Towards Enhancing the Multimodal Interaction of a Social Robot to Assist Children with Autism in Emotion Regulation

Marcelo Rocha[1(✉)], Pedro Valentim[1], Fábio Barreto[1], Adrian Mitjans[2], Dagoberto Cruz-Sandoval[3], Jesus Favela[2], and Débora Muchaluat-Saade[1]

[1] MÍdiaCom Lab - UFF, Niterói, Brazil
{marcelo_rocha,pedroalvesvalentim,fbarreto,debora}@midiacom.uff.br
[2] CICESE, Ensenada, Mexico
favela@cicese.mx
[3] Computer Science and Engineering, UC San Diego, San Diego, USA
dcruzsandoval@eng.ucsd.edu

Abstract. Assistive robots are expected to become ubiquitous by transforming everyday life and are expected to be widely used in healthcare therapies. SARs (Socially Assistive Robots) are a class of robots that are at an intersection between the class of assistive robots and that of interactive social robots. SARs are being explored to assist in the diagnosis and treatment of children with ASD (Autism Spectrum Disorder). A SAR called EVA has been used to assist non-pharmacological interventions based on verbal, non-verbal communication and social interaction. The EVA robot can currently speak, listen and express emotions through looking. Towards offering immersive therapies for autistic children, this work enhances EVA's capabilities to recognize user emotions through facial expression recognition and also to create light sensory effects in order to make the therapy more attractive to users. A therapy session was developed through a serious game where the child should recognize the robot's emotions. During the game, EVA recognizes the child's facial expression to check his/her learning progress. We invited a neurotypical 6-year-old child to play the game, with the consent of her parents, and recorded videos of the game session. Those videos were evaluated by 48 expert physicians and psychologists in therapies for ASD using the Technology Acceptance Model (TAM). They considered our work useful and agreed it would help them doing their job more effectively.

Keywords: Socially Assistive Robots (SAR) · Multimodal interaction · Serious game · Emotion regulation · Autism Spectrum Disorder (ASD)

1 Introduction

Robotic technologies are no longer used only in factories and have been increasingly used to assist people perform activities of daily living. Assistive robots

H. Lewy and R. Barkan (Eds.): PH 2021, LNICST 431, pp. 398–415, 2022.
https://doi.org/10.1007/978-3-030-99194-4_25

have been used to improve the quality of life in society and are becoming more and more common [28]. Robots are used in our lives as assistive devices and robotics is a method capable of enhancing the physical and cognitive abilities of humans [7]. Robots can help with housework or assist people with some type of physical disability. However, advances have emerged in a new field of robotics where the objective is not only to provide assistance, but to provide stimuli to humans through interaction with the robot [27]. Robots are expected to become ubiquitous by transforming everyday life and are expected to be widely used in healthcare therapies. SARs (Socially Assistive Robots) [7] are a class of robots that are an intersection between the class of assistive robots (of the types that provide user assistance) and that of interactive social robots (those that communicate with the user through social and not physical interaction). SARs have been used in various types of healthcare therapies, such as non-pharmacological interventions, which can improve the quality of life for patients and those around them [18]. A Socially Assistive Robot called EVA [4,32] has been used to assist non-pharmacological interventions based on verbal, non-verbal communication and social interaction with patients with dementia and Alzheimer's disease (AD). A user study with five patients with (AD) was conducted and showed that EVA was effective in engaging therapy participants. The EVA robot can currently speak, listen and express emotions through looking.

The development and use of interactive technologies for individuals with ASD (Autism Spectrum Disorder) has also been growing rapidly. Those technologies can enrich interventions, facilitate communication, support data collection and have the potential to improve the assessment and diagnosis process of individuals with ASD [12]. In [26] there is a proposal for a robotic coaching platform for social training, motor and cognitive capabilities. The robot used in this work is the NAO robot, which has been widely used in social therapy with children with ASD. As the NAO is a humanoid robot, it looks like a human without being one, and it can provide audio and visual stimuli. All these characteristics are favorable for interaction with ASD children, who tend to prefer simplified stimuli to avoid focusing on details. The work [30] compares the results from robot-based interventions with those from human-based applied to children with ASD and intellectual disabilities. In [6], positive results of the interaction of children with ASD with a robot are presented, indicating a greater incidence of eye contact, proximity and interaction than the interactions between children and a human. There is evidence supporting the claim that autistic children prefer robots to humans [13,14,21]. The results of these studies indicate that robots can generate several positive effects in this type of therapy and that human-robot interaction can help to solve the problems that occur in human-human interaction. Deficits in social and communication skills in patients with autism encompass a number of other disabilities, such as difficulty in recognizing body language or demonstrating eye contact with whom they talk to and problems in expressing their own emotions and understanding those of others [6].

In order to enhance robot capabilities to interact with users, multimodal interaction can be very useful. Besides voice interaction, which is frequently used by SARs such as EVA, video interaction could be provided to help capture user emotion, for example. In addition, sensory effects could be used to create immersive therapies and make them more attractive to users, specially children.

