-Future Robots. *Child Development Perspectives*, *7*(1), 32–37. https://doi.org/10.1111/cdep.12011

Kanner, L. (1943). Autistic disturbances of affective contact. In *Nervous Child* (Vol. 2, pp. 217–250). https://doi.org/10.1105/tpc.11.5.949

Ke, Q., Bennamoun, M., An, S., Sohel, F., & Boussaid, F. (2017). A new representation of skeleton sequences for 3D action recognition. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 4570–4579. https://doi.org/10.1109/CVPR.2017.486

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049. https://doi.org/10.1007/s10803-012-1645-2

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, *59*(9), 809. https://doi.org/10.1001/archpsyc.59.9.809

Knight, V., McKissick, B. R., & Saunders, A. (2013). A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. In *Journal of Autism and Developmental Disorders* (Vol. 43, Issue 11, pp. 2628–2648). https://doi.org/10.1007/s10803-013-1814-y

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, *23*(14). https://doi.org/10.1163/016918609X12518783330360

Koutras, P., Potamianos, G., Efthymiou, N., Maragos, P., Filntisis, P. P., & Tsiami, A. (2018). *Multi3:*

Multi-Sensory Perception System for Multi-Modal Child Interaction with Multiple Robots. https://doi.org/10.1109/icra.2018.8461210

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2008). Keepon. *International Journal of Social Robotics*, *1*(1), 3–18. https://doi.org/10.1007/s12369-008-0009-8

Lange, C. (1887). *Ueber gemuthsbewgungen* (3rd ed., Vol. 8).

Laptev, I., Marszałek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587756

Laraba, S., Tilmanne, J., & Dutoit, T. (2019). Leveraging Pre-trained CNN Models for Skeleton-Based Action Recognition. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11754 LNCS* (Issue November). Springer International Publishing. https://doi.org/10.1007/978-3-030-34995-0_56

Lecciso, F., Levante, A., Fabio, R. A., Caprì, T., Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., & Petrocchi, S. (2021). Emotional Expression in Children With ASD: A Pre-Study on a Two-Group Pre-Post-Test Design Comparing Robot-Based and Computer-Based Training. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.678052

Lee, J., & Ahn, B. (2020). Low-Cost RGB Camera and Mobile Robot Platform. *Sensors (Switzerland)*, *20*, 1–12. https://doi.org/doi:10.3390/s20102886

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5). https://doi.org/10.1016/j.jpsychires.2010.10.008

Li, C., Hou, Y., Wang, P., & Li, W. (2017). Joint Distance Maps Based Action Recognition with Convolutional Neural Networks. *IEEE Signal Processing Letters*, *24*(5), 624–628. https://doi.org/10.1109/LSP.2017.2678539

Li, W., Zhang, Z., & Liu, Z. (2010). Action recognition based on a bag of 3D points. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*. https://doi.org/10.1109/CVPRW.2010.5543273

Lima, C. B. de, & Levy, P. de Q. (2015). A Perturbação do Espetro do Autismo. In *Perturbações do Neurodesenvolvimento: Manual de orientações diagnósticas e estratégias de intervenção*.

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

Liu, B., Yu, H., Zhou, X., Tang, D., & Liu, H. (2017). Combining 3D joints Moving Trend and Geometry

property for human action recognition. *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*. https://doi.org/10.1109/SMC.2016.7844262

Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016). Spatio-temporal LSTM with trust gates for 3D human action recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9907 LNCS*, 816–833. https://doi.org/10.1007/978-3-319-46487-9_50

Loveland, K. A., Tunali-Kotoski, B., Chen, Y. R., Ortegon, J., Pearson, D. A., Brelsford, K. A., & Gibbs, M. C. (1997). Emotion recognition in autism: Verbal and nonverbal information. *Development and Psychopathology*, *9*(3). https://doi.org/10.1017/s0954579497001351

Ludl, D., Gulde, T., & Curio, C. (2019). Simple yet efficient real-time pose-based action recognition. *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*. https://doi.org/10.1109/ITSC.2019.8917128

Lund, H. H. (2009). Modular playware as a playful diagnosis tool for autistic children. *2009 IEEE International Conference on Rehabilitation Robotics, ICORR 2009*, 899–904. https://doi.org/10.1109/ICORR.2009.5209606

Lund, H. H., Dam Pedersen, M., & Beck, R. (2009). Modular robotic tiles: Experiments for children with autism. *Artificial Life and Robotics*, *13*(2), 394–400. https://doi.org/10.1007/s10015-008-0623-4

Lund, H. H., Klitbo, T., & Jessen, C. (2005). Playware technology for physically activating play. *Artificial Life and Robotics*, *9*(4), 165–174. https://doi.org/10.1007/s10015-005-0350-z

LuxAI. (2023). *LuxAI - Award winning social robots for autism and special needs education*. https://luxai.com/

Málaga, I., Lago, R. B., Antonio, H.-F., Álvarez-Álvarez, N., Oreña-Ansonera, V. A., & Baeza-Velasco, M. (2019). Prevalencia de los trastornos del espectro autista en niños en Estados Unidos, Europa y España: coincidencias y discrepancias. *Medicina (Buenos Aires)*, *79*, 1–6. http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0025-76802019000200003&lng=es&tlng=es.

Mammarella, I. C., Giofrè, D., Caviola, S., Cornoldi, C., & Hamilton, C. (2014). Visuospatial working memory in children with autism: The effect of a semantic global organization. *Research in Developmental Disabilities*, *35*(6), 1349–1356. https://doi.org/10.1016/j.ridd.2014.03.030

Marques, C. (1998). *Perturbações do espectro do autismo: ensaio de uma intervenção construtivista e desenvolvimentalista com mães*. Universidade de Coimbra.

Mazzei, D., Greco, A., Lazzeri, N., Zaraki, A., Lanata, A., Igliozzi, R., Mancini, A., Stoppa, F., Scilingo, E. P., Muratori, F., & de Rossi, D. (2012). Robotic social therapy on children with autism: Preliminary evaluation through multi-parametric analysis. *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*. https://doi.org/10.1109/SocialCom-PASSAT.2012.99

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., Muratori, F., & de Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2011*, 4515–4518. https://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., Muratori, F., & De Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Engineering in Medicine, Proceedings of the Annual International Conference of the IEEE*, *2011*, 4515–4518. https://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Hanson, D., & de Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, 195–200. https://doi.org/10.1109/BioRob.2012.6290687

McDaniel, J., Brady, N. C., & Warren, S. F. (2022). Effectiveness of Responsivity Intervention Strategies on Prelinguistic and Language Outcomes for Children with Autism Spectrum Disorder: A Systematic Review and Meta-Analysis of Group and Single Case Studies. *Journal of Autism and Developmental Disorders*, *52*(11). https://doi.org/10.1007/s10803-021-05331-y

McQuarrie, E. F., Stewart, D. W., & Shamdasani, P. N. (1991). Focus Groups: Theory and Practice. *Journal of Marketing Research*. https://doi.org/10.2307/3172875

Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., Seidel, H. P., Rhodin, H., Pons-Moll, G., & Theobalt, C. (2020). XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera. *ACM Transactions on Graphics*, *39*(4). https://doi.org/10.1145/3386569.3392410

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H. P., Xu, W., Casas, D., & Theobalt, C. (2017). VNect: Real-time 3D human pose estimation with a single RGB camera. *ACM Transactions on Graphics*, *36*(4). https://doi.org/10.1145/3072959.3073596

Melo, F. S., Sardinha, A., Belo, D., Couto, M., Faria, M., Farias, A., Gambôa, H., Jesus, C., Kinarullathil,

M., Lima, P., Luz, L., Mateus, A., Melo, I., Moreno, P., Osório, D., Paiva, A., Pimentel, J., Rodrigues, J., Sequeira, P., ... Ventura, R. (2019). Project INSIDE: towards autonomous semi-unstructured human–robot social interaction in autism therapy. *Artificial Intelligence in Medicine*, *96*. https://doi.org/10.1016/j.artmed.2018.12.003

Meta. (2023). *Meta Quest Pro: Our Most Advanced New VR Headset | Meta Store | Meta Store*. https://www.meta.com/quest/quest-pro/

Michaud, F., Laplante, J.-F., Larouche, H., Duquette, a., Caron, S., Letourneau, D., & Masson, P. (2005). Autonomous spherical mobile robot for child-development studies. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 471–480. https://doi.org/10.1109/TSMCA.2005.850596

Microsoft. (2017). *Developing with Kinect*. https://developer.microsoft.com/en-us/windows/kinect/develop

Microsoft. (2021). *Azure Kinect DK Windows comparison | Microsoft Docs*. https://docs.microsoft.com/en-us/azure/kinect-dk/windows-comparison

Mundy, P., Ungerer, J., & Sherman, T. (1986). DEFINING THE SOCIAL DEFICITS OF AUTISM. *Journal of Child Psychology and Psychiatry*, *27*(5), 657–669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(4). https://doi.org/10.1109/TPAMI.2008.106

Myles, Brenda Smith; Swanson, Terri Cooper, & Holverstott, J. (2013). Autism spectrum disorders: a handbook for parents and professionals. *Choice Reviews Online*. https://doi.org/10.5860/choice.45-5017

