Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Facial expression analysis in children with autism spectrum disorder using a refined Human-Robot-Game platform for active learning

Nayeth Idalid Solórzano Alcívar, Dennys Fabián Paillacho Chiluiza, Michael Xavier Arce Sierra, Anthony Jair Pincay Lino & Edwin Andrew Eras Zamora

To cite this article: Nayeth Idalid Solórzano Alcívar, Dennys Fabián Paillacho Chiluiza, Michael Xavier Arce Sierra, Anthony Jair Pincay Lino & Edwin Andrew Eras Zamora (2025) Facial expression analysis in children with autism spectrum disorder using a refined Human-Robot-Game platform for active learning, Behaviour & Information Technology, 44:13, 3152-3164, DOI: 10.1080/0144929X.2024.2434896

To link to this article: https://doi.org/10.1080/0144929X.2024.2434896

Published online: 06 Dec 2024.

BEHAVIOUR & INFORMATION TECHNOLOGY
2025, VOL. 44, NO. 13, 3152–3164
https://doi.org/10.1080/0144929X.2024.2434896

Taylor & Francis
Taylor & Francis Group

Check for updates

Facial expression analysis in children with autism spectrum disorder using a refined Human-Robot-Game platform for active learning

Nayeth Idalid Solórzano Alcívar, Dennys Fabián Paillacho Chiluiza, Michael Xavier Arce Sierra, Anthony Jair Pincay Lino and Edwin Andrew Eras Zamora

ESPOL, ESPOL Polytechnic University, Guayaquil, Ecuador.

ABSTRACT

The use of serious video games and robotics in children's education and therapy is rapidly increasing. Innovative Human-Robot-Game (HRG) platforms are particularly promising for assessing learning and behaviour in children with autism spectrum disorder (ASD), who often face challenges with attention, communication, and socialisation. Research shows that when children with ASD engage with social robots connected to interactive educational games, their learning, attention, and communication skills improve. This study aims to monitor the psychosocial and cognitive progress of children with ASD using an HRG platform, refining it as a tool for active learning. The article focuses on enhancing the facial recognition metrics of the HRG platform, which is identified as LOLY-MIDI and designed to measure attention and emotions in children while playing serious video games with the assistance of a social robot. Using a mixed-methods experimental approach strategy, reviewing previous studies, and employing tools like OpenFace and the Facial Action Coding System (FACS) for facial recognition, the research achieved greater precision in designing these metrics, providing accurate measurements of attention and emotional responses. The findings offer valuable insights for psychology and education professionals in assessing the socio-educational progress of children with ASD.

ARTICLE HISTORY
Received 19 January 2024
Accepted 23 November 2024

KEYWORDS
Special needs; inclusive education; HCI; serious games; facial recognition; video games

1. Introduction

Several educational centers are developing academic innovations using Information and Communication Technology (ICT) tools (Turrado-Sevilla and Cantón-Mayo 2022). This type of innovation is the case of digital games and social robotics use, which has advanced exponentially and is important in different interdisciplinary areas, such as neurorehabilitation (Hoffmann and Krämer 2021). Also, for their assistance in therapies and home environments, as indicated in their studies by Jeong et al. (2018), as well as companion robots, according to de Wolf and Li (2020), or even to interact with people (Kubota et al. 2020). What is more, several educational institutions are investigating new processes to treat children with autism spectrum disorder (ASD) through human-robot-games (HRG) systems, considering that these children are more attracted to this type of technological resources considered no intrusive multimedia. The HRG systems may include social interaction, considering virtual facial expressions that convey emotional interaction, improved immersive guidance and evaluation (Ji et al. 2023), and helping experts understand their behaviour in different assessments.

Children with ASD have social interaction problems, communication issues, and short focus time that leads to less concentration. Multi-use scenarios technologies such as social robots, mobile learning, serious games, virtual reality, edutainment, virtual learning environment, and augmented reality keep individuals interested in learning and maintaining focus. Another key factor is the usability and design of easier applications for ASD children because autistic people are more visual learners (Khan et al. 2019). In this sense, as indicated by Kuutti et al. (2022), the multi-use scenarios that involve growth in the development of human-robot interaction (HRI) help the potential benefit of autonomous intervention by the robot for obtaining information about the behaviour and assessment of the progress of participating individuals. Also, these scenarios highlight the importance of analysing non-verbal interactions such as gestures or facial expressions.

On the other hand, video games used as educational support tools help to improve cognitive skills, attention, and academic performance (Amukune, Barrett, and Józsa 2022), whose use can be complemented with robotic accompaniment to further stimulate the interest

CONTACT Nayeth Idalid Solórzano Alcívar nsolorza@espol.edu.ec ESPOL Polytechnic University, ESPOL, Campus Gustavo Galindo, Km. 30.5 Vía Perimetral, Guayaquil 090902, Ecuador.
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

of neurotypical or neurodivergent children, with innovative resources for learning through play.

An example of this type of complemented technology use of video games and social robotics is the case of the LOLY-MIDI studies linked to the MIDI-AM series game (acronym of Multimedia Interactive Didactic for Children – Mobile Applications). Within this study developed in several stages, a human-robot-game (HRG) platform, including visual monitoring, has been developed to analyse expressions for the evaluation of emotions and attention levels of children while they are playing with educational videogames and interacting with a social robot (Solorzano Alcivar et al. 2021). The LOLY-MIDI research has been mainly addressed for children with autism spectrum disorder (ASD), as the emphasis population chooses to test the designed HRG platform. The evaluation of the HRG platform developed in previous stages of this study is the technological focus tool aimed to be shaped and continually tested until refined for greater precision. The expectation is to obtain accurate values in the report outcomes as essential insights for professionals in psychology and education to assess ASD children's socio-educational progress.

Overall, the current investigation aims to monitor the psychosocial and cognitive progress of children with ASD using the designed LOLY-MIDI HRG platform to be tested and refined as a tool useful for the active learning process. Initially, progressive technical evaluation for analysing and determining how to refine specific facial recognition metrics of a designed HRG platform that helps to measure users' degrees of attention and emotions is discussed. Then, proposed technical changes in metrics, evaluation procedures, and tools are applied until accurate results are obtained.

2. Revised literature and previous studies

This research addresses an experimental study that analyses facial expressions and assesses the level of attention metrics, particularly in ASD children. Thus, revising previous studies, stating definitions, and explaining complementary tools, such as OpenFace and FACS, were required to be considered as part of the study process.

2.1. Children with ASD linked to the technology

Children with ASD manifest three main characteristics that are represented in social interaction difficulties, communication, and behaviour, so their learning is different compared to neurotypical children. These difficulties in ASD children can be effectively addressed early with the integrity of computer technology strategies rather than conventional assessments because it has more predictable and less intrusive multimedia content on electronic devices (Ji et al. 2023). The studies of Khan et al. (2019) indicate that different ICT tools can be used for assessment and intervention to enhance functionality and cognitive and social abilities in individuals with ASD. For example, serious games, also known as educational video games, are based on a learning approach that supports theoretical, curricular content or vocabulary for learning in autistic children.