Towards offering immersive therapies for autistic children, this work enhances EVA's capabilities to recognize user emotions through facial expression recognition and also to create light sensory effects in order to make the therapy more attractive to users. The idea is to use the interactive robot as the first stage of an emotion recognition therapy for children with ASD.

We developed a therapy session through a serious game where the child should recognize the robot's emotions. During the game, EVA recognizes the child's facial expression to check his/her learning progress. We invited a neurotypical 6-year-old child to play the game, with the consent of her parents, and recorded videos of the game session. Those videos were evaluated by expert physicians and psychologists in therapies for ASD using the Technology Acceptance Model (TAM). We made a statistical analysis to compare evaluation results between expert and beginner ASD therapists.

The remainder of this paper is structured as follows. Section 2 discusses related work. Section 3 presents EVA's architecture and main functionalities. Section 4 presents our extensions to EVA. Section 5 discusses the autism therapy session that we have implemented. Section 6 presents an evaluation done by ASD therapists. Section 7 concludes this paper and presents future work.

2 Related Work

SARs have become popular tools in interventions with patients with ASD. Robots have been used in special education schools and care centers for autistic people. In [11], a framework was developed to engage children with ASD in sensory interactions. They used two robots, the first called Romo and the second was a small humanoid robot. They were capable of producing visual and audio stimuli. The Romo robot could express its emotions through three components: movement, facial expressions and sound effects. The framework was used as a tool for emotion regulation therapies and worked as follows. The patient interacted with the robot through a game, which ran on a notebook. The game uses three key states, which are the user's emotion state, the penguin character's emotion state, and a fixed predetermined goal emotion state towards which the penguin steers the user. To build a relationship with the user, the penguin first dynamically allocates a temporary goal state that approximates the user's state, then it gradually moves that temporary goal state closer to the predetermined goal to facilitate emotion regulation. Thus, based on the continuous interaction with the penguin character, the algorithm tries to guide the user towards the desired target state. The participant was then guided, progressively, through the robot's social behavior (movement/facial expression/sound effects) until the final facial expression. The results of a user study confirmed the viability of the framework as a tool for regulating emotion.

In [25], a mobile robot was developed and used to investigate its potential in stimulating social interactions in children with autism. Those stimuli were provided through games or therapeutic activities. During sessions with the robot, specialists in autism observed some capabilities in patients who had never been seen in other therapies, indicating that the robot could assist those specialists in finding new capacities in children with autism, creating more effective therapies, adapted to each patient.

A proposal to use a robot, which is capable of measuring the social behavior of a child with ASD during an interaction, was made in [9]. That proposal uses a wearable device to detect the child's smile. Although the device allows a smile to be detected even without the child facing the robot, the use of a wearable device can be uncomfortable, especially for children with ASD who tend to have tactile hypersensitivity.

The authors in [2] present an interactive robotic system that provides emotional and social behaviors for multisensory therapy for children with ASD. The system uses two robots, one of which is capable of expressing emotions through facial expressions and the other demonstrates its emotions using body language and gestures. The authors' proposal is that the robots go through a maze-like scenario and during the circuit find objects that stimulate the five senses, i.e. sight, hearing, smell, taste and touch. The idea is that children can learn from robots to deal with sensory overload.

A study using two robots was conducted in [8]. The aim was to examine the robot's behavior and whether or not it affected the behavior of children with ASD. Evidence was found that the robot's behavior does affect the child's social behavior, both in human-human interaction and in human-robot interaction.

The use of sensory effects has also been applied for autism therapies. In [19] there is a proposal for an adaptive physical environment that allows children with severe autism to successfully interact with multimodal stimuli. This environment generates stimuli of various types (visual, aural, and vibrotactile) in real-time. In [34] a system of sensory devices is presented to support therapists of children with ASD. The Sensor Box could be handled by children and was capable of generating a series of sensory stimuli, such as sounds and light effects.

A large-scale elastic multisensory surface that allows users to make music when tapping and touching on top of the canvas was developed in [3]. On this surface, users could play sounds of different musical instruments. BendableSound is a system using a Kinect sensor, speakers, and an ultra-short throw projector placed behind a spandex fabric.

Previous studies [17] point to a link between playing action games and cognitive and perceptual improvement. However, that improvement is not only associated with playing action games. Different types of games can improve different cognitive aspects. In [20] an implementation of the Stroop game with light sensory effects was made for the Brazilian digital TV system. A serious game (SG) was developed to train the selective attention of the elderly through the Stroop effect. The game displays on the screen the name of a color written with another color, then, the user through the TV remote control buttons, selects the correct color to answer. The light sensory effect is used for helping the elderly user with the correct answer.