Nachson, I. (1995). On the Modularity of Face Recognition: The Riddle of Domain Specificity*. *Journal of Clinical and Experimental Neuropsychology*, *17*(2). https://doi.org/10.1080/01688639508405122

Noldus | Innovative solutions for behavioral research. (2015). http://www.noldus.com/

Noor, H. A. M., Shahbodin, F., & Che Pee, N. (2012). Serious Game for Autism Children: Review of Literature. *World Academy of Science, Engineering and Technology*, *64*(4), 647–652. https://doi.org/10.5176/2251-1679_CGAT30

Nuitrack. (2019). *Nuitrack Full Body Skeletal Tracking Software - Kinect replacement for Android, Windows, Linux, iOS, Intel RealSense, Orbbec*. https://nuitrack.com/

Nuitrack Face Tracking. (2019). *nuitrack-sdk/Unity_Face_Tracking.md at master · 3DiVi/nuitrack-sdk · GitHub*. https://github.com/3DiVi/nuitrack-sdk/blob/master/doc/Unity_Face_Tracking.md

Nuitrack Frame Rate. (2018). *Skeleton tracking specs ? 60 - 90 fps? - Nuitrack SDK - Nuitrack*. https://community.nuitrack.com/t/skeleton-tracking-specs-60-90-fps/916

OOM with CUDNN installed and GPU requirements met · Issue #1929. (2021). https://github.com/CMU-Perceptual-Computing-Lab/openpose/issues/1929

OpenCV. (2020). *OpenCV: Deep Neural Networks (dnn module)*. https://docs.opencv.org/master/d2/d58/tutorial_table_of_content_dnn.html

OpenFace CPU. (2016). *High CPU Usage · Issue #82 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/82

OpenFace Pupil and Gaze Estimations. (2021). *Inaccurate pupil (and gaze) estimations · Issue #958 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/958

OpenFace Pupil Estimation. (2018). *Inaccurate Iris/Pupil Localization · Issue #359 · TadasBaltrusaitis/OpenFace · GitHub*. https://github.com/TadasBaltrusaitis/OpenFace/issues/359

OptiTrack. (2023). *OptiTrack - Motion Capture Systems*. https://optitrack.com/

ORBBEC. (2019). *Astra Series – Orbbec*. https://orbbec3d.com/product-astra-pro/

Orbbec. (2020). *Astra | Orbbec 3D*. https://shop.orbbec3d.com/Astra

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. In *Autism Research* (Vol. 9, Issue 2, pp. 165–183). https://doi.org/10.1002/aur.1527

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2018). Exploiting deep residual networks for human action recognition from skeletal data. *Computer Vision and Image Understanding*. https://doi.org/10.1016/j.cviu.2018.03.003

Pham, H. H., Salmane, H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2019). Spatio–temporal image representation of 3D skeletal movements for view-invariant action recognition with deep convolutional neural networks. *Sensors (Switzerland)*, *19*(8). https://doi.org/10.3390/s19081932

Piaget. (1976). *Piaget and His School* (B. Inhelder, H. H. Chipman, & C. Zwingmann, Eds.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-46323-5

Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, *14*(Spitz 1945), 141–169.

Picard, R. W. (1997a). Affective Computing. In *Rosalind W. Picard*. http://www.amazon.com/exec/obidos/ASIN/0262161702/acmorg-20

Picard, R. W. (1997b). Affective Computing. In *Pattern Recognition* (Vol. 73). https://doi.org/10.1007/BF01238028

Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1535), 3575–3584. https://doi.org/10.1098/rstb.2009.0143

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. http://arxiv.org/abs/1804.02767

Ricks, D. J., & Colton, M. B. (2010a). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. https://doi.org/10.1109/ROBOT.2010.5509327

Ricks, D. J., & Colton, M. B. (2010b). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. https://doi.org/10.1109/ROBOT.2010.5509327

Robins, B., Dautenhahn, K., & Dickerson, P. (2009). From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009*, 205–211. https://doi.org/10.1109/ACHI.2009.32

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., & Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *2005*. https://doi.org/10.1109/ROMAN.2005.1513864

Robins, B., Dickerson, P., Stribling, P., & Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction Studies*, *5*(2). https://doi.org/10.1075/is.5.2.02rob

Robins, B., Ferrari, E., & Dautenhahn, K. (2008). Developing scenarios for robot assisted play. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*. https://doi.org/10.1109/ROMAN.2008.4600663

Robinson, P., & el Kaliouby, R. (2009). Computation of emotions in man and machines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3441–3447. https://doi.org/10.1098/rstb.2009.0198

Robokind. (2023). *Robokind | Advanced Social Robots*. https://www.robokind.com/

Robots4Autism. (2018). *Milo Robot*. https://robots4autism.com/

Rodriguez, M. D., Ahmed, J., & Shah, M. (2008). Action MACH: A spatio-temporal maximum average correlation height filter for action recognition. *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587727

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, *46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

Ruiz, N., Chong, E., & Rehg, J. M. (2018). Fine-grained head pose estimation without keypoints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2018-June*. https://doi.org/10.1109/CVPRW.2018.00281

Saleh, M. A., Hanapiah, F. A., & Hashim, H. (2021). Robot applications for autism: a comprehensive review. In *Disability and Rehabilitation: Assistive Technology* (Vol. 16, Issue 6, pp. 580–602). Taylor and Francis Ltd. https://doi.org/10.1080/17483107.2019.1685016

Samuels, R., & Stansfield, J. (2012). The effectiveness of social stories™ to develop social interactions with adults with characteristics of autism spectrum disorder. *British Journal of Learning Disabilities*, *40*(4). https://doi.org/10.1111/j.1468-3156.2011.00706.x

Sanders, J. L. (2009). Qualitative or quantitative differences between asperger's disorder and autism? Historical considerations. *Journal of Autism and Developmental Disorders*, *39*(11). https://doi.org/10.1007/s10803-009-0798-0

Sansosti, F. J., Powell-Smith, K. A., & Kincaid, D. (2004). A Research Synthesis of Social Story Interventions for Children With Autism Spectrum Disorders. *Focus on Autism and Other Developmental Disabilities*, *19*(4). https://doi.org/10.1177/10883576040190040101

Scassellati, B. (2007). How social robots will help us to diagnose , treat , and understand autism. *Robotics Research*, 552–563. https://doi.org/10.1007/978-3-540-48113-3_47

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., Ventola, P., & Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, *3*(21). https://doi.org/10.1126/SCIROBOTICS.AAT7544

Scassellati, B., Henny Admoni, & Matarić, M. (2012). Robots for Use in Autism Research. *Annual Review of Biomedical Engineering*, *14*(1), 275–294. https://doi.org/10.1146/annurev-bioeng-071811-150036

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state.

Psychological Review., *69(5)*, 379–399. https://doi.org/10.1037/h0046234

Schadenberg, B. R., Heylen, D. K. J., & Evers, V. (2017). Affect bursts to constrain the meaning of the facial expressions of the humanoid robot Zeno. *CEUR Workshop Proceedings*, *2059*.

Schopler, E., & van Bourgondien, M. E. (2020). Treatment and education of autistic and related communication handicapped children. In *Autistic Adults at Bittersweet Farms*. https://doi.org/10.4324/9781003075240-8

Schuetze, M., Rohr, C. S., Dewey, D., McCrimmon, A., & Bray, S. (2017). Reinforcement learning in autism spectrum disorder. In *Frontiers in Psychology* (Vol. 8, Issue NOV). https://doi.org/10.3389/fpsyg.2017.02035

Shahroudy, A., Liu, J., Ng, T. T., & Wang, G. (2016). NTU RGB+D: A large scale dataset for 3D human activity analysis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 1010–1019. https://doi.org/10.1109/CVPR.2016.115

Shotton, J., Sharp, T., Fitzgibbon, A., Blake, A., Cook, M., Kipman, A., Finocchio, M., & Moore, R. (2013). Real-Time human pose recognition in parts from single depth images. *Communications of the ACM*, *56*(1). https://doi.org/10.1145/2398356.2398381

Siegel, B. (2008). *O mundo da criança com autismo: compreender e tratar perturbações do espectro do autismo.* Porto Editora.

Silva, V., Queirós, S., Soares, F., Esteves, J. S., & Matos, D. (2019). A supervised autonomous approach for robot intervention with children with autism spectrum disorder. *ICINCO 2019 - Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics*, *2*, 497–503. https://doi.org/10.5220/0007958004970503

Silva, V., Soares, F., Esteves, J., Pereira, A., Leão, C., & Queirós, S. (2021). Hybrid Approach to Promote Social Interaction with Children with Autism Spectrum Disorder. *7th International Conference on Information and Communication Technologies for Ageing Well and E-Health*, 69–77. https://doi.org/10.5220/0010468600690077

Silva, V., Soares, F., Esteves, J. S., Pereira, A. P., & Matos, D. (2020). Adequacy of Game Scenarios for an Object with Playware Technology to Promote Emotion Recognition in Children with Autism Spectrum Disorder. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-27928-8_44

S.Messinger, D. (2014). *Affective Computing, Emotional Development, and Autism*. 516–536.

Soares, F. O., Costa, S. C., Santos, C. P., Pereira, A. P. S., Hiolle, A. R., & Silva, V. (2019). Socio-emotional development in high functioning children with Autism Spectrum Disorders using a humanoid robot.