The studies of Chinchay et al. (2024) indicate that nowadays, ICT tools seem to show great advantages for people with ASD, such as the use of mobile devices and, within these, the development of web or mobile applications (app), allowing the production of special interactive games oriented to motivate the attention and learning of children with this spectrum. Many of these apps have been adapted to similar or alternative forms of virtual teaching and adopted emergently during the COVID-19 pandemic. Using serious games as a methodology, already known as _Learning by Playing_ has greatly impacted the development of knowledge and the orientation of social behaviours (Shohieb et al. 2022). Complementarily, it is estimated that the use of these mobile games' applications in a controlled manner contributes to the treatment of children with ASD, seeking to identify their impact in evaluations carried out by psychologists, psychotherapists, or educators.

Conversely, developing social skills, functional adaptive skills, or reducing disruptive behaviour are early childhood teaching goals for children with ASD (Myers et al. of Pediatrics Council on Children with Disabilities 2007). The teaching interventions emphasise the diagnosis of understanding and generation of nonverbal cues as an essential tool for social interaction (Drimalla et al. 2021). In this context, it has been studied how non-human social support can interact and positively influence socio-emotional development, knowing that autistic children enjoy and can benefit from activities that do not require social interaction (Syrjämäki et al. 2023). In this sense, Dubois-Sage et al. (2024) assert that children with ASD are more interested in robots than humans, as they attribute human characteristics to robots. Other cases show pet therapies are one of the most influential examples in this area. Works such as (Jeong et al. 2018) and (Hamzah et al. 2014) incorporate social robotics, obtaining positive results in the collective development of the child.

Overall, different technology tools and mechanisms have been used in autism for treatment and education. This technology includes social robots, serious games, mobile learning, virtual or augmented reality, and virtual learning environments (Khan et al. 2019).

2.2. Social robots studies in ASD individuals

It has been evidenced that children with ASD relate better with social robot models with autonomous behaviour as their peers and non-autonomous robots without movement or feedback because it reduces their anxiety and have less disruptive behaviours like humans (Hamzah et al. 2014).

In robotic research, there is a diversity of literature on the potential uses of social robots to help neurotypical or neurodivergent children. Some of these studies are particularly focused on children with ASD (see Table 1).

2.3. The HRG platform LOLY-MIDI

The LOLY-MIDI studies include the design of an HRG platform with an architecture based on the SAR concept used for therapeutic interventions and educational purposes. The designed platform allows the analysis of facial expressions and assesses attention metrics by monitoring and measuring the level of attention and emotion (Solorzano Alcivar et al. 2021). The main components of the HRG platform are detailed in Figure 1, which includes robot, educational game, and cloud services for data processing:

The data are extracted from videos collected from a social robot named 'Loly', shaped as a robotic bust developed by the researcher team. This robotic bust was designed to evaluate human-robot interaction while children play a serious game application. The Loly robot contains a speaker, head, wing with movements, and a face-level screen displaying images of expressing eyes for the non-verbal language. The updated version of this robotic prototype has two cameras at the forehead or at a bust level to capture images from users. Loly is linked to children's educational digital games from a series MIDI-AM, which contains mobile interactive applications for primary education, produced as part of this research component. The mobile video games are designed to understand the guidance of Loly, who follows the instructions and tries to catch the attention of the child with ASD (Solorzano Alcivar et al. 2022).

The LOLY-MIDI platform also contains the control board MIDI-API dashboard_ developed as a supporting and monitoring tool that records the progress and the behaviour of children with ASD while playing. This platform facilitates the reading and interpretation of a set of emotions and the level of attention metrics aimed at psychologists and psychotherapists working directly with children with ASD. The dashboard receives information on a set of variables related to facial point coding, collected through videos recorded by a fisheye camera on Loly's forehead or by the camera on Loly's bust. The open-source tool for facial analysis, OpenFace 2.0, measures degrees of attention according to face and head positions (Baltrusaitis et al. 2018).

2.4. Tools used for emotional analysis

The interpretation of nonverbal signals, such as facial expressions, is of special interest in children with ASD. The therapists could develop new personalised interventions by creating immersive experiences for children with ASD (Alhakbani 2024). OpenFace 2.0 is used for image and video processing, incorporating computer vision and machine learning (ML) to perform measurements and analyse emotions. It allows for

Table 1. Scientific studies in social robotics.

Authors	Description
Kubota et al. (2020)	JESSIE is a robotic system designed to programme social robots under high-level specifications. It has three variable states: by objectives (monitoring, educational, or service delivery), by type of interaction (active or passive), or by research context. Supported by its reproducibility, this work allows the community to customise the robot's behaviour, promoting a longitudinal HRI.
Mejía Silva and Arroba (2016); Espinoza Erazo and Gallardo (2018)	CODY is an Ecuadorian social robot developed for human-computer interaction. It has a design focused on educating children from two to five years old, was developed under the VDI-2206 standard, and changed its name to NAR.
Jeong et al. (2018)	HUGGABLE is a robot designed to provide socioemotional support for patients and their families during hospital care. It engages the pediatric patient in playful interactions.
Westlund et al. (2016)	TEGA, a social robot designed for playful learning implementable in various environments, simulates language, behaviour, and attitudes for interaction with children.
Auliawan and Ong (2020)	MUSIO, a robot with artificial intelligence focused on English language teaching, works through direct conversation and can recite and listen using didactic elements.
Pakkar et al. (2019)	KIWI was developed under a design and systems intended for long-term analysis of the robot to complement home therapy for children with ASD.
N. I. Solorzano Alcivar et al. (2021)	LOLY-MIDI, A human-robot-game (HRG) platform using educational videogames and social robotics with visual monitoring and metrics to analyse expressions for evaluating emotions and attention levels in children with ASD.

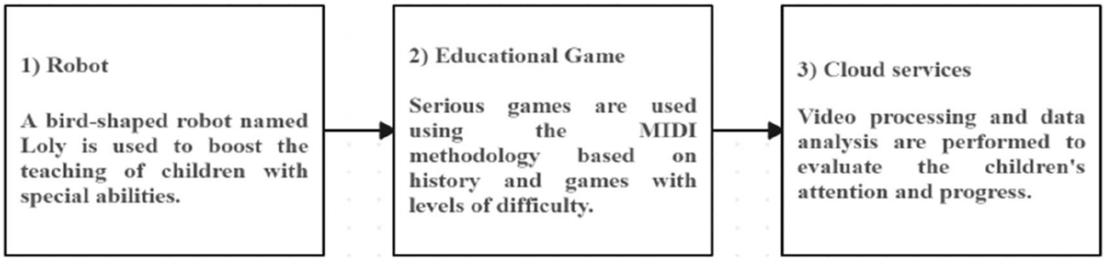

Figure 1. Main components for the HRG platform.

analysing people's facial behaviours depending on four fields: location of facial landmarks, head position, gaze, and facial expressions (Baltrusaitis et al. 2018).

The OpenFace analysis process registers the following sequence: input frame, facial detection, 3D facial landmark location, gaze estimation, head pose estimation, facial alignment extraction, normalisation of the person's fusion characteristics, and AU facial recognition (see Table 2). At the end of this process, a CSV (comma-separated values) format file is generated with the results of the video analysis, of which the ayes gaze, head pose, and facial points are highlighted in head location and facial landmarks, as it is shown in Figure 2.