Through the use of games it is possible to provide interactivity, increase mental activity and promote social interaction between various users. The work in [24] describes the development of an SG that can be used by children with ASD in order to improve communication and social interaction. In [1] an SG for children with ASD was proposed with the aim of investigating the effects of a game as a play therapy. Through an empirical study, conventional games (non-computer block-games) were compared with the proposed SG. Evidence showed an improvement in the social interaction of children, in the process of collaboration between them and a decrease in solitary games.

The emotional development of a child involves the ability to understand his/her own feelings as well as those around him/her. For a child with ASD, the process of understanding and expressing feelings is very difficult. An SG was designed in [31] with the aim of helping children with ASD, through an imitation process, to recognize different facial expressions.

Unlike the previously mentioned studies, our work extends the multimodal interaction capabilities of a SAR, integrating sensory effects and facial expression recognition in a single robot-based platform. The authors of [2,8,9,11,25] propose the use of robots for therapies with children with ASD, without offering the possibility of capturing the child's facial expression through video. In [3,19,34] sensory effects are used for ASD therapies, including light effects, but they do not use a robot-based platform. The studies [1,17,20,24,31] deal with SG, however, this work proposes an SG offered by the robot. These points are summarized in Table 1.

Table 1. Related work comparison

Related work	Use a robot	Light sensory effects	Voice recognition	Use facial expression	Serious game	Facial expression recognition
[2]	✓	–	–	✓	–	–
[8,11]	✓	–	–	✓	✓	–
[9]	✓	–	–	–	–	✓
[25]	✓	✓	–	✓	✓	–
[19]	–	✓	✓	–	–	–
[3]	–	✓	–	–	–	–
[34]	–	✓	✓	✓	–	–
[1,17,24]	–	–	–	–	✓	–
[20]	–	✓	–	–	✓	–
[31]	–	–	–	–	✓	✓
Our Work	✓	✓	✓	✓	✓	✓

3 EVA: An Open-Source Social Robotics Platform

EVA (Embodied Voice Assistant) is an open-source robotics platform. EVA provides most of the resources to build a fully functional social robot at an affordable cost, and design and create interactions for specific contexts and populations. EVA's repository provides all the elements to make your own social robot, such as 3D models, software, and guidelines to assemble it using open-hardware solutions such as the Raspberry Pi and Arduino-based card-boards. The basic version of the EVA robot includes a voice interface (microphone array and speaker) and a 5-inch touchscreen to display a set of basic facial expressions and manage the robot's basic features. Furthermore, EVA integrates a ring of LEDs in its chest to display light animations. In addition to the above elements, the intermediate version (see Fig. 1) includes two servo-motors to provide 2 degrees of freedom (DOF) to the robot's head and a depth camera (Intel RealSense) for tasks involving computer vision. The complete (mobile) version includes a mobile platform based on TurtleBot to explore more complex body gestures and mobility features. Although these three versions have been proposed by the creative and development team, makers and research teams can create, combine, or add new elements to create their new EVA-based solutions.

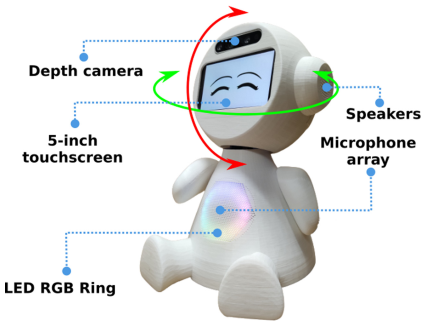

Fig. 1. Intermediate version of EVA

3.1 Software Architecture

The software architecture of EVA includes two main modules. The first one is a graphical interface accessible using the touchscreen directly. The operator uses it to configure and manage the basic features, such as Wi-Fi connection, deploy pre-programmed interaction, network information, restart, and shutdown. The second one, called Core Module, has all the logic and core features of EVA, including back-end components: server, behavioral controller; and front-end components: operator console, WoZ design, programming, and configuration. Figure 2 illustrates EVA's main software components, which are detailed as follows.

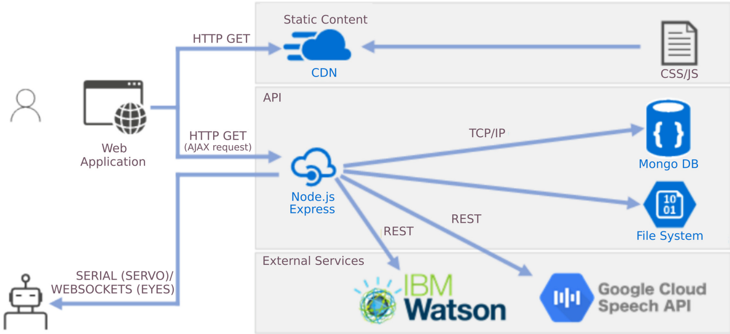

Fig. 2. EVA's main software components [16]

Core Module. The back-end components of the Core Module are responsible for the logic and operation of EVA. These components are the core of the EVA and implement tasks such as: running a server to establish communication between the operator and robot; controllers to interact with the sensors and actuators; and directly managing the database. The back-end components have been developed with NodeJS, Python, and C++.