Interaction StudiesInteraction Studies Social Behaviour and Communication in Biological and Artificial Systems, *20*(2), 205–233. https://doi.org/10.1075/is.15003.cos

Softbank Robotics. (2006). *NAO Robot*. https://www.ald.softbankrobotics.com/en/robots/nao

Solomon, R. C., & Griffiths, P. E. (1999). What Emotions Really are: The Problem of Psychological Categories. *The Philosophical Review*, *108*(1). https://doi.org/10.2307/2998275

Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Review*, *81*(2). https://doi.org/10.1037/h0036128

Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2017). An end-to-end spatio-temporal attention model for human action recognition from skeleton data. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*.

Sosnowski, S., Kuehnlenz, K., & Buss, M. (2006). EDDIE - An Emotion Display with Dynamic Intuitive Expressions. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3113–3118. https://doi.org/10.1109/IROS.2006.282208

Stake, R. E. (2010). Qualitative Research: Studying How Things Work. In *Qualitative research: Studying how things work*. NY: Guilford Press.

Stewart, D. W., & Shamdasani, P. N. (2014). *Focus Groups: Theory and Practice* (Third). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/focus-groups/book239690

Suzuki, K. (2014). Social Playware: Device-mediated social interaction for therapeutic activities. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*. https://doi.org/10.1109/ROMAN.2014.6926232

Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., Drew, A., Charman, T., Rees, L., & Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(5). https://doi.org/10.1017/S0021963098002595

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308

Tasnim, N., Islam, M. K., & Baek, J. H. (2021). Deep learning based human activity recognition using spatio-temporal image formation of skeleton joints. *Applied Sciences (Switzerland)*, *11*(6). https://doi.org/10.3390/app11062675

Toshev, A., & Szegedy, C. (2014). DeepPose: Human pose estimation via deep neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2014.214

Vicon. (2023). *Vicon | Award Winning Motion Capture Systems*. https://www.vicon.com/

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2010). Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. *2010 10th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2010*. https://doi.org/10.1109/ICHR.2010.5686346

Wang, H., Hu, J., & Deng, W. (2018). Face Feature Extraction: A Complete Review. *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2017.2784842

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang, J., Liu, Z., Wu, Y., & Yuan, J. (2012). Mining actionlet ensemble for action recognition with depth cameras. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2012.6247813

Wang, Y. A., Yu, X., Silverman, P. M., Harris, R. L., & Edward, H. (2010). *NIH Public Access*. *385*(1), 22–29. https://doi.org/10.1016/j.jmb.2008.10.054.The

Weber, M. (2011). The Nature of Social Action. In *Max Weber: Selections in Translation*. https://doi.org/10.1017/cbo9780511810831.005

webQDA – Qualitative Data Analysis Software. (2022). https://www.webqda.net/

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. In *Neuroscience and Biobehavioral Reviews* (Vol. 25, Issue 4, pp. 287–295). https://doi.org/10.1016/S0149-7634(01)00014-8

Wilson, G. (1996). Mindblindness—an essay on autism and theory of mind. *Personality and Individual Differences*, *20*(2). https://doi.org/10.1016/s0191-8869(96)80006-7

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29. https://doi.org/10.1007/BF01531288

Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., & Syrdal, D. S. (2013). Robot-Mediated Interviews - How Effective Is a Humanoid Robot as a Tool for Interviewing Young Children? *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059448

Wu, Y., & Ji, Q. (2019). Facial Landmark Detection: A Literature Survey. *International Journal of Computer*

Vision, *127*(2). https://doi.org/10.1007/s11263-018-1097-z

Xia, L., Chen, C. C., & Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3D joints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2012.6239233

Yang, T. Y., Chen, Y. T., Lin, Y. Y., & Chuang, Y. Y. (2019). Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*. https://doi.org/10.1109/CVPR.2019.00118

Yannakakis, G. N., Lund, H. H., & Hallam, J. (2007). Modeling children's entertainment in the playware playground. *Proceedings of the 2006 IEEE Symposium on Computational Intelligence and Games, CIG'06*, 134–141. https://doi.org/10.1109/CIG.2006.311692

Yannakakis, G. N., M, O., Hallam, J., & Lund, H. H. (2006). Comparative Fun Analysis in the Innovative Playware Game Platform. *In Proceedings of the 1st World Conference for Fun 'n Games*.

Ye, M., Wang, X., Yang, R., Ren, L., & Pollefeys, M. (2011). Accurate 3D pose estimation from a single depth image. *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2011.6126310

Yirmiya, N., Sigman, M. D., Kasari, C., & Mundy, P. (1992). Empathy and Cognition in High-Functioning Children with Autism. *Child Development*, *63*(1). https://doi.org/10.1111/j.1467-8624.1992.tb03603.x

Zaraki, A., Wood, L., Robins, B., & Dautenhahn, K. (2018). Development of a Semi-Autonomous Robotic System to Assist Children with Autism in Developing Visual Perspective Taking Skills. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 969–976. https://doi.org/10.1109/ROMAN.2018.8525681

Zhang, H., Wang, M., Liu, Y., & Yuan, Y. (2020). FDN: Feature decoupling network for head pose estimation. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v34i07.6974

Zhang, Y., Pezeshki, M., Brakel, P., Zhang, S., Laurent, C., Bengio, Y., & Courville, A. (2016). Towards end-to-end speech recognition with deep convolutional neural networks. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *08-12-Sept*, 410–414. https://doi.org/10.21437/Interspeech.2016-1446

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE Computer Society Conference on Computer*

Vision and Pattern Recognition, *2016-Decem*, 2921–2929. https://doi.org/10.1109/CVPR.2016.319

Zou, J., Gauthier, S., Gargot, T., Archambault, D., Cohen, D., & Anzalone, S. M. (2021). Development of a wizard-of-Oz based framework for QTRobot to facilitate the learning of handwriting in children with dysgraphia. *CEUR Workshop Proceedings*, *2806*.

APPENDIXES

A.1 Technical Drawings of the Robot

A.2 Social Stories

A.3 Visual Cues used with the Social Stories

A.4 Informed Consent – children's parents and the professionals

A.5 Informed consent for the participants of the Focus Group and SSIs

A.6 Questionnaire to characterize the child

A.7 Ethics Committee Documents

A.8 FCT Statement – exceptional COVID-19 impact mitigation grant call

A.1 Technical Drawings of the Robot

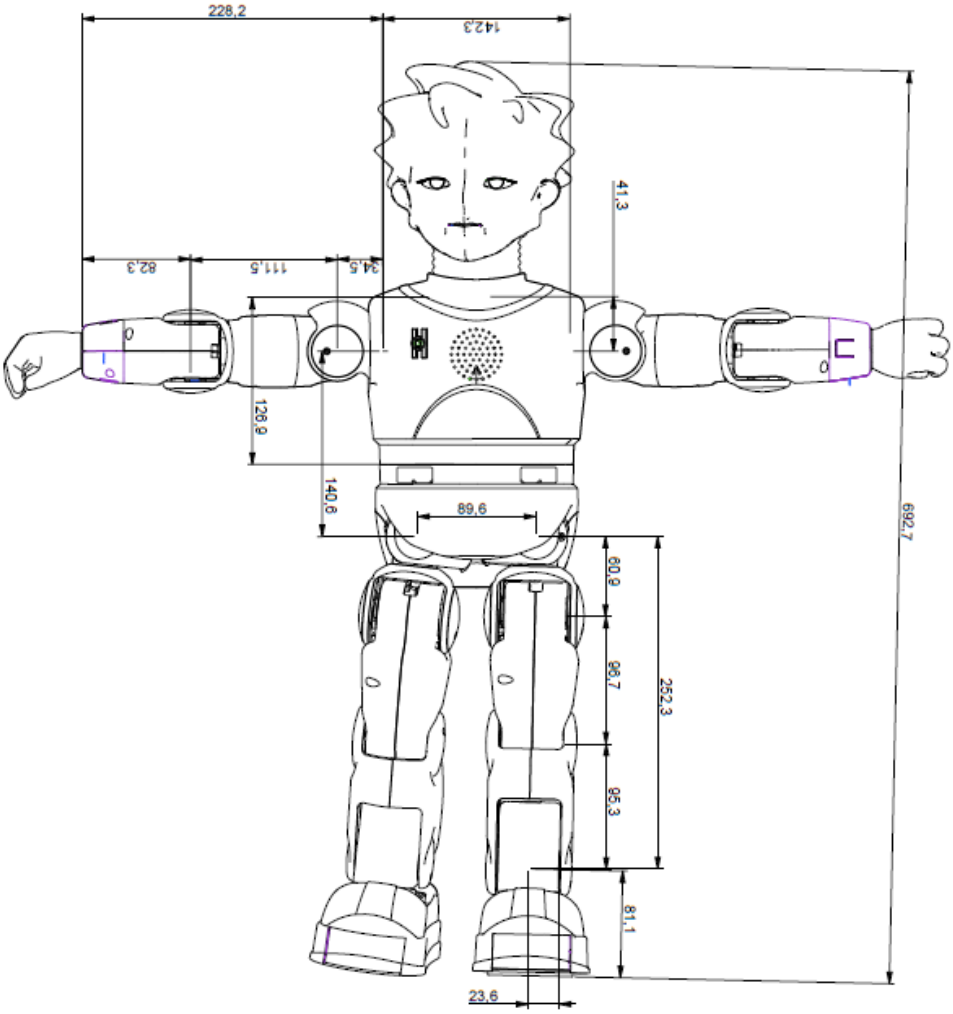


Figure A. 1: Robot technical drawing showing its entire body in a front view (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