The Facial Action Coding System (FACS) is a system for categorising human facial movements by their gestures on the face. FACS codes movements of individual facial muscles from slight instantaneous changes in facial appearance. Using FACS, it is possible to encode almost any anatomically generated facial expression by deconstructing it into the specific _Facial Action Units (Aus)_ that produced the expression (Baltrusaitis et al. 2018). Aus or Action Units (AU) is a common standard for objectively describing facial expressions of which OpenFace can recognise a subset, specifically (see Table 2):

- Frame: Frame number in case of videos.
- Face_id: In the case that the face of several people is detected.
- Timestamp: The time it takes for a video to be processed.
- Confidence: How reliable is the estimation of facial landmarks?
- Success: If a person's face could be detected in the input box (= 1, 0 otherwise).
- Gaze_0_x, gaze_0_y, gaze_0_z: Estimated gaze vector concerning the left eye.
- Gaze_1_x, gaze_1_y, gaze_1_z: Estimated gaze vector concerning the right eye.
- Pose_Tx, pose_Ty, pose_Tz: The location of the head concerning the camera is in millimetres.
- Pose_Rx, pose_Ry, pose_Rz: Head rotation in radians around the X, Y, and Z axes.
- AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r, AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, AU45_r: Detects the intensity of the AU's

Table 2. Action units (Baltrusaitis et al. 2018).

AU	Description	Illustration
AU01	Raise inner eyebrows	
AU02	Raise outer eyebrows	
AU04	Lower eyebrows	
AU05	Lift upper eyelid	
AU06	Lift lower eyelid	
AU07	Tight eyelids	
AU09	Nose wrinkle	
AU10	Lift upper lip	
AU12	Removing lip corners	
AU14	Dimples	
AU15	Lower lip corners	
AU17	Chin lift	
AU20	Stretched lips	
AU23	Tight lips	
AU25	Split lips	
AU26	Dropped jaw	
AU28	Lips sucked	
AU45	Blinking	

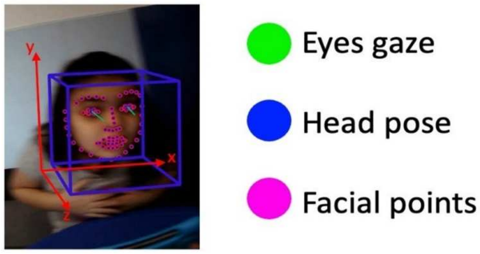

Figure 2. Gaze, head location, and facial landmarks.

which vary from 0–5, there are 17 AU's in total. It is also labeled as 'au_n_r'.

- AU01_c, AU02_c, AU04_c, AU05_c, AU06_c, AU07_c, AU09_c, AU10_c, AU12_c, AU14_c, AU15_c, AU17_c, AU20_c, AU23_c, AU25_c, AU26_c, AU28_c, AU45_c: Detect the presence (1) or absence (0) of the (AU) describing human expressions which are more accurate when only one person's face is analysed, there are 18 AU in total. Also XXX labelled as 'au_n_c' (Baltrušaitis, Mahmoud, and Robinson 2015).

The LOLY-MIDI platform incorporates two back-end and two front-end services, as shown in Figure 3. The back end incorporates the tools FACE-API, which performs facial analysis of videos using the OpenFace tool, and MIDI-API stores the information of the analysed videos in LOLY-API. The front-end LOLY-MIDI dashboard displays the information obtained from FACE-API. The data processed from the sessions, recorded by the Loly robot, is received, and the metrics for head position, degrees of attention, and emotions are defined for each child based on the results obtained from the CSV file generated by OpenFace. The MIDI dashboard presents data generated from MIDI-AM sets regarding how long and how children play the games.

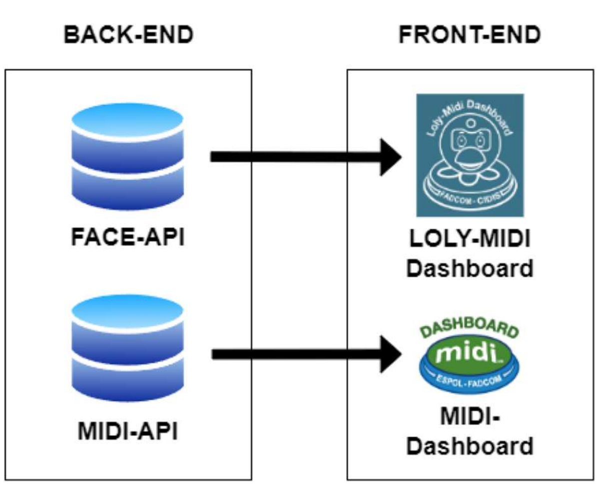

Figure 3. Back-end and front-end services are incorporated into the LOLY-MIDI Platform.

2.5. Metrics analysis for the head positions

For the perpendicular head position metrics, the poseRx field is used with values in radians between (−57.30, 20.05), poseRy with values between (−57.30, 57.30), and poseRz with values between (−57.30, 57.30). For the head tilt position metrics, the poseRx field is used with values in radians between (20.05, 114.19), poseRy with values between (−57.30, 57.30), and poseRz with values between (−57.30, 57.30).

2.6. Metrics analysis for the degrees of attention

For the metrics of the degrees of attention, we have in radians the fixed gazeAngleX field with values between (−42.97, 42.97) for the attention to Loly, the bust and tablet, while for the attention to Loly, we have the gazeAngleY field with values between (−42.97, 5. 73) and that the head is perpendicular, for the attention to the bust we have the gazeAngleY field with values between (5.73, 14.32) and that the head is perpendicular, finally for the attention to the tablet we have the gazeAngleY field with values between (14.32, 42.97) and that the head is tilted downwards.

The past emotion metrics are based on the 67 facial benchmarks (see Figure 2) obtained from the OpenFace analysis, considering measuring at least four of the six basic universal emotions defined by psychologist Paul Ekam and mentioned in (Solorzano Alcívar et al., 2021). Also, this research included two more metrics: one for a neutral emotion when no facial expression can be detected and another for disinterest when a child is outside the attention ranges to detect the face during an evaluation process. See Table 3.

The points used in the definitions of Table 3 can be seen in Figure 4, where all the points that OpenFace recognised are numbered.

Table 3. Emotion metrics are determined by OpenFace parameters.

Emotions	Description
Surprise	It is determined by the distance between point 21 and point 39 and between points 22 and 40. If both distances are greater than zero, it must be a surprise.
Happiness	It is determined by validating that the point y48 is less than the point y61 and that the point y54 is less than the point y63.
Sadness	It is determined by validating that the point y48 is greater than the point y67 and that the point y54 is greater than the point y65.
Anger	Determined by validating that point y21 is greater than point y17 and that point y22 is greater than point y26.
Neutral	Determined by validating that point y17 is greater than point y18, that point y26 is greater than point y25, and that surprise and anger are 0, i.e. these emotions are undetected.
Disinterest	Determined by attention degree metrics, it is said there is interest if they are outside the attention ranges.

3. Methodology

The framework bases for this research include applying a mixed research method, including lab tests for the data analysis as part of an experimental methodology. The study initiated revising related literature, followed by evaluating complementary tools expected to be used, such as OpenFace and FACS, for facial recognition to achieve greater precision in the designed metrics.