Server Controller. This component enables a web server that deploys a REST API used by the front-end components to manage and configure EVA's features and behavior. The server controller manages the request and handles those which are related to configuration and storage. Those requests related to EVA's features, operation, and behavior are sent to the behavioral controller.

Behavioral Controller. This is the main component of EVA. It enacts the behavior (utterances, gestures, movements, facial expressions, light animations) of the robot controlling the actuators (speaker, display, servo-motors, and ring of LEDs). Furthermore, it manages and processes the inputs of the sensors (microphone, camera) in order to enact the specific response or behavior of the robot. The component uses third-party cognitive services from Watson (Text-to-Speech, Vision) and Google (Speech, Dialogflow) to process inputs from sensors and outputs to actuators. Moreover, open-source solutions can be integrated, such as Vosk (speech recognition), Mimic (Text-to-Speech), and Padatious (Natural Language Understanding).

The front-end components of the Core module deploy user interfaces to configure and manage the behavior and features of the social robot. The frond-end has been developed using the framework AngularJS. The description of each front-end component is provided as follows.

Operator Console. The operator can manage the behavior of the robot directly using this component. The operator console includes modules to send personalized utterances, enact facial expressions, make personalized and pre-programmed movements, and deploy light animations. Using this module, a user can operate the robot in real-time.

WoZ Design. Wizard of Oz (WoZ) is one of the most popular techniques to conduct early interactions to design and define future autonomous interactions. The EVA platform considers the importance of this approach and provides a component to design and deploy this kind of interaction. The component allows creating WoZ interactions using (pre)defined utterances, sounds, music tracks, pre-programmed facial expressions, and gestures (movements). These interactions are stored in the database for future uses.

Visual Programming. In addition to the WoZ module, the platform EVA includes this component to design and deploy autonomous interactions. The user can design the interaction scripts defining the flow (sequence, conditions, loops) using a visual programming language (VPL). This includes the definition of actions such as utterances, voice recognition, time to response, atom responses (utterances, sounds, facial expressions, gestures, light animations), or composed responses - a combination of atom responses. These autonomous interactions are stored in the database for future applications.

Configuration. The configuration module is used to define the robot's name (wake-word), voice (language, gender), and voice recognition (language). Supported languages are based on Watson (TTS, STT) and Google Cloud Speech. Moreover, we can include credentials to use third-party cognitive services.

4 Extending EVA's Multimodal Interaction

As we mentioned previously, the EVA robot is able to conduct a personalized session through interaction scripts. These scripts can be easily created using a VPL designed exclusively for EVA [16]. VPL is not concerned with the syntactic details of a conventional programming language, it was created to facilitate the construction of interaction scripts allowing people who are not specialized in programming to create their own scripts.

With the use of a graphical tool, building a script is done in a simple way, simply dragging and dropping the control components into the application window. The language has several components, including those that control the flow of script execution, counters, timers, conditional controls and the robot's speech and listening controls. Executing the interaction script, the robot can interact with the user, obtaining information through listening and expressing itself through speech, facial expressions, body gestures and light animations.

This work uses the EVA platform and extends its core and front-end capabilities for user interaction with video capture and sensory effects. Particularly, we used the basic version of EVA in our work. We extended VPL to include light sensory effects into therapy scripts, so that EVA can communicate with a smart bulb using a Wi-Fi interface to turn it on or off. We also extended VPL to capture the user emotion through facial expression recognition, therefore, an integrated camera is used to take the user's picture and analyze his/her facial expression to identify if he/she is happy, sad, angry or neutral. Section 4.1 presents our proposal for including light sensory effects and Sect. 4.2 discusses our proposal for facial expression recognition.

4.1 Light Sensory Effects

The first component added to the language was *Light*, which gives the robot the ability to control lighting. With the addition of this component, EVA can turn on the light, turn it off and select the desired color. The color selection can be done with the help of an RGB color palette. At the moment a *Light* component is inserted into a script, a configuration window appears and the script developer can select the light color and bulb status, which can be turned on or off.

All lighting control is done through a TCP connection between the robot and the smart bulb. A Xiaomi Yeelight smart bulb was used for our implementation. The control commands use a JSON string that must contain the lamp identifier, the type of command and its parameters. All command strings must end with "\r\n". These parameters are defined when the *Light* component is inserted into the script. Figure 3 shows the connection scheme between EVA and a smart bulb with identifier equal to 1 and IP address 192.168.1.105.