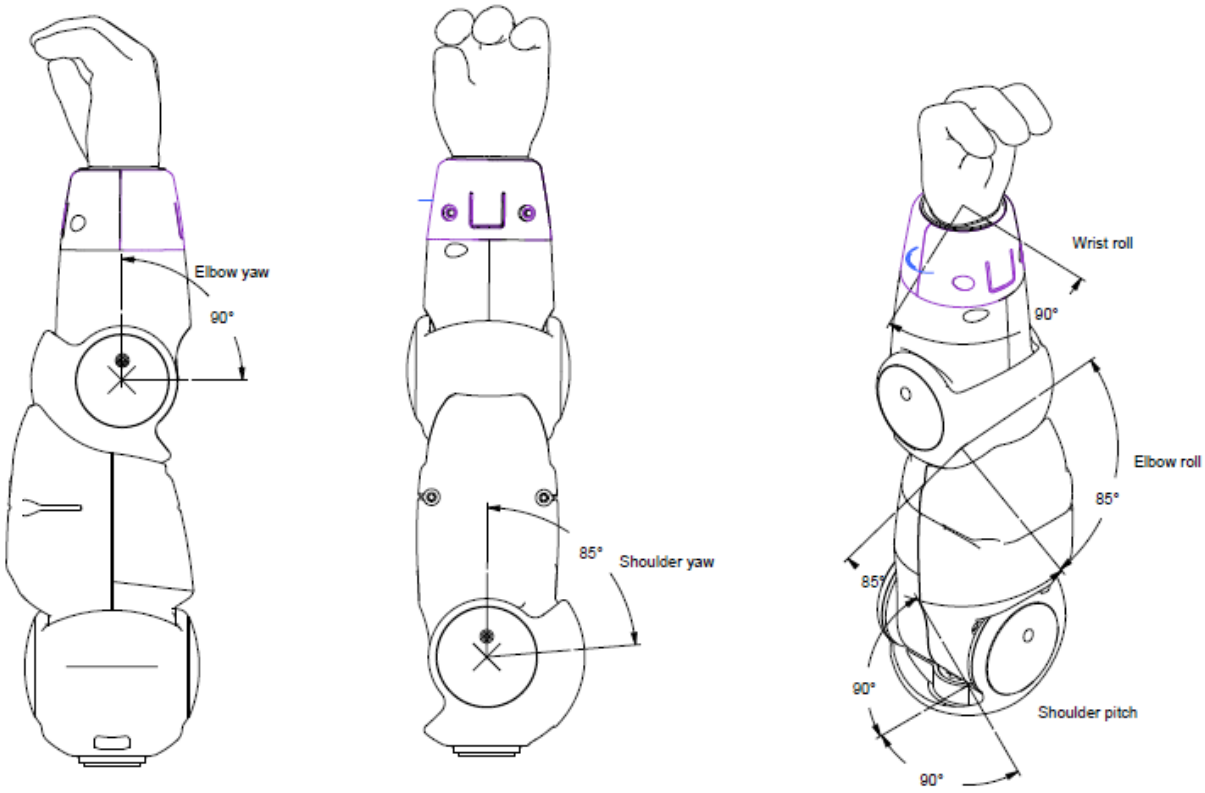


Figure A. 2: Technical drawing of the robot showing its arms (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

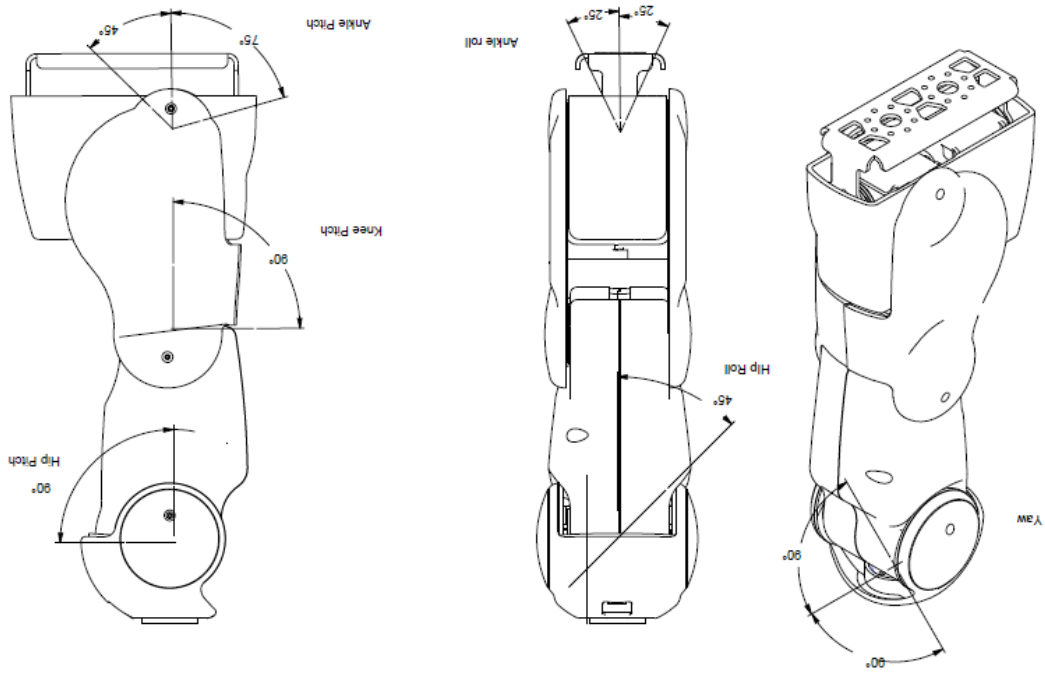


Figure A. 3 Technical drawing of the robot showing its legs (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

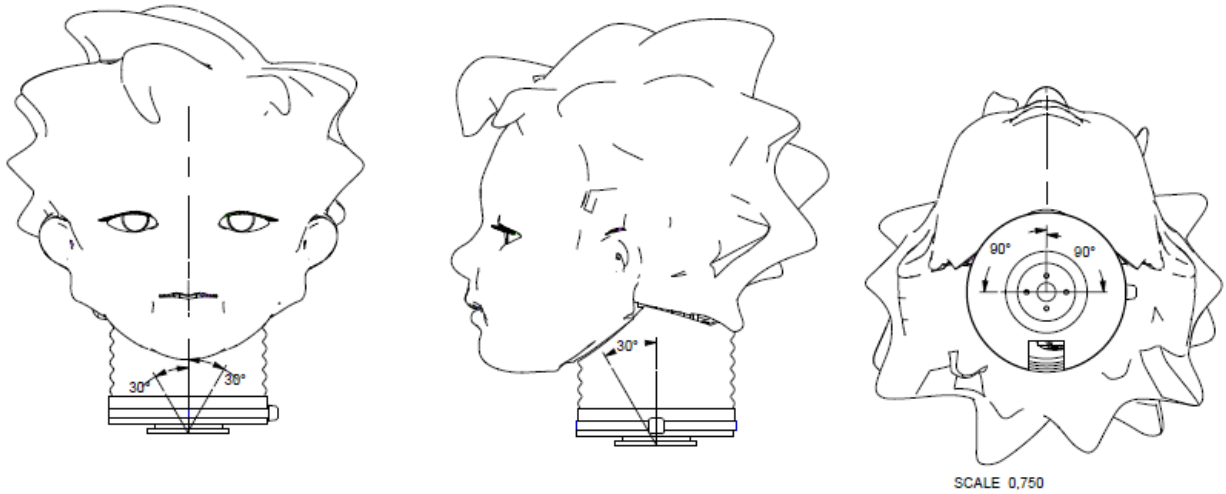


Figure A. 4: Robot technical drawing showing specifically its head (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