3.1. Method

A mixed quantitative-qualitative method study was carried out with observational and quantitative experiments performed with artificial intelligence (AI) generated images and both neurotypical children and those with ASD. AI images were used to define and validate AU's performance in the metrics, and then final tests with real children were carried out to validate the functionality of the new metrics. The final tests were carried out in an area of the robotics and video game development laboratory, equipped to perform observational processes of evaluation of children during their interaction with the robot and the games.

3.2. Sample

This study was carried out using three different sources. A significant sample size was required to validate the study and overcome challenges related to using authentic data and privacy in facial recognition. Synthetic images were used in the first instance, that is, images created artificially. Next, the research work team carried out some first experimental tests. Finally, longitudinal tests were carried out over approximately six months with both regular children and those with ASD.

Initially, 278 artificially generated images based on the Karras, Laine, and Aila (2019) dataset on the website 'thispersondoesnotexist'. were used to study AUs. The images are classified according to the type of emotion they represent. The result of processing with OpenFace is the set of action units representing happiness (n: 218), surprise (23), anger (18), and sadness (19). Other classifications are made on the neutral state of the face and situations where a facial expression cannot be effectively identified; the latter will be categorised as a face with disinterest.

This resource explored the diversity of emotions and common characteristics such as facial expressions, lip movement/position, and gaze. Multiple images were generated, and the variation in facial expressions and action units was observed for subsequent classification. Action units refer to the specific parts of the face that indicate an emotion or characteristic, and the combination of these forms a facial representation that is then used to classify emotions.

For the longitudinal and the final test included 12 children, seven with ASD limited to medium-low levels between four and seven years old, and five neurotypicals between five to ten years old, were included in the sample testing.

As proof of concept, the interaction in which the children played with a tablet guided by the Loly Robot was recorded with their parents' or representatives' accompaniment and respective consent. Loly's intervention was to give instructions and cheers during the game, i.e. explain the rules of a level and give encouragement in case of mistakes.

3.3. Procedure

For the development and metric adjustments procedures, first, the child's video is taken from both cameras, the fisheye camera on the head of the Loly robot and the camera at the bust level (see Figure 5). Both videos are stored in Google Drive and then downloaded and processed locally with the OpenFace tool.

Once the processed videos generate CSV format files useful to analyse and redesign the metrics using STATA and R languages, to validate them on a server with node.js later using the JavaScript language, the information is sent to a test database where the results of the charts analysed by OpenFace are stored and are required by the LOLY-MIDI dashboard. Statistic pie charts are produced with the dashboard to measure attention and head posture. Besides, radar charts are produced to detect the presence of the six emotions mentioned in Table 3. The procedure is explained in Figure 6.

For the analysis carried out in this research, improvements were made to the control panel, mainly defining parameters of attention metrics and facial behaviour recognition. The values generated using the facial analysis tool are based on the data analysing videos of the interaction between the child, the Loly robot, and the serious game. For this evaluation, we consider performing the analysis with two cameras installed on the Loly robot, one non-static on the Loly's forehead and the other static on the robot's chest. The purpose is to examine the attention and expressions that differ according to their position and movement to obtain data from more than one perspective to reconfirm the veracity of the results, testing their effectiveness from different facial evaluation tools.

3.4. Instruments

As instruments, the OpenFace tool was used for facial expression analysis, and the R programming language

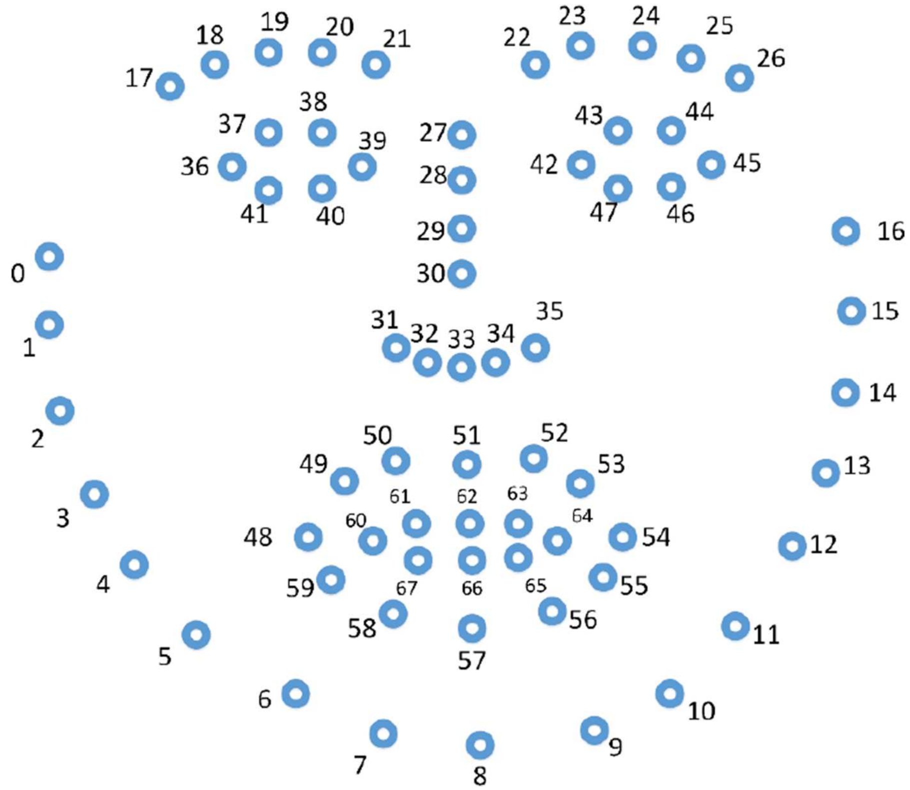

Figure 4. Facial landmark from OpenFace by Baltrusaitis et al. (2018).

was used for processing and understanding the data set generated by the tool derived from videos recorded by the Loly robot to run the tests and make real comparisons. Statistical analysis of the data obtained from the recordings was developed in local environments with R and STATA programming languages to create emotion identification metrics. These were then translated into JavaScript programming language on a local test server with node.js for graphical testing from the LOLY-MIDI dashboard.

The tests were performed with the current metrics, where facial points are used for emotion detection by applying new metrics using AU from FACS. Then, the degrees of attention and head positions were compared to determine improved aspects or changes in the metrics. Additionally, a set of free-use images of people with different expressions was used to process them in OpenFace, evaluate the results of the metrics, and improve them.

3.5. Data analysis

The CSV files analysis generated by OpenFace was carried out to eliminate outliers or processing errors, as well as identify a standard scenario on which to study the changes in the expressions of the test subjects according to the indications of the robot since these are naturally dynamic to their position throughout each session.

Once the changes in location and rotation in terms of depth, horizontal, and vertical movements have been captured, a 95% confidence interval is established to exclude values that fall outside this range from the development of the emotion metrics.

In order to determine which AUs generated by OpenFace correspond to each emotion, secondary literature is used in addition to an empirical test by the presence and intensity of each AU.

Once the metrics are updated, they will be used in the node.js server code locally, where calculations, AU combinations, and parameter changes will be made and then sent to the Loly server test database using a JSON data type to make comparisons from the dashboard between the metrics being developed and those already used, emphasising the emotions, attention, and posture of the child's head in the processed video (see Figure 5).