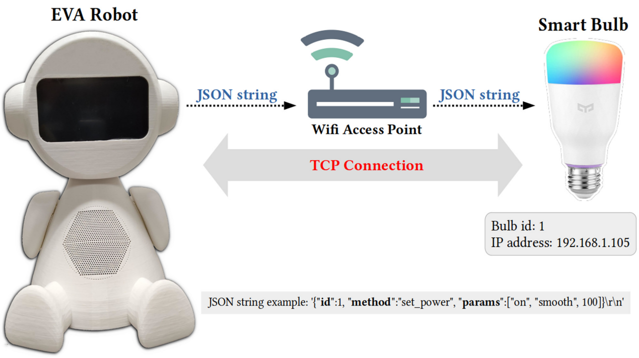

Fig. 3. EVA controlling a smart bulb

4.2 Facial Expression Recognition

Being able to identify the user's emotion through the recognition of facial expressions and using this information within the application is an important facility, which expands the power of the robot's interaction with the user [33]. For this purpose, a second extension was added to VPL giving the robot the ability to recognize facial expressions. A new component was created and added to the visual programming language, named *User Emotion*. It uses the services of an external module developed in Python. Both the Python module and the EVA software run on the same device, a Raspberry PI 4.

Communication between the module and the robot takes over a TCP connection. When started, the face recognition module creates a TCP server that keeps waiting for the connection with the robot. After establishing the connection, the facial recognition module activates the webcam, thus initiating the image capture and facial expression inference processes. At the end of the procedure, the module returns to the TCP client (EVA), a string containing the identified expression. The facial recognition module can return the following expressions: "NEUTRAL", "ANGRY", "DISGUST", "FEAR", "SURPRISE", "HAPPY" and "SAD". In order to obtain the user's facial expression, the robot, through the *User Emotion* component, sends a request to the Python facial recognition module. The process works as illustrated in Figure 4.

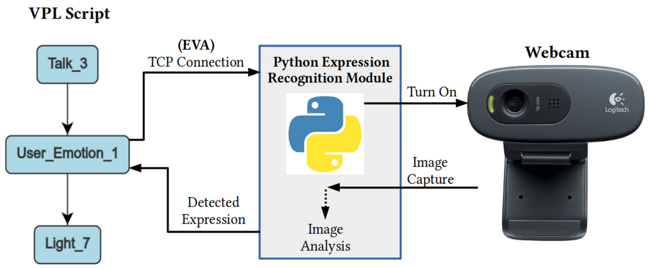

Fig. 4. Communication between EVA and the Facial Expression Recognition module

We have done tests to evaluate the accuracy of the facial expression module to recognize user's emotion when "HAPPY", "ANGRY" and "SAD". We chose those three facial expression types, because they are the ones that can currently be expressed by EVA and will be used in the therapy session we implemented as going to be discussed in the following section. We implemented a program with EVA to recognize 30 facial expressions and two users have tested this program with 5 expressions of each type "HAPPY", "ANGRY" and "SAD". Therefore, considering 30 facial expression recognition events, EVA correctly recognized 24, obtaining an accuracy of 80%. Considering only the "HAPPY" expression, EVA's accuracy was 100%. The tests were performed by two adults. Thus, we considered it as a satisfactory result that can be used in practice.

5 Serious Game for Autism Therapy

In order to serve as the object of evaluation of our proposal, we developed a serious game [1] using the robot's programming language. This game is aimed at assisting in emotion regulation therapies for children with ASD.

During the game design phase, we talked to an ASD expert therapist to discuss our ideas. The therapist recommended us to only emphasize correct answers given by a child. EVA should not say, for example, "your answer is wrong" to the child. On the other hand, EVA should compliment the child when the given answer is fine. This procedure is used in interventions in Applied Behavior Analysis called Discrete Trial Training (DTT) [29]. DTT is a practice that uses instructions with repeated teaching attempts, where each shot presents an answer and a reinforcement to that answer, if appropriate. It is a fully structured training, with a clearly defined beginning and end. This type of discrete experimental intervention increases the likelihood of correct patient responses based on error-free learning. The interaction between therapist and child must be carried out in an environment without distractions, with clear, concise and objective instructions and immediate reinforcement for each correct answer.

We designed a game with three stages. The first stage is called *game of colors*, in which the child is asked to match the colors that are presented by the robot using light sensory effects. In the second stage, called *game of emotions*, the robot presents several facial expressions, while the child tries to identify them. In the third stage, called *imitation game*, the child needs to imitate the robot's emotions with his/her own facial expressions.

Each stage includes three questions. If the child gives an incorrect response, he/she is provided with a new chance to try the correct answer. Stage 1 follows the flowchart shown in Fig. 5.