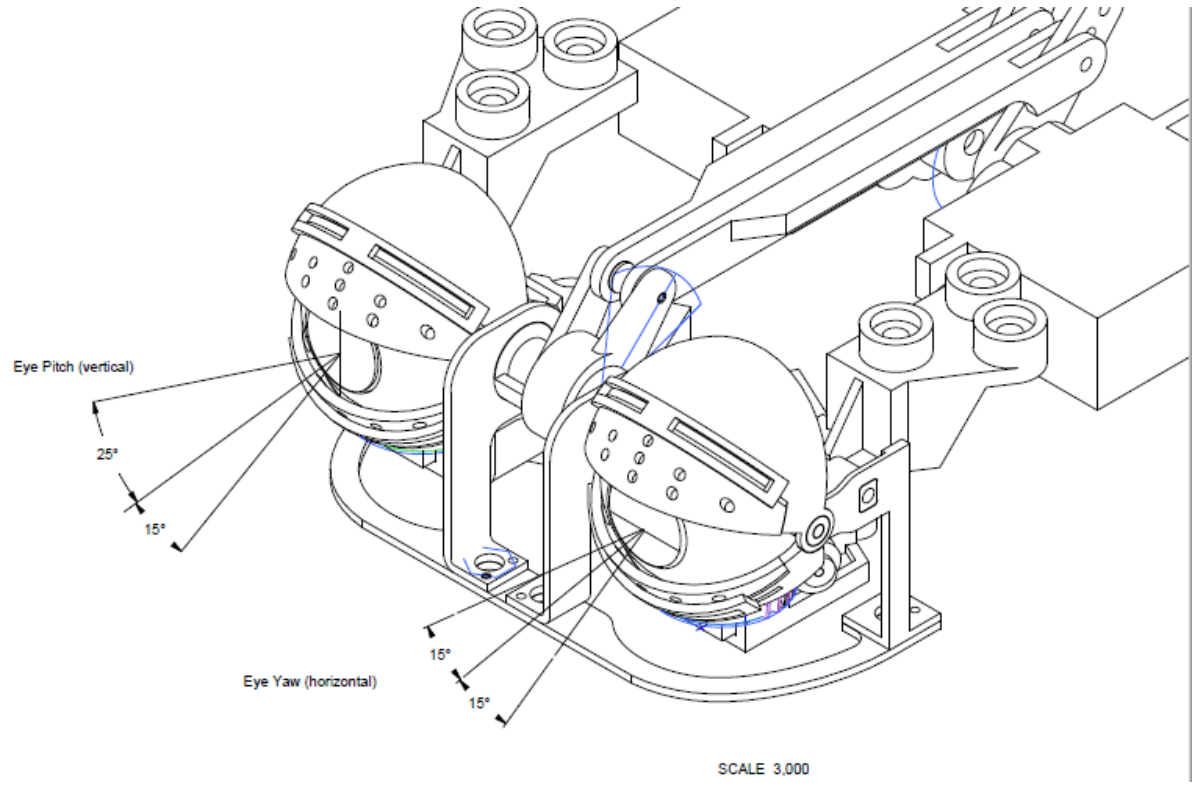


Figure A. 5: Technical drawing of the robot showing the detailing the eyes (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

A.2 Social Stories

Table A. 1: Social Stories used in the Storytelling game scenario.

#	Emotion	Story
1	Anger	My sister is called Alice. Alice plays with me in the playground. Today, when we were playing, Alice took my ball. I hate when Alice does this, so I was very angry.
2	Anger	In the classroom, I like to play with my blocks. My sister Alice sits next to me every day. Alice knocked down my blocks, on purpose. I was very upset.
3	Anger	Every day I play games with my friends in the classroom. I love to do puzzles, but when I lose a piece, I can not finish the puzzle and it makes me very angry.
4	Fear	At the end of dinner I go to my room. When I'm lying in bed, ready to sleep, my mom turns off the light. So, I see shadows on the wall, I can not sleep and I'm scared.
5	Fear	On Sundays I go for a walk with my parents. When the weather is nice, we go to the park and we play a lot. One day, while playing football, I was away and when I did not see my parents any more. I was very scared.
6	Fear	I go with my mother shopping. I like to choose the yogurts that I eat. Today, at the exit of the supermarket, a very large dog began barking very loud. I was shaking.
7	Joy	Every day in the morning I go to school. I enjoy playing with my friends. My teacher said I did a good job. It is good to do what we like.
8	Joy	Today I went for a walk and to play football with my friends from school. To play football is great fun and makes me smile. I smile when I do activities I like, and that makes me happy.
9	Joy	In my school, there are many computers. Sometimes another child is using the computer and I have to wait. It is so nice when it is my turn to use the computer.
10	Sadness	When I was playing in the playground, I fell on the floor. My arm and my leg were hurting a lot. I had a big scratch, and I could not stop crying.
11	Sadness	I like to play when I'm home. Today I took my ball and played with the ball in the living room. I kicked strongly the ball and broke a window. My mother me grounded and I cried.
12	Sadness	My friend is called John. John did not come to school today. I asked my teacher for John, and the teacher said that John is sick. That day I had to play alone.
13	Surprise	Every day, I go to school. One day when I entered the classroom, everyone screamed "Congratulations Zeca!" Because it was my birthday. I was so amazed.
14	Surprise	Every day, my father picks me up from school. One day, my father gave me an ice cream of my favorite flavor and I was very much in awe.
15	Surprise	I love Christmas. And like very much when me and my sister Alice open our Christmas presents. When I opened the greatest gift and I saw the toy I wanted I was really amazed.

A.3 Visual Cues used with the Social Stories



Figure A. 6: Story #1.



Figure A. 7: Story #2.



Figure A. 8: Story #3.



Figure A. 9: Story #4.

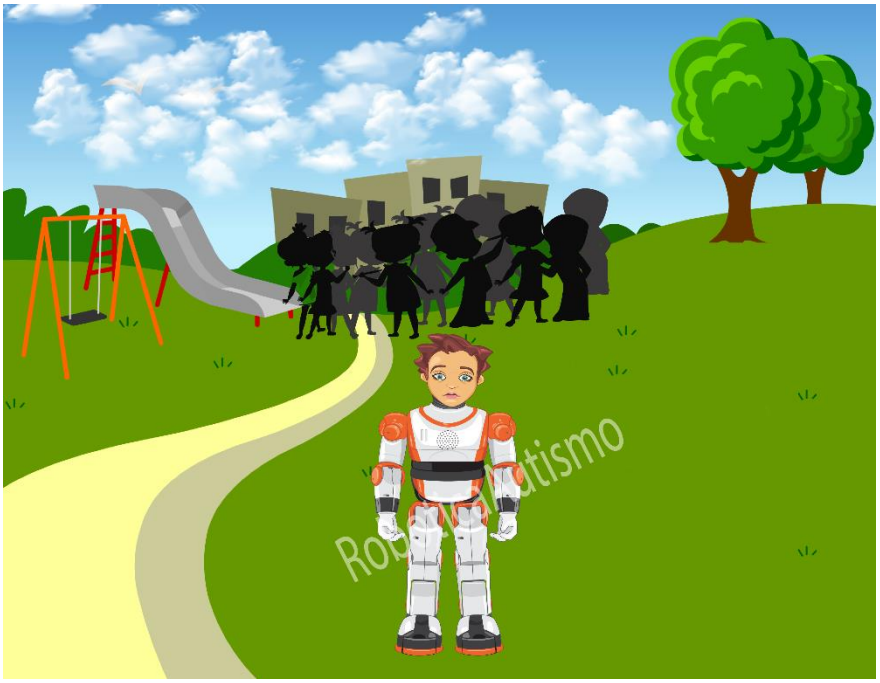


Figure A. 10: Story #5.



Figure A. 11: Story #6.

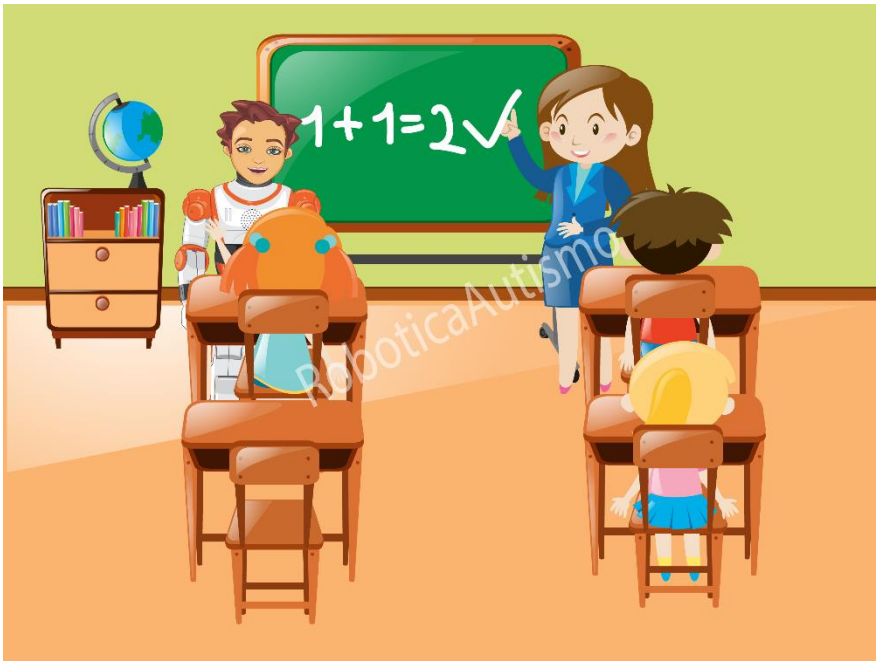


Figure A. 12: Story #7.



Figure A. 13: Story #8.



Figure A. 14: Story #9.



Figure A. 15: Story #10.



Figure A. 16: Story #11.



Figure A. 17: Story #12.



Figure A. 18: Story #13.



Figure A. 19: Story #14.



Figure A. 20: Story #16.

A.4 Informed Consent – children's parents and the professionals

ANEXO VII - Exemplar do consentimento informado e da informação fornecida aos encarregados de educação

22 de Outubro de 2021

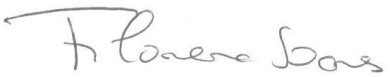
Exmo. Senhor (a) Encarregado de Educação/Tutor (a),

O **projeto Robótica-Autismo** (http://robotica-autismo.dei.uminho.pt/) visa a aplicação de ferramentas robóticas como forma de melhorar a vida social de crianças e jovens com **perturbação do espectro do autismo (**PEA). Em particular, o objetivo é **melhorar as habilidades de interação e comunicação com o meio-ambiente e com outras pessoas**. Em particular, no âmbito de doutoramento de Vinícius Corrêa Alves da Silva pretendemos investigar a capacidade da ferramenta robótica potencializar o reconhecimento de emoções e expressões faciais.