4. Results and discussion

The initial results come from the study with the facial behaviour analysis tool OpenFace, the action units of the FACS, and data processed by the dashboards. The presence and intensity data of action units indicate whether they activate an image or sequence of images and how strong or pronounced the facial action is (Baltrusaitis et al. 2018).

Figure 5. Loly Robots with head fisheye and bust cameras.

4.1. Action Units

In the first stage, the presence AUs (au_n_c) are used to classify facial expressions into five emotions: (1) neutral, (2) happiness, (3) surprise, (4) anger, and (5) sadness. In the second testing stage, the intensity measures (au_n_r) associated with each facial action presence variable are used to resolve conflicts in the classification generated by measures that have an ambiguous interpretation (see Figure 7). The reference points shown in the chart are the average of these parameters. It is understood that each action unit above the average represents the specific facial action. By analysing the AUs' presence and intensity, we can analyse how they are combined and how they change over time, allowing us to study the dynamics of facial expressions.

For all classifications, AUs are segmented concerning face features (upper, middle, and lower zones) (See Table 4). The disinterest result corresponds to attention to objects (robot and tablet) in the test sessions of the metrics.

For expressions identifying happiness, AUs (6, 7, 12, 14) are consistent with the literature developed around the study of facial expressions (Gavrilescu and Vizireanu 2017; Katembu et al. 2022; Namba et al. 2017; Tian, Kanade, and Cohn 2001). These are positively associated with appraisals of amusement and satisfaction and negatively related to other emotionally similar expressions such as anger, frustration, pain, or contempt. At the same time, AUs (1, 2, and 10) are associated with expressing interest in something, this being an emotionally positive load. AU05 is particularly ambiguous, contributing to expressions of happiness, surprise, and anger.

AUs (5, 25, and 26) combined are part of the characteristic set of expressions associated with the emotion of surprise; 10 and 23 can be interpreted alone as 23 (Namba et al. 2017) for contributing to similar expressions such as confusion. Regarding the expressions of anger and sadness, a conservative combination is used that does not erroneously capture the transition between emotions as a particular set that overestimates these categories.

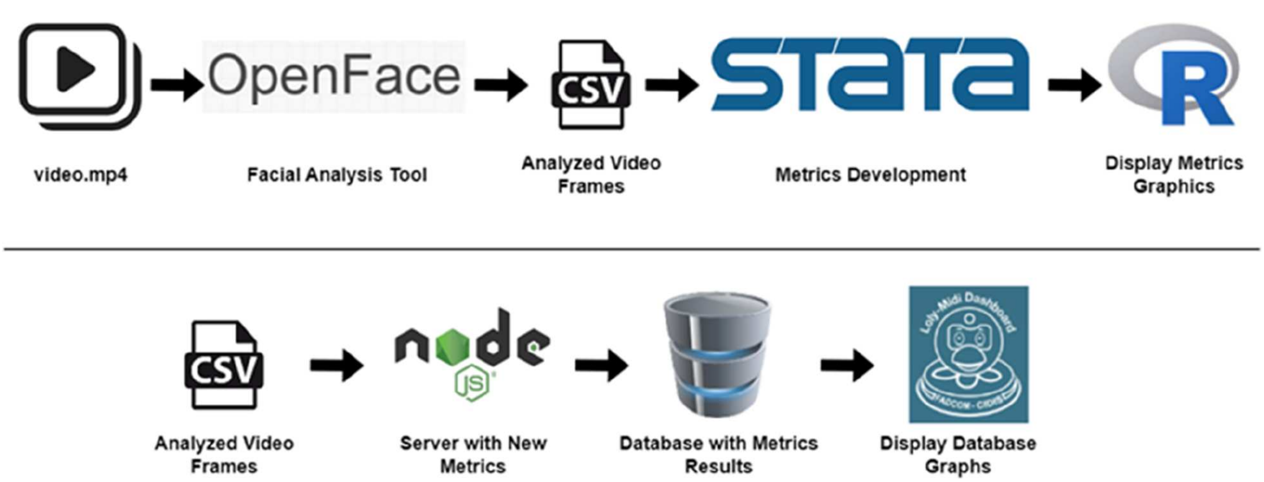

Figure 6. Procedure of the development of metrics.

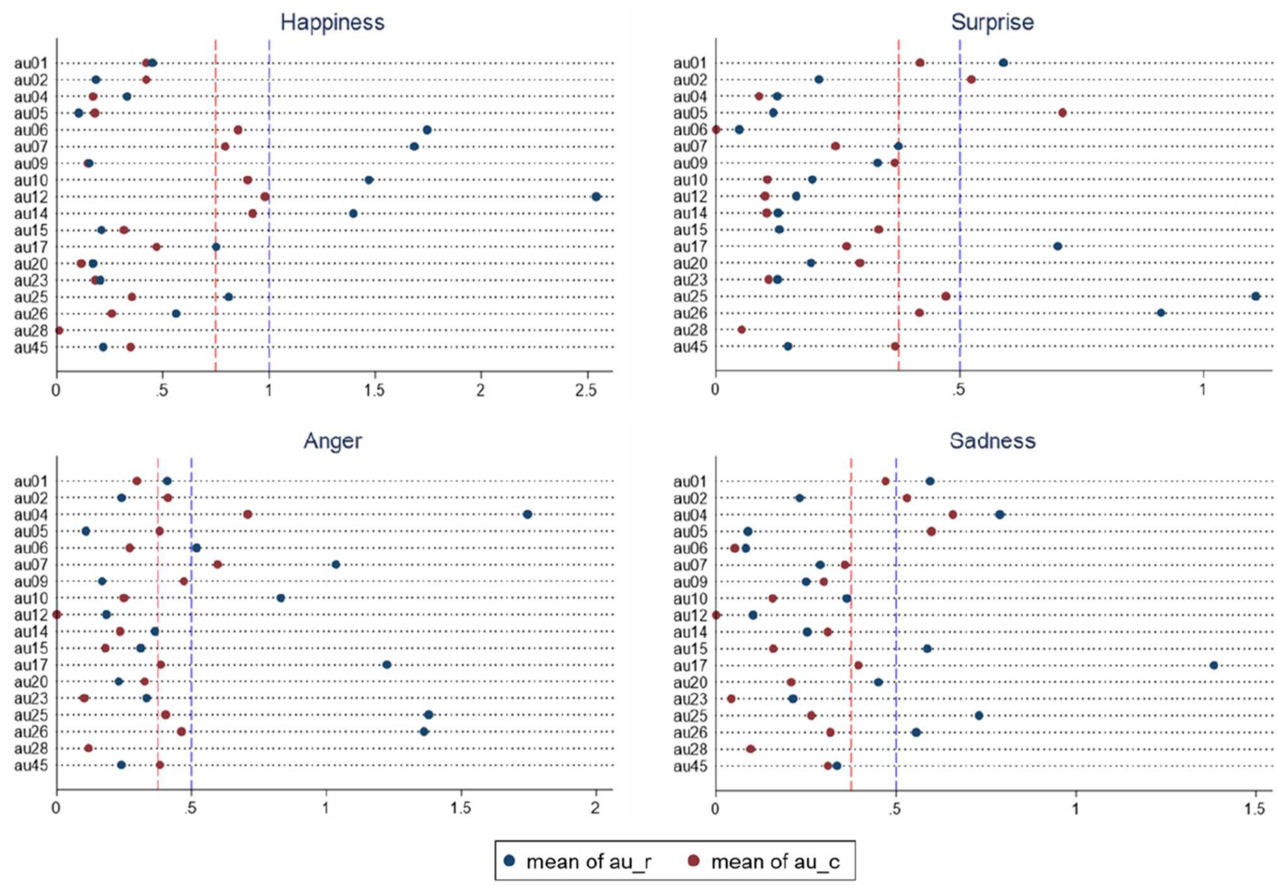

Figure 7. Action units by type of emotion identified.