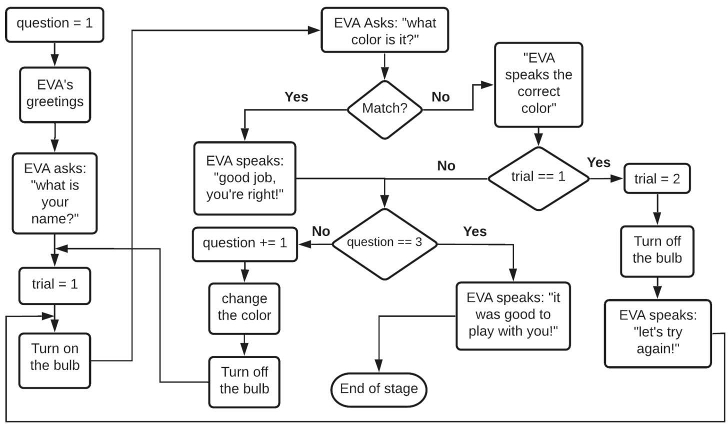

Fig. 5. Flowchart of stage 1 of the game

Using the robot's TTS (Text-to-Speech) feature, a greeting text is transformed into speech and the game starts with the robot greeting the child. Then, the robot listens, waiting for the child's response and through the STT resource (Speech-to-Text) the participant's name is captured, transformed into a string and stored in a variable. Then, the robot, using the *Light* component proposed in this work, presents the first color, lighting the bulb and asking the child to name its color. At this point, using the VPL's conditional testing feature, the child's response, already transformed into a string, is compared with the string that represents the correct answer. If the child gets it right, the robot congratulates him/her and checks if it has reached the maximum number of questions, if not, a new color is selected and the process of presenting is repeated by presenting

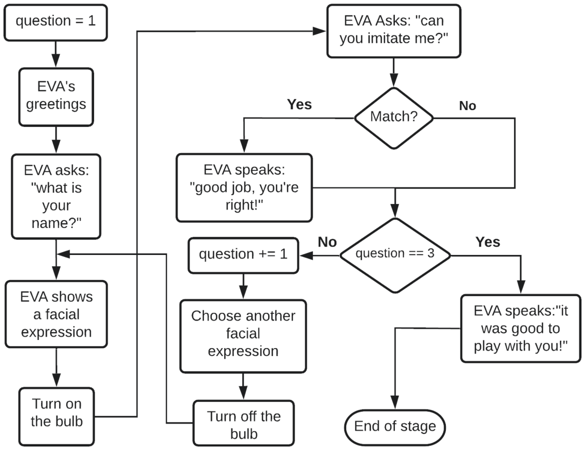

Fig. 6. Flowchart of stage 3 of the game

a new color. If the child makes a mistake, the robot speaks the correct name of the color and gives another opportunity for the child to try again. After three questions, the robot gives a final greeting and the game ends.

Stage 2 follows the same logic as stage 1, but now the child has to recognize the robot's emotions. Therefore, instead of presenting the colors and turning on the bulb, the robot presents emotions through its own gaze expressions. Three emotions are represented: *happy, anger and sadness.*

In stage 3, the logic of the game changes as shown in Fig. 6 . The robot shows a facial expression and asks the child to imitate it. Unlike the previous stages where the robot takes the child's voice as input, in this stage, the robot obtains the child's response by recognizing his/her facial expression using the *User Emotion* component proposed in this work. The robot captures the child's expression and checks if he/she provided the correct response. EVA congratulates the child when his/her imitation is recognized correctly. Three facial expressions are presented in a match.

6 Proposal Evaluation

In order to evaluate our proposal, we used the Technology Acceptance Model (TAM) [15]. This instrument is widely used for evaluating technology applied to healthcare [10]. TAM is based on a questionnaire that assesses Intention to Use the technology by evaluating Perceived Usefulness (PU), "the degree to which a person believes that using a particular system would enhance his or her job performance"; and Perceived Ease-Of-Use (PEOU), "the degree to which a person believes that using a particular system would be free from effort" [5].

As suggested by [22], the game was tested by a Brazilian neurotypical 6-year-old child with the consent of her parents for participating in the experiment. The robot spoke Portuguese in the therapy session. As not all physical parts of the robot were already available, we connected a TV to the Raspberry Pi to show the robot's eyes to the child. However, at the moment, we already have the robot fully built with all the parts that make it up[1]. We recorded videos of the child playing each game stage with EVA[2,3,4], with the consent of her parents. Figure 7 shows the child playing game stage 3.

Fig. 7. Child playing the imitation game

We have adapted a TAM questionnaire for our evaluation and asked healthcare professionals and students to watch those videos and complete the questionnaire. The results are shown in Table 2. Questions for evaluating perceived usefulness (PU) and easy-of-use (PEOU) used a likert scale varying from (1) completely disagree to (5) completely agree.