Uma vez que as sessões de trabalho implicarão a participação do seu educando, tal só será possível com a sua autorização. Garantimos o anonimato dos dados e das imagens e que os dados recolhidos serão apenas utilizados nesta investigação e na sua divulgação científica, nunca explorando os mesmos noutras situações que não estejam relacionadas com o projeto. Solicitamos, assim, a sua colaboração dando o seu consentimento através da devolução do destacável anexo.

Agradecendo desde já a sua participação, com os melhores cumprimentos,

A Coordenadora Científica do Projeto e orientadora do doutoramento, UM



_____

Filomena Maria da Rocha Menezes de Oliveira Soares

O aluno de Doutoramento responsável pelo estudo, UM



_____

Vinícius Corrêa Alves da Silva

Contactos: roboticaautismo@gmail.com ou a65312@alunos.uminho.pt; Vinícius Silva: 917672420

Eu_____ encarregado(a) de educação do(a)/tutor(a) de _____ declaro ter compreendido os objetivos de quanto me foi proposto e explicado pelo investigador que assina este documento, ter-me sido dada oportunidade de fazer todas as perguntas sobre o assunto e para todas elas ter obtido resposta esclarecedora, ter-me sido garantido que não haverá prejuízo para os direitos assistenciais se eu recusar esta solicitação, e ter-me sido dado tempo suficiente para refletir sobre esta proposta.

Declaro também que autorizo o meu (minha) educando(a) a participar no Projeto de Investigação "Ferramentas Robóticas na Educação Especial", em particular no "Jogo de Emoções com o Zeca", o que implicará a realização de experiências de 10 a 20 minutos com a supervisão de um adulto. As experiências serão filmadas para posterior análise de desempenho do educando. As filmagens e os dados decorrentes destas serão utilizados única e exclusivamente na divulgação científica do projeto. Os dados pessoais e os dados obtidos na investigação não serão divulgados e serão mantidos por um período de dez anos, ao fim do qual serão destruídos. Todas as informações de caráter pessoal recolhidas no decurso da investigação serão consideradas confidenciais e tratadas de acordo com as regras relativas à proteção de dados e da vida privada. Se o encarregado de educação o entender, o aluno pode abandonar o projeto em qualquer altura. A sua participação, e/ou a do seu dependente, ou recusa em participar ou posterior abandono, não prejudicarão a relação com a equipa de clínicos e/ou investigadores. Não se preveem quaisquer riscos para os participantes durante as sessões. Caso a criança demonstre desconforto, a sessão será terminada.

Os dados pessoais constantes nesta ficha de autorização serão tratados de forma confidencial, sendo registados na base de dados do Projeto Robótica-Autismo da Universidade do Minho para posterior utilização na investigação no âmbito do Doutoramento de Vinícius Silva. Os dados fornecidos não serão partilhados com terceiros, sendo conservados durante o tempo necessário para a finalidade indicada ou até que o titular manifeste vontade em contrário. Os titulares dos dados podem, a todo o momento, aceder à informação que lhes respeite e solicitar por escrito a sua atualização, correção ou eliminação.

_____ de _____ de 2021

_____

(Assinatura Completa do Encarregado de Educação e/ou tutor)

Este documento é composto de 2 páginas e feito em duplicado: uma via para o/a investigador/a, outra para a pessoa que consente.

ANEXO VIII - Exemplar do consentimento informado e da informação fornecida aos profissionais que diariamente acompanham as crianças

____ de _____ de 2021

Exmo. Senhor (a) Professor (a),

O **projeto Robótica-Autismo** (http://robotica-autismo.dei.uminho.pt/) visa a aplicação de ferramentas robóticas como forma de melhorar a vida social de crianças e jovens com **perturbação do espectro do autismo (**PEA). Em particular, o objetivo é **melhorar as habilidades de interação e comunicação com o meio-ambiente e com outras pessoas**.

Em particular, no âmbito de doutoramento de Vinícius Corrêa Alves da Silva pretendemos investigar a capacidade da ferramenta robótica potencializar o reconhecimento de emoções e expressões faciais.

Uma vez que as sessões de trabalho implicarão a participação do seu aluno, tal só será possível com a sua autorização. Garantimos que os dados recolhidos serão apenas utilizados nesta investigação e divulgação científica da mesma, nunca explorando os mesmos noutras situações que não estejam relacionadas com o projeto. Solicitamos, assim, a sua colaboração dando o seu consentimento através da devolução do destacável anexo.

Agradecendo desde já a sua participação, com os melhores cumprimentos,

A Coordenadora Científica do Projeto e orientadora, UM

_____

Filomena Maria da Rocha Menezes de Oliveira Soares

O aluno de Doutoramento responsável pelo estudo, UM

_____

Vinícius Corrêa Alves da Silva

Contactos: roboticaautismo@gmail.com ou a65312@alunos.uminho.pt; Vinícius Silva: 917672420

Eu_____ professor(a) do(a) aluno (a) _____ declaro ter compreendido os objetivos de quanto me foi proposto e explicado pelo investigador que assina este documento, ter-me sido dada oportunidade de fazer todas as perguntas sobre o assunto e para todas elas ter obtido resposta esclarecedora, ter-me sido garantido que não haverá prejuízo para os direitos assistenciais se eu recusar esta solicitação, e ter-me sido dado tempo suficiente para refletir sobre esta proposta.

Declaro também que autorizo o meu (minha) aluno(a) a participar no Projeto de Investigação "Ferramentas Robóticas na Educação Especial", em particular no "Jogo de Emoções com o Zeca", o que implicará a realização de experiências de 10 a 20 minutos com a supervisão de um adulto. As experiências serão filmadas para posterior análise de desempenho do educando. As filmagens e os dados decorrentes destas serão utilizadas única e exclusivamente na divulgação científica do projeto. Os dados pessoais e os dados obtidos na investigação não serão divulgados e serão mantidos por um período de dez anos, ao fim do qual serão destruídos. Todas as informações de caráter pessoal recolhidas no decurso da investigação serão consideradas confidenciais e tratadas de acordo com as regras relativas à proteção de dados e da vida privada. Se o encarregado de educação o entender, o aluno pode abandonar o projeto em qualquer altura. A sua participação, e/ou a do seu aluno, ou recusa em participar ou posterior abandono, não prejudicarão a relação com a equipa de clínicos e/ou investigadores. Não se preveem quaisquer riscos para os participantes durante as sessões. Caso a criança demonstre desconforto, a sessão será terminada.

Os dados pessoais constantes nesta ficha de autorização serão tratados de forma confidencial, sendo registados na base de dados do Projeto Robótica-Autismo da Universidade do Minho para posterior utilização na investigação no âmbito do Doutoramento de Vinícius Silva. Os dados fornecidos não serão partilhados com terceiros, sendo conservados durante o tempo necessário para a finalidade indicada ou até que o titular manifeste vontade em contrário. Os titulares dos dados podem, a todo o momento, aceder à informação que lhes respeite e solicitar por escrito a sua atualização, correção ou eliminação.

_____ de _____ de 2021

_____

(Assinatura Completa do Professor)

Este documento é composto de 2 páginas e feito em duplicado: uma via para o/a investigador/a, outra para a pessoa que consente.

A.5 Informed consent for the participants of the Focus Group and SSIs

CONSENTIMENTO INFORMADO, ESCLARECIDO E LIVRE

Eu, Vinícius Silva, doutorando do Programa Doutoral em Engenharia Eletrónica e de computadores, da Escola de Engenharia da Universidade do Minho, encontro-me neste momento a realizar um estudo científico, sob a orientação da Doutora Filomena Soares da Universidade do Minho.

O presente estudo tem por finalidade estudar o melhor conceito que um Playware, com diferentes configurações, deve possuir para ser utilizado no apoio a crianças com Perturbação do Espectro do Autismo.

Para a concretização deste estudo é necessária a sua colaboração num Grupo de Discussão. É ainda solicitada a sua anuência para a gravação em áudio e vídeo da respetiva sessão.

Todas as informações e resultados serão confidenciais e, por isso, não existirá a possibilidade de identificação, sendo que o seu nome e dados pessoais não serão incluídos no estudo.

Disponibilizo-me para esclarecimentos e informações que considere pertinentes, através do contacto:a65312@alunos.uminho.pt.

Li e concordo participar no estudo e autorizo o Vinícius Silva a usar os dados para fins de investigação.

Data:__/__/_____ Assinatura:

_____

CONSENTIMENTO INFORMADO, ESCLARECIDO E LIVRE

Eu, Vinícius Silva, doutorando do Programa Doutoral em Engenharia Eletrónica e de computadores, da Escola de Engenharia da Universidade do Minho, encontro-me neste momento a realizar um estudo científico, sob a orientação da Doutora Filomena Soares da Universidade do Minho.