4.2. Local comparison in STATA and R

A comparison was made on the Loly test server between the data with the new metrics generated and those previously established to compare the results in a production-like environment with the metrics in the JavaScript language and the data visualisation on the LOLY-MIDI dashboard. An improvement could be noticed concerning the previous metrics in the part of emotions using the radar graph, since before, the emotion that was always detected mostly was neutral, which was not always the case. It should be noted that in the fisheye camera on Loly's head, more disinterest can be detected since this camera, having a higher viewing angle and moving with the head of the Loly robot, loses information when analysed in OpenFace.

For the degrees of attention and head posture in the pie chart, the range of vision of the child towards the tablet, the bust of the Loly robot, and Loly's head are more limited by confidence intervals that define a range where the child's vision must be for it to be detected. It should be noted that the fisheye camera on the head increases the confidence interval for detecting head posture or attention because this camera moves and increases the detection range of these measurements, increasing the bias of the estimates.

Table 4. Classification of action units by type of emotion.

Emotions	Action Units	Additional expressions
Neutral	Exclusion (if the picture does not fall into any emotion category)	None
Happiness	Upper AU (01, 02, 05, 06, 07); Lower AU (10, 12, 14)	Fun, rejoicing, joy
Surprise	Upper AU (01, 02, 05); Lower AU (25, 26)	Astonishment, shock
Anger	Upper AU (04, 05, 07); Intermediate AU (09); Lower AU (10, 23)	Disgust, contempt, indignation
Sadness	AU superior (01, 04); AU inferior (15)	Depression, melancholy, homesickness
Disinterest	Degree of attention (= 0), excluding success (= 0)	None

4.3. Emotions

The current identification methodology consists of distance measurements between facial landmarks. Results for the set of characteristic images of each emotion are compared with the result of the action units (see Figure 8).

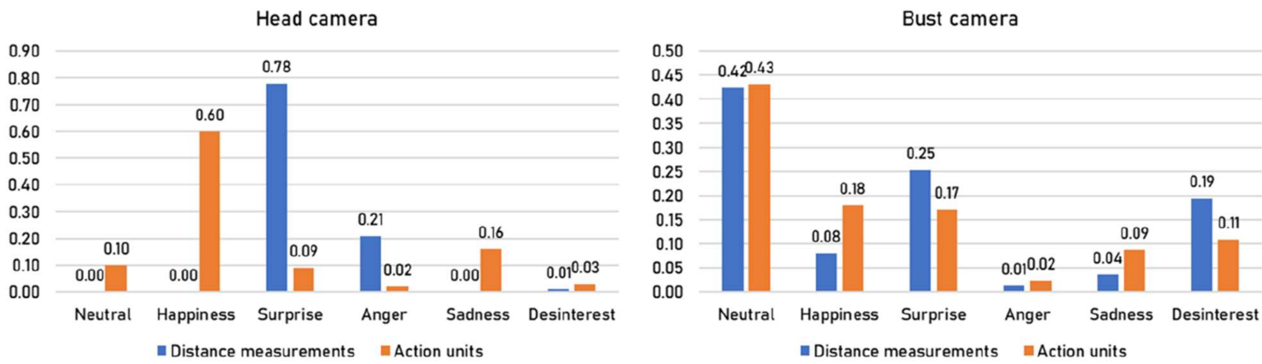

Figure 8. Comparison between specifications.

The results in terms of proportion are not comparable due to the independence of emotions. Action units recognise only one emotion for each row, validated by the presence and intensity of the AUs. Distance measurements recognise more than one emotion per row. The problem with distance measurements is its percentage of disinterest (less than 5%), differentiating it from neutral emotion (absence of other emotions). Disinterest is the impossibility of recognising an emotion by the movement of the study subject or movement in the camera. Both specifications tend to recognise emotions more frequently with a neutral charge, considering that the transition between emotions can be attributed to this situation.

The significant differences between emotion recognition with the head camera compared to the camera located on the robot's chest are attributable to the movement of the former; this is due to the capture of distortions of facial characteristics, movement speed, and tracking of reference points.

4.4. Tolerance to movement

The workspace is validated by the average movement of the user whenever a face is perfectly identifiable to give depth, lateral movements, and height. At the same time, it delimits the observational range for the elements of interest (robot and tablet). These delimitations are dynamic concerning each session, using a 95% confidence interval to reduce the main identification problems arising from the analysis tool or the user's interaction with the Loly robot. This interaction favours the principle of adaptability mentioned in (Aguiar et al. 2022), as it allows to adapt to the user's movement while retaining acceptable results. The validation process of the interaction consists of identifying mitigation mechanisms by distance of the results between the user and the robot or by the presence of other identifiable people in the video; the greater the distance, the lower the accuracy of the estimates for all parameters (see Figure 9). This result means there is a lower reliability in

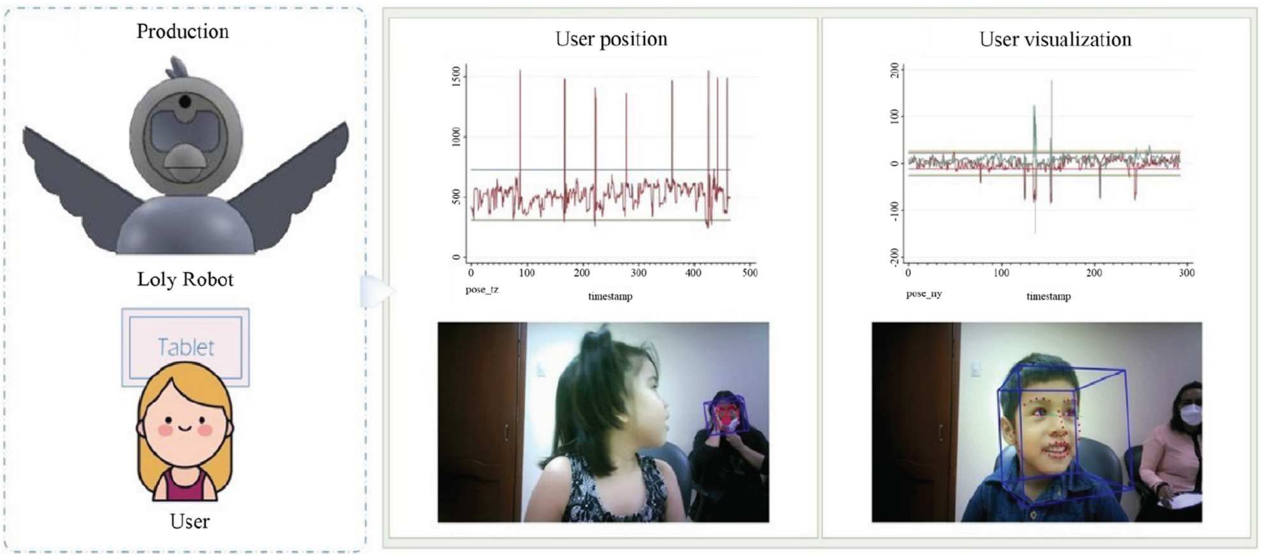

Figure 9. Interaction process with the Loly robot.

identifying emotions and a higher probability of going out of the robot's vision range.