A group of 48 adult users, aging from 19 to 61 years old, including healthcare professionals and students, such as physicians, psychiatrists, psychologists, nurses, psychopedagogues, educators, speech therapists, and occupational therapists, have evaluated our proposal. All of them have given their consent to participate in the experiment, otherwise they would not be able to watch the videos and answer the questionnaire. We considered professionals with two or more years of experience in taking care of children with ASD as experts and the ones with less than two years of experience, including students, as beginners [23]. Four participants were excluded from the experiment because they had not specified their graduation course. Therefore, our final group was divided into 24 experts in ASD and 20 beginners in ASD. Table 2 shows the average (AVG) and standard deviation (SD) results for PU and PEOU questions, considering the whole group of 44 participants (All) and also specific subgroup results (experts and beginners). Average results greater than or equal to 4,0 indicate participants

1 https://bit.ly/3kNt4u3.
2 https://bityli.com/0ceIRU.
3 https://bityli.com/IhTkIc.
4 https://youtu.be/PU8BLwTkGaw.

Table 2. TAM questionnaire and results

Technology Acceptance Model - TAM									
		All		24 Experts		20 Beginners		Mann-Whitney	
Perceived Usefulness (PU)		AVG	SD	AVG	SD	AVG	SD	U	p-exactly
PU1	The use of technology helps me to do my job	4,39	0,72	**4,46**	0,72	4,30	0,73	206,5	0,217
PU2	The EVA interactive robot would be useful to assist me in therapies for autism spectrum disorder (ASD)	4,23	1,16	**4,25**	1,11	4,20	1,24	238,5	0,486
PU3	The EVA interactive robot would hinder my therapy sessions for autism spectrum disorder (ASD)	2,32	1,39	**2,08**	1,18	2,60	1,60	196	0,154
PU4	The color recognition game would be useful to assist me in therapies for autism spectrum disorder (ASD)	4,25	1,06	**4,29**	1,08	4,20	1,06	218	0,308
PU5	The game to recognize emotions would be useful to assist me in therapies for autism spectrum disorder (ASD)	4,39	0,97	**4,54**	0,88	4,20	1,06	186	0,105
PU6	The game to mimic the emotions of the robot would be useful to assist me in therapies for autism spectrum disorder (ASD)	4,39	0,97	**4,46**	0,88	4,30	1,08	227	0,385
PU7	The EVA interactive robot could help me increase the effectiveness of therapies for autism spectrum disorder (ASD) in recognizing emotions	4,25	1,10	**4,33**	1,13	4,15	1,09	202	0,190
PU8	The EVA interactive robot could be useful in other types of therapies	4,59	0,90	**4,58**	1,02	4,60	0,75	222	0,341
Perceived Ease-Of-Use (PEOU)									
PEOU1	A child with ASD would easily interact with the robot through voice	3,82	1,04	**3,75**	1,11	3,90	0,97	223	0,350
PEOU2	The light effect makes the therapy session with children with ASD LESS attractive	2,57	1,35	**2,71**	1,46	2,40	1,23	207	0,224
PEOU3	**Sound effects make the therapy session with children with ASD MORE attractive**	**4,05**	**1,14**	**3,79**	**1,18**	**4,35**	**1,04**	**169,5**	**0,048**
PEOU4	I think it would be easy to use the robot in my therapy sessions with children with ASD	4,14	1,00	**4,00**	1,10	4,30	0,86	205	0,210
PEOU5	I would need a lot of effort to use the robot in my therapy sessions with children with ASD	2,30	1,32	**2,17**	1,24	2,45	1,43	213,5	0,268

completely/partially agree with the question, and results less than or equal to 3,0 indicate they completely/partially disagree with it.

Analyzing the obtained PU results, we can conclude that most experts agree that Eva would be useful to help them during therapy sessions (PU1, PU2, PU3 and PU8). Notice that PU3 is a negative question where the opposite answer (disagree) is desired and obtained. Experts also agree that the extensions proposed in our work are useful and would help them doing their job more effectively (PU4, PU5, PU6 and PU7).

Considering PEOU results, most experts agree it would be easy to use EVA in their ASD therapy sessions (PEOU4) without much effort (PEOU5). On the other hand, some experts indicated that patients may have difficulty to interact with EVA using voice (PEOU1), depending on the level of social communication compromising they have. Most experts disagree that light effects makes the therapy less attractive (PEOU2), which is a promising result. On the other hand, considering sound effects, experts also have few concerns about it (PEOU3).

We did a statistical analysis using the Mann-Whitney nonparametric test to compare TAM responses between the groups of experts and beginners for each question. Table 2 shows U and p-exactly one tail results in the last column. A one-tailed test is appropriate if we want to determine if there is a difference between groups. We only found a statistical difference in PEOU3 ($U = 169{,}5$; $p < 0{,}05$), where beginners found sound effects make the therapy session more attractive than experts. In all other questions beginners and experts agree on their answers. Analyzing TAM results, we conclude that EVA is useful and easy to use according to healthcare professionals.

Participants have also given suggestions to improve our work, such as: asking the child's age before starting the therapy, including a test phase before the game starts, increasing the time the smart bulb stays turned on, implementing a tablet or smartphone version for EVA, implementing EVA as an avatar, including activities with music, improving the robot's expressions including its mouth besides its eyes, providing games to recognize vowels, consonants and numbers, providing simple math and Portuguese questions, and personalizing the therapy session according to the child. Those suggestions are considered future work.