O presente estudo tem por finalidade estudar o melhor conceito de adaptação do comportamento do robô as possíveis ações das crianças com Perturbação do Espectro do Autismo durante uma sessão de interação Humano-robô.

Para a concretização deste estudo é necessária à sua colaboração numa entrevista semiestruturada. É ainda solicitada a sua anuência para a gravação em áudio e vídeo da respetiva sessão.

Todas as informações e resultados serão confidenciais e, por isso, não existirá a possibilidade de identificação, sendo que o seu nome e dados pessoais não serão incluídos no estudo.

Disponibilizo-me para esclarecimentos e informações que considere pertinentes, através do contacto: a65312@alunos.uminho.pt.

Li e concordo participar no estudo e autorizo o Vinícius Silva a usar os dados para fins de investigação.

Data:__/__/_____ Assinatura:

_____

A.6 Questionnaire to characterize the child

ANEXO V - Exemplar do questionário para recolha de dados

Identificação da pessoa que responde ao questionário:

Nome: Click or tap here to enter text.

Habilitação Académica: Click or tap here to enter text.___(Licenciatura, Mestrado, Doutoramento)
Categoria profissional: Click or tap here to enter text.(Prof. Ed. Especial, Terapeuta da Fala, etc.)
Horas de contacto semanal com a criança: Click or tap here to enter text.horas

---

O projeto Robótica-Autismo (http://robotica-autismo.dei.uminho.pt/) visa a aplicação de ferramentas robóticas como forma de melhorar a vida social de crianças e jovens com perturbação do espectro do autismo (PEA). Em particular, o objetivo é melhorar as habilidades de interação e comunicação com o meio-ambiente e com outras pessoas. Em particular, no âmbito de doutoramento de Vinícius Corrêa Alvesda Silva pretendemos investigar a capacidade de a ferramenta robótica potencializar o reconhecimento de emoções e expressões faciais.

Identificação da Criança:

Nome (primeiro e último): Click or tap here to enter text.

Data de Nascimento (aaaa/mm/dd): Click or tap to enter a date.

Escola: Click or tap here to enter text.

Caracterização da Criança:

Diagnóstico dentro da Perturbação do Espectro do Autismo (nível de severidade, ...):Click or tap here to enter text.

Tipo de recompensa/reforço preferido pela criança (quando os professores pretendem estimular acriança depois de uma resposta correta)

- □ Verbal (por exemplo, "Muito bem, João!"); qual?Click or tap here to enter text.
- □ Movimento (p.e., agitar os braços no ar simulando vitória); qual? Click or tap here to enter text.
- □ Som (p.e, xilofone), qual? Click or tap here to enter text.
- □ Verbal + Movimento qual?Click or tap here to enter text.
- □ Verbal + Som; qual? Click or tap here to enter text.
- □ Som + Movimento; qual? Click or tap here to enter text.
- □ Verbal + Movimento + Som Click or tap here to enter text.
- □ Outro; qual? Click or tap here to enter text.
- □ Experiência anterior no desenvolvimento da competência de reconhecimento de emoções; Tipos de iniciativas utilizadas (p.e. jogo de correspondência de cartões, jogos virtuais, ...), quais?

Click or tap here to enter text.

Avaliação subjetiva da criança:

Numa escala de 1 a 5 (em que **1 - Nunca, 2 - Raramente, 3 - Ocasionalmente, 4 - Frequentemente e 5 - Sempre; NA - Não aplicável**), por favor classifique, na sua opinião, o seguinte conjunto de afirmações, colocando uma cruz na coluna correspondente:

Afirmação	1	2	3	4	5	NA	Comentários
A criança tem a capacidade de perceber o ponto de vista do outro.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança está consciente do seu próprio corpo e identidade enquanto brinca.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança reconhece a presença de diferentes companheiros de jogo.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança brinca sozinha sem se preocupar com as atividades das pessoas à sua volta.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança observa os seus colegas a brincar (p.e., observa as ações de jogo de outros) mas não se junta a essas atividades.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança reage quando uma ou mais crianças estão à sua volta.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.

A criança brinca separadamente em vez de brincar com outras crianças, com brinquedos e materiais similares.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança segue os movimentos das outras crianças com o seu olhar.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança tenta fazer contacto social com outras crianças.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança junta-se a outras crianças em atividades com objetos, brinquedos ou materiais com um objetivo comum.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança não compreende porque algumas situações aborrecem tanto outras pessoas.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança não chora ou fica perturbada se um personagem de um filme morre.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança é rápida a perceber quando as pessoas estão a brincar (ironizar).	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança é franca a dar as suas opiniões, mesmo quando estas perturbam alguém.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança é muitas vezes rude ou indelicada sem perceber.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança mostra preocupação quando os outros estão tristes.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança pode parecer tão preocupada com os seus próprios pensamentos que não percebe que os outros estão aborrecidos.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança fica irritada quando as atividades não são feitas na hora.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança fica perturbada se vê um animal com dor.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança empurra ou belisca se alguém a estiver a aborrecer.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
Se outra criança não é convidada para uma festa, a criança preocupa-se como ela se sente.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.

A criança fica incomodada ao ver outros a chorar ou com dor.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.
A criança ajuda as novas crianças a se integrarem na sala de aula.	☐	☐	☐	☐	☐	☐	Click or tap here to enter text.

Comentários sobre esta criança: Click or tap here to enter text.

A.7 Ethics Committee Documents



Universidade do Minho

Conselho de Ética

Conselho de Ética

Comissão de Ética para a Investigação em Ciências da Vida e da Saúde (CEICVS)

1. Identificação do documento: SECVS 028/2014 (ADENDA)

2. **Título do projeto:** *Projeto robótica-autismo*

3. Equipa de investigação: Dra. Filomena Maria da Rocha Menezes de Oliveira Soares, da Escola de Engenharia, Departamento de Eletrónica Industrial, Centro de Investigação Algoritmi, Universidade do Minho; Dra. Sandra Cristina Cunha Costa, Escola de Engenharia, Departamento de Eletrónica Industrial, Centro de Investigação Algoritmi; Dra. Cristina Manuela Peixoto dos Santos - Escola de Engenharia, Departamento de Eletrónica Industrial, Centro de Investigação Algoritmi; Dra. Ana Paula da Silva Pereira, Instituto da Educação, Centro de Investigação dos Estudos da Criança, Universidade do Minho

4. Unidade Orgânica Promotora: Centro Algoritmi, Universidade do Minho

5. Outras Unidades: Escolas do primeiro ciclo do ensino básico com unidades de ensino estruturado para alunos com Perturbações do Espectro do Autismo, Centros de desenvolvimento e Instituições que prestam apoios a estes alunos

6. Âmbito e objetivos do projeto: De acordo com a documentação apresentada, o projeto intitulado "Projeto robótica-autismo", trata-se de um pedido de adenda ao projeto "Projeto robótica-autismo" com prévia análise e emissão de Parecer pela Subcomissão de Ética para as Ciências da Vida e Saúde (SECVS) da Universidade do Minho (SECVS 028/2014) a 3 de março de 2014, cujo objetivo geral prende-se com utilizar a robótica social como promotora do desenvolvimento sócio emocional em crianças com perturbações do espectro do autismo (PEA). Eram objetivos específicos do projeto promover o reconhecimento de expressões faciais e as emoções correspondentes, fomentando interações sociais e de comunicação, através do uso de um robô humanoide "ZECA" (*Zeno Engaging Children with Autism*) que irá interagir em contexto individual com as crianças. Trata-se de um estudo intervencional e longitudinal, sendo a população-alvo crianças com Perturbações do Espectro do Autismo (PEA) de alto funcionamento, entre 6 e 10 anos de idade (num tamanho amostral previsto de n = 45, dividido em 3 grupos: G1 - Grupo Experimental, G2 - Grupo

de Controlo 1 e G3 - Grupo de Controlo 2; sendo que não existem ainda estudos de prevalência em Portugal). A colaboração das Unidades de Saúde Familiar consiste apenas em permitir a aplicação do questionário apresentado em anexo, de forma a recolher as informações pretendidas.

1. Justificação de alteração a protocolo inicial

Trata-se de um pedido de alteração/adenda ao protocolo inicial do estudo "Projeto robótica-autismo" que foi alvo de Deliberação da SECVS, a 3 de março de 2014.

Foi submetido sumário do protocolo de investigação, com pedido de extensão do projeto até 31 de janeiro de 2022, sendo resumidas as seguintes alterações ao protocolo inicial:

· Investigadores

Equipa de Investigação: Vinícius Corrêa Alves da Silva - Escola de Engenharia, Departamento de Eletrónica Industrial, Centro de Investigação Algoritmi; João Miguel Clemente Sena Esteves - Escola de Engenharia, Departamento de Eletrónica Industrial, Centro de Investigação Algoritmi; Ana Paula da Silva Pereira - Instituto da Educação, Centro de Investigação em Educação; Sandra Queirós, Colaboradora externa.

· Autorizações

O projeto de investigação foi submetido junto da DGEstE DSRN (número do pedido de registo 2187/2014) e foi aprovado a 3 junho de 2014 (autorização 5344/2914).

Autorização CNPD de 3 de junho de 2014 (autorização 2187/2014).