The results added to the research findings of Ji et al. (2023) show that human-robot-games platforms have improved the treatment of children with ASD, especially long-term training. With the recognition of emotions and attention, it is possible to study the children's progress along with six basic emotions between time intervals and the attention on each played game.

5. Conclusions

The current research undertaken allowed us to upgrade and validate metrics for the LOLY-MIDI HRG platform, using the Facial Action Coding System (FACS) obtained from the analysis of the videos made in OpenFace 2.0 and adapt it to an HRG interface. This validation, contrasted with facial points, was initially applied to evaluating metrics. Two cameras are used from different perspectives to validate the best type of camera and camera position for the facial analysis. The applications also examine parameters for detecting head postures and the child's attention to obtain more accurate results from the HRG platform managed on the LOLY-MIDI website.

Obtaining and presenting more accurate and easier-to-interpret parameters for psychologists and educational therapists can facilitate and help in the treatment of children with ASD to support their learning processes and attention levels. Once the laboratory and field tests were completed, using the HRG platform with visual monitoring, it was possible to analyse the refined metrics to evaluate facial expressions that allow identifying the degree of attention and emotions in neurodivergent individuals, particularly in ASD children from 4 to 7 years old.

Two metrics were used and refined to obtain results for the completed test. With the first optimizations established in the metrics, the parameter 'success' is considered equal to 1, excluding those parts of the video where the OpenFace application does not detect the child's face, eliminating information that alters the confidence intervals and means in the developed metrics.

Then, several of the proposed metrics were updated by widening the confidence intervals for attention and head posture detection. These metrics help to analyse more information from the video and decrease the 'no information' and 'no attention' parameters in the LOLY-MIDI dashboard pie charts.

In summary, obtaining and presenting classified results by type of emotion using action units allows us to establish more accurate and visible parameters concerning what the videos reflect. These results can be simpler and more reliably interpreted by non-technical scientists. For example, psychologists and educational therapists, most of whom require this information to analyse the behavioural use of HRG platforms and their impact on learning processes and attention levels, especially in children with ASD.

6. Future works

Future work within this research is planned to incorporate some processes that allow the use of the camera mounted on the robot head without losing accuracy. For example, exclude in the videos the times in which the robot's head performs some movement since this generates variation in the results of the videos recorded with the camera on Loly's bust. If the movement time of Loly's head is excluded, there will be less data to analyse, so the metrics will be accurate but will have less information. Another alternative is to apply some filtering algorithm to the motion of the camera mounted on the robot's head.

In addition, it is recommended that new parameters be set to recognise which video or camera each analysis performed in OpenFace belongs to so that the appropriate metrics for each video, either with the fisheye camera or the Loly bust camera, can be selected. It should be noted that other video processing tools, such as OpenCV, have more support (Bradski 2000). These have more documentation and are widely used for image processing. Additionally, modularising the metrics into different components for emotion analysis, attention, and posture detection through artificial intelligence would give more control over the important fields to evaluate when processing the video as required. It would also help to rely on something other than a third-party tool over which there is limited control for the LOLY-MIDI HRG platform processes. This matter is added to the fact that delivering a large amount of data requires more analysis time, reducing processing time if metrics are modularised.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Aguiar, Yuska Paola Costa, Edith Galy, Anaïs Godde, Maëla Trémaud, and Carole Tardif. 2022. "AutismGuide: Usability Guidelines to Design Software Solutions for Users with Autism Spectrum Disorder." *Behavior & Information Technology* 41 (6): 1132–1150. https://doi.org/10.1080/0144929X.2020.1856927.

Alhakbani, Noura. 2024. “Facial Emotion Recognition-Based Engagement De-Tection in Autism Spectrum Disorder.” *International Journal of Advanced Computer Science and Applications* 15 (3). https://doi.org/10.14569/IJACSA.2024.0150395.

Amukune, Stephen, Karen Caplovitz Barrett, and Krisztián Józsa. 2022. “Game-Based Assessment of School Readiness Domains of 3-8-Year-Old-Children: A Scoping Review.” *Journal of New Approaches in Educational Research* 11 (1): 146–166. https://doi.org/10.7821/naer.2022.1.741.

Auliawan, Althaf Gauhar, and Susy Ong. 2020. “The Usage of AI Robot in English Language Teaching for City Revitalization Case Study: Toda Daini Elementary School, Toda City, Saitama, Japan.” *IOP Conference Series: Earth and Environmental Science* 436 (1): 012022. https://doi.org/10.1088/1755-1315/436/1/012022.

Baltrušaitis, Tadas, Marwa Mahmoud, and Peter Robinson. 2015. “Cross-Dataset Learning and Person-Specific Normalization for Automatic Action Unit Detection.” In *2015, the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 06:1–6. https://doi.org/10.1109/FG.2015.7284869.

Baltrusaitis, Tadas, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. “OpenFace 2.0: Facial Behavior Analysis Toolkit.” In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG, 2018)*, 59–66. Xi’an: IEEE. https://doi.org/10.1109/FG.2018.00019.

Bradski, G. 2000. The openCV library. In *Dr. Dobb’s Journal: Software Tools for the Professional Programmer*, 120–23.

Chinchay, Yussy, Juan C. Torrado, Javier Gomez, and Germán Montoro. 2024. “Towards more Supportive ICT for Children with Autism Spectrum Disorders: Lessons Learned from COVID-19 Pandemic.” *Behaviour & Information Technology* 43 (12): 3027–3046. https://doi.org/10.1080/0144929X.2023.2268734.

Drimalla, Hanna, Irina Baskow, Behnoush Behnia, Stefan Roepke, and y Isabel Dziobek. 2021. “Imitation and Recognition of Facial Emotions in Autism: A Computer Vision Approach.” *Molecular Autism* 12 (1): 27. https://doi.org/10.1186/s13229-021-00430-0.

Dubois-Sage, Marion, Baptiste Jacquet, Frank Jamet, and Jean Baratgin. 2024. “People with Autism Spectrum Disorder Could Interact More Easily with a Robot than with a Human: Reasons and Limits.” *Behavioral Sciences* 14 (2): 131. https://doi.org/10.3390/bs14020131

Espinoza Erazo, Nathaly Andrea, and Ricardo Patricio Almeida Gallardo. 2018. “Rediseño y construcción de un robot social para interacción hombre máquina orientado a la educación de niños/as de 2 - 5 años.” http://repositorio.espe.edu.ec/jspui/handle/21000/14865.

Gavrilescu, Mihai, and Nicolae Vizireanu. 2017. “Predicting the Sixteen Personality Factors (16PF) of an Individual by Analyzing Facial Features.” *EURASIP Journal on Image and Video Processing* 2017 (1): 59. https://doi.org/10.1186/s13640-017-0211-4.