6.1 Limitations

The accuracy of the robot's facial recognition module depends mainly on two factors: the correct positioning of the child in front of the camera and the lighting conditions in the room. It is necessary for the child to be positioned properly so that his/her face can be captured from the front. It is also very important that the room is well lit.

Another limitation that we noticed, when using the EVA platform, is that there is a delay in the audio capture process in the step that precedes the STT (Speech-to-Text) step. The robot system cannot capture the child's speech instantly after executing the Listen component. Thus, it is necessary to wait, up to two seconds, for the system to be ready to capture the child's voice.

A limitation of our study is that we have used a TV set and not the robot itself to interact with the child in the experiment. We are currently requesting the Ethics Research Committee Approval to conduct future tests with children with ASD. We hope to run those tests in a near future.

7 Conclusion

This work presented an extension of the EVA robot designed for Autism Spectrum Disorder (ASD) therapies. Our proposal enhanced EVA's capabilities for multimodal interaction using light sensory effects with a smart bulb and facial expression recognition with a webcam to recognize children's emotions.

We developed a serious game with three stages that can help emotion regulation therapies for children with ASD. The first stage is a game where the child has to match light colors. The second stage is a game where the child has to match the robot's emotions. And the third stage asks the child to imitate the robot's emotions with his/her own facial expressions. The game was evaluated by 44 healthcare professionals using the TAM model. Results were very promising indicating that they considered our proposal useful and easy to use in therapy sessions for ASD children.

As future work, we intend to show video or images on the screen of the robot. It could be useful for showing people's emotions as part of the game. We are also going to extend EVA's script language to include our extensions for light effects and user emotion recognition to facilitate creating and editing therapy sessions. During the game scripting process, we missed a component in the VPL that could generate random numbers. It is our goal to extend the language by adding this type of component. For some types of scripts, the use of a graphical programming interface does not seem to be the best option. It would be interesting if therapy session scripts could also be written in pseudocode or in XML and could be imported into the robot system. Another idea is to integrate the robot into other mulsemedia applications. In this case, the robot would work as an avatar and could interact with the player using voice, giving tips and motivating the participant, promoting more engagement.

Acknowledgment. This work was partially supported by CNPq, FAPERJ, CAPES, CAPES Print and INCT-MACC.

References

1. Barajas, A.O., Al Osman, H., Shirmohammadi, S.: A serious game for children with autism spectrum disorder as a tool for play therapy. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–7. IEEE (2017)
2. Bevill, R., et al.: Multisensory robotic therapy to promote natural emotional interaction for children with ASD. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), p. 571 (2016). https://doi.org/10.1109/HRI.2016.7451861
3. Cibrian, F.L., Peña, O., Ortega, D., Tentori, M.: BendableSound: an elastic multisensory surface using touch-based interactions to assist children with severe autism during music therapy. Int. J. Hum. Comput. Stud. **107**, 22–37 (2017)
4. Cruz-Sandoval, D., Favela, J.: A conversational robot to conduct therapeutic interventions for dementia. IEEE Pervasive Comput. **18**(2), 10–19 (2019). https://doi.org/10.1109/MPRV.2019.2907020
5. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 319–340 (1989)
6. Fachantidis, N., Syriopoulou-Delli, C.K., Zygopoulou, M.: The effectiveness of socially assistive robotics in children with autism spectrum disorder. Int. J. Dev. Disabil. **66**(2), 113–121 (2020)
7. Feil-Seifer, D., Mataric, M.: Defining socially assistive robotics. In: 9th International Conference on Rehabilitation Robotics, ICORR 2005, pp. 465–468 (2005). https://doi.org/10.1109/ICORR.2005.1501143
8. Feil-Seifer, D., Mataric, M.: Robot-assisted therapy for children with autism spectrum disorders. In: Proceedings of the 7th International Conference on Interaction Design and Children, pp. 49–52 (2008)
9. Hirokawa, M., Funahashi, A., Pan, Y., Itoh, Y., Suzuki, K.: Design of a robotic agent that measures smile and facing behavior of children with autism spectrum disorder. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 843–848 (2016). https://doi.org/10.1109/ROMAN.2016.7745217
10. Hu, P.J., Chau, P.Y., Sheng, O.R.L., Tam, K.Y.: Examining the technology acceptance model using physician acceptance of telemedicine technology. J. Manag. Inf. Syst. **16**(2), 91–112 (1999)

11. Javed, H., Park, C.H.: Interactions with an empathetic agent: regulating emotions and improving engagement in autism. IEEE Robot. Autom. Mag. **26**(2), 40–48 (2019). https://doi.org/10.1109/MRA.2019.2904638
12. Kientz, J.A., Goodwin, M.S., Hayes, G.R., Abowd, G.D.: Interactive