· Procedimento

A ferramenta utilizada nesta investigação consiste num robô humanoide, ZECA (Zeno Engaging Children with Autism) cujas especificações estão definidas no novo protocolo de investigação submetido e em http://www.hansonrobotics.com/. Este robô vai interagir em contexto individual com as crianças. Está agora disponibilizado um website com informação sobre o projeto (http://robotica-autismo.dei.uminho.pt/).

· Metodologia

Foi sumarizado o procedimento, que segue o previamente descrito e se resume: A amostra será dividida em 3 grupos: G1 - Grupo Experimental, G2 - Grupo de Controlo 1 e G3 - Grupo de Controlo 2. Todos os grupos realizam, no início e no fim do período experimental, um pré- e um pós-teste para comparar as suas competências antes e depois da intervenção. A tarefa do pré- e pós-teste foi definida com apoio de profissionais com experiência na área, e consiste

na correspondência de uma imagem Picture Exchange Communication System com a fotografia de uma pessoa desconhecida, tendo em conta a expressão facial que representam. O Grupo G1 realiza três tarefas com o robô, que têm como objetivo o desenvolvimento da competência do reconhecimento de expressões faciais correspondendo a 5 emoções: alegria, tristeza, surpresa, medo e raiva. O grupo G2 realiza as mesmas tarefas sem a intervenção do robô. O grupo G3 realizará apenas a tarefa de pré- e pós-teste. As crianças estão envolvidas num total de 6 sessões, duas vezes por semana, em contexto individual com o robô e o experimentador, numa sala de aula e durante 10 a 20 minutos. Os pais não estão presentes durante as sessões, mas pelo menos um profissional que acompanha regularmente a criança está disponível para colaborar caso seja necessário, devido à especificidade dos alunos com espectro envolvidos na amostra deste estudo. Durante todo o projeto de investigação são recolhidos os seguintes dados: informação sobre as crianças utilizando um questionário administrado a um dos profissionais que acompanha diariamente a criança e vídeos das sessões.

· População-Alvo

Mantem-se a população alvo e os critérios de inclusão e exclusão que se sumarizam: os critérios de inclusão na amostra incluem crianças com Perturbação do Espectro do Autismo (PEA) de nível 1, entre 5 e 10 anos, com o consentimento dos pais e professores para participação no estudo e que estejam incluídas em contexto escolar (possivelmente também numa unidade estruturada de apoio ao autismo). Os critérios de exclusão da amostra são crianças com PEA com a comorbidade das dificuldades intelectuais.

- Financiamento

Este projeto é financiado pela Fundação Ciência e Tecnologia através da bolsa de doutoramento atribuída ao investigador Vinícius Corrêa Alves da Silva, cuja referência é SFRH/BD/133314/2017.

O Documento de Consentimento Informado foi revisto de forma a seguir as recomendações do documento guia a ARS Norte e do Documento CEIC sobre o Regulamento Geral de Proteção de Dados (RGPD) no contexto da Investigação Clínica" de 17 de outubro de 2018 que incide sobre procedimentos a adotar para o cumprimento do estabelecido no RGPD, no contexto da investigação clínica em geral, embora que mais especificamente sobre a informação a constar no Consentimento Informado (CI) relativo a matéria de confidencialidade e proteção de dados.

Assegurou-se que os compromissos de boas práticas na investigação dadas pelo IR do estudo e por outros investigadores ou colaboradores na investigação, incluindo garantias de confidencialidade, continuarão a ser seguidas.

Apresentou-se em anexo o protocolo de investigação com fundamentação teórica e metodológica para as alterações ao protocolo original supracitadas.

1. Parecer

Após verificação e análise dos documentos associados ao processo de pedido de emissão de parecer ético sobre o projeto em apreço, a que reporta sumariamente a "Grelha de verificação e avaliação ética" acima apresentada, considera-se que (i) o processo está devidamente instruído, (ii) a análise dos documentos apresentados sobre o estudo a realizar obedecem às regras de conduta ética e requisitos exigidos para as boas práticas na experimentação com humanos e (iii) estão em conformidade com o Guião para submissão de processos a pedido de Parecer Ético na UMinho.

Relatoras: Nadine Santos (Investigadora doutorada/EM e Centro Clinico Académico - HB)

Face ao exposto, A CEICVS nada tem a opor à realização do projeto, emitindo o seu parecer favorável.

Braga, 6 de maio de 2019.

A Presidente da CEICVS

MARIA CECÍLIA DE LEMOS PINTO ESTRELA LEÃO

Assinado de forma digital por MARIA CECÍLIA DE LEMOS PINTO ESTRELA LEÃO
Dados: 2019.05.13 09:46:02 +01'00'

CNPD

Referência: **FORM-D.20200609105705.4Z37VN**

Exmo Senhor Vinícius Silva,
Em resposta à sua exposição, informamos que, desde 25 de maio de 2018, com a aplicação do Regulamento Geral sobre a Proteção de Dados (Regulamento UE 2016/679) - RGPD, a Comissão Nacional de Protecção de Dados - CNPD deixou de ter poderes para autorizar tratamentos de dados pessoais.

Para compensar a ausência de controlo prévio da CNPD, o RGPD faz recair a obrigação sobre quem realiza os tratamentos de dados pessoais de verificar se o tratamento está conforme com as regras e princípios do RGPD (cf. artigo 24.º do RGPD). Para o efeito, deverá, enquanto responsável pelo tratamento, considerar as normas do artigo 6.º e do artigo 9.º do RGPD (em função das categorias de dados pessoais objeto de recolha, sendo que se algum dos dados recolhidos corresponder a uma das categorias de dados elencados no n.º 1 do artigo 9.º, só com base numa das condições estabelecidas no n.º 2 desse artigo é que poderá tratar os dados). Em especial, se o fundamento que legitima o tratamento for o consentimento, sublinha-se que o consentimento deve ser uma declaração informada e específica, relativa ao concreto tratamento de dados pessoais, não se confundindo com o consentimento para a sujeição ou participação no estudo. Chama-se a atenção para o direito de informação previsto no artigo 13.º do RGPD.

Deve ainda verificar se o tratamento cumpre os princípios previstos no artigo 5.º do RGPD, em especial, o princípio da minimização dos dados e da indispensabilidade da sua recolha, e eventualmente, em função da natureza e âmbito do tratamento, considerar a obrigação de realizar uma avaliação prévia do impacto sobre a proteção de dados (prevista no artigo 35.º do RGPD).

Com os melhores cumprimentos
CNPD

mime.noreply@min-educ.pt

Para: Vinícius Corrêa Alves da Silva; Vinícius Corrêa Alves da Silva

sex, 09/07/2021 16:11

Exmo(a)s. Sr(a)s.

O pedido de autorização do inquérito n.º 0741100002, com a designação *Projeto Robótica Autismo*, registado em 28-06-2021, foi aprovado.

Avaliação do inquérito:

> Exmo.(a) Senhor(a) Vinícius Corrêa Alves da Silva
> Venho por este meio informar que o pedido de realização de inquérito em meio escolar é autorizado uma vez que, submetido a análise, cumpre os requisitos, devendo atender-se às observações aduzidas.
> Com os melhores cumprimentos
> José Vitor Pedroso
> Diretor-Geral
> DGE

A.8 FCT Statement – exceptional COVID-19 impact mitigation grant call

Caro/a Bolseiro/a,
Vinicius Corrêa Alves da Silva
SFRH/BD/133314/2017

Na sequência do requerimento apresentado ao abrigo do regulamento 560-A/2021, e após análise do processo por parte da Comissão de Avaliação designada para o efeito, comunicamos a decisão sobre o mesmo.

Parecer de Avaliação:

De acordo com a documentação apresentada, os trabalhos no âmbito deste doutoramento tiveram início a 01/10/2017. Segundo o parecer emitido pelos orientadores, em agosto de 2020, para pedido de prorrogação de 12 meses, já se encontravam previstas medidas de contingência que permitiriam adaptar o plano de trabalho às restrições causadas pelo covid-19, nomeadamente através da substituição de sessões de teste alargadas por estudos de caso. Nesse mesmo parecer é referido já existirem publicações de âmbito internacional e que estavam a ser preparadas outras para submissão em revistas internacionais (no requerimento, o bolseiro menciona 7 artigos em capítulo de livro, 6 artigos em conferência e 1 em revista internacional). É ainda indicado existir um atraso motivado por não ter sido antecipada a necessidade de um maior volume de dados para a utilização de metodologias de deep learning, não sendo este atraso imputável às restrições covid-19. De acordo com os termos de referência para atribuição de bolsa excecional, é necessário existir "grave prejuízo provocado nas atividades de investigação pela suspensão de atividades presenciais e outras limitações provocadas pela pandemia da doença covid-19". Contudo, neste caso não é evidente que tenha existido grave prejuízo e que, consequentemente, a obtenção do grau de doutor tenha sido comprometida.

Face ao exposto, propõe-se a não concessão de bolsa ao abrigo do Regulamento n.º 560-A/2021.
Requerimento Aprovado: Não

Com os melhores cumprimentos,
Helena Pereira

Presidente do Conselho Diretivo
President of the Board of Directors
FCT – Fundação para a Ciência e a Tecnologia
Av. D. Carlos I, 126, 1249-074 Lisboa, Portugal
T: [+351] 213 924 377/84/85
presidencia@fct.pt | www.fct.pt