Hamzah, Mohd Shah Johan, Syamimi Shamsuddin, Mohd Azfar Miskam, Hanafiah Yussof, and Khadijah Said Hashim. 2014. “Development of Interaction Scenarios Based on Pre-School Curriculum in Robotic Intervention for Children with Autism.” *Procedia Computer Science* 42:214–221. https://doi.org/10.1016/j.procs.2014.11.054.

Hoffmann, Laura, and Nicole C. Krämer. 2021. “The Persuasive Power of Robot Touch. Behavioral and Evaluative Consequences of Non-Functional Touch from a Robot.” Editado by Chen Lv. *PLoS One* 16 (5): e0249554. https://doi.org/10.1371/journal.pone.0249554.

Jeong, Sooyeon, Cynthia Breazeal, Deirdre Logan, and Peter Weinstock. 2018. “Huggable: The Impact of Embodiment on Promoting Socio-Emotional Interactions for Young Pediatric Inpatients.” In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. Montreal QC Canada: ACM. https://doi.org/10.1145/3173574.3174069

Ji, Hao-Yu, Bo-Wen Chen, Zhi-Yong Wang, Han-Wei Ma, Rui-Han Lin, Yi-Xuan Sheng, and Hong-Hai Liu. 2023. “A Humanoid Robot to Assess the Affective Imitative Expression Abilities of Children with Autism Spectrum Disorder.” In *2023 International Conference on Machine Learning and Cybernetics (ICMLC)*, 352–57. https://doi.org/10.1109/ICMLC58545.2023.10327992

Karras, Tero, Samuli Laine, and Timo Aila. 2019. “A Style-Based Generator Architecture for Generative Adversarial Networks.” arXiv. https://doi.org/10.48550/arXiv.1812.04948

Katembu, Stephen, Qiang Xu, Hadiseh Nowparast Rostami, Guillermo Recio, and Werner Sommer. 2022. “Effects of Social Context on Deliberate Facial Expressions: Evidence from a Stroop-like Task.” *Journal of Nonverbal Behavior* 46 (3): 247–267. https://doi.org/10.1007/s10919-022-00400-x.

Khan, Mohammad Shadab, Noor Azizah K. S. Mohamadali, Zulfikar Ahmed Maher, Syed Qamrun Nisa, Humaiz Shaikh, and Asadullah Shah. 2019. “Information Technology (IT) Based Intervention Among Individuals with ASD (Autism Spectrum Disorder): A Review.” In *2019 IEEE International Conference on Innovative Research and Development (ICIRD)*, 1-3. https://doi.org/10.1109/ICIRD47319.2019.9074749.

Kubota, Alyssa, Emma I. C. Peterson, Vaishali Rajendren, Hadas Kress-Gazit, and Laurel D. Riek. 2020. “JESSIE: Synthesizing Social Robot Behaviors for Personalized Neurorehabilitation and Beyond.” In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 121-30. Cambridge United Kingdom: ACM. https://doi.org/10.1145/3319502.3374836.

Kuutti, Tiina, Nina Sajaniemi, Piia M. Björn, Nina Heiskanen, and Jyrki Reunamo. 2022. “Participation, Involvement and Peer Relationships in Children with Special Educational Needs in Early Childhood Education.” *European Journal of Special Needs Education* 37 (4): 587–602. https://doi.org/10.1080/08856257.2021.1920214.

Mejía Silva, Michelle Carolina, and Jennifer Alexandra Núñez Arroba. 2016. “Diseño y construcción de un robot social para interacción hombre máquina.” http://repositorio.espe.edu.ec/jspui/handle/21000/11733.

Namba, Shushi, Russell S. Kabir, Makoto Miyatani, and Takashi Nakao. 2017. “Spontaneous Facial Actions Map onto Emotional Experiences in a Non-social Context: Toward a Component-Based Approach.” *Frontiers in*

Psychology 8 (mayo): 633. https://doi.org/10.3389/fpsyg.2017.00633.

Pakkar, Roxanna, Caitlyn Clabaugh, Rhianna Lee, Eric Deng, and Maja J Mataricc. 2019. "Designing a Socially Assistive Robot for Long-Term In-Home Use for Children with Autism Spectrum Disorders." In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–7. New Delhi, India: IEEE. https://doi.org/10.1109/RO-MAN46459.2019.8956468.

Shohieb, Samaa M., Ceymi Doenyas, and Abdelghafar M. Elhady. 2022. "Dynamic Difficulty Adjustment Technique-Based Mobile Vocabulary Learning Game for Children with Autism Spectrum Disorder." *Entertainment Computing* 42 (mayo): 100495. https://doi.org/10.1016/j.entcom.2022.100495

Solorzano Alcivar, Nayeth I., Luis C. Herrera Paltan, Leslie R. Lima Palacios, Dennys F. Paillacho Chiluiza, and Jonathan S. Paillacho Corredores. 2022. "Visual Metrics for Educational Videogames Linked to Socially Assistive Robots in an Inclusive Education Framework." In *Perspectives and Trends in Education and Technology*, edited by Anabela Mesquita, António Abreu, and João Vidal Carvalho, Vol. 256, 119–32. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-16-5063-5_10

Solorzano Alcivar, Nayeth I., Luis C. Herrera Paltan, Leslie R. Lima Palacios, Dennys F. Paillacho Chiluiza, Jonathan S. Paillacho Corredores, and Dennys F. Paillacho Chiluiza. 2021. "Metrics Design of Usability and Behavior Analysis of a Human-Robot-Game Platform." In *Applied Technologies*, edited by Miguel Botto-Tobar, Sergio Montes León, Oscar Camacho, Danilo Chávez, Pablo Torres-Carrión, and Marcelo Zambrano Vizuete, 164–78. Communications in Computer and Information Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-71503-8_13

Syrjämäki, Marja, Jyrki Reunamo, Henri Pesonen, Raija Pirttimaa, and Elina Kontu. 2023. "The Involvement of Autistic Children in Early Childhood Education." *European Journal of Special Needs Education* 38 (6): 879–893. https://doi.org/10.1080/08856257.2023.2179310.

Tian, Y.-I., T. Kanade, and J.F. Cohn. 2001. "Recognizing Action Units for Facial Expression Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2): 97–115. https://doi.org/10.1109/34.908962.

Turrado-Sevilla, María Ángeles, and Isabel Cantón-Mayo. 2022. "Design and Validation of an Instrument to Measure Educational Innovations in Primary and Pre-Primary Schools." *Journal of New Approaches in Educational Research* 11 (1): 79–96. https://doi.org/10.7821/naer.2022.1.727.

Westlund, Jacqueline Kory, Jin Joo Lee, Luke Plummer, Fardad Faridi, Jesse Gray, Matt Berlin, Harald Quintus-Bosz, et al. 2016. "Tega: A social robot." In 2016 *11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 561–561. Christchurch: IEEE. https://doi.org/10.1109/HRI.2016.7451856.

Wolf, Edo de, and Jamy Li. 2020. "Double Trouble: The Effect of Eye Gaze on the Social Impression of Mobile Robotic Telepresence Operators." In *Social Robotics*, edited by Alan R. Wagner, David Feil-Seifer, Kerstin S. Haring, Silvia Rossi, Thomas Williams, Hongsheng He, and Shuzhi Sam Ge, Vol. 12483, 357–68. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-62056-1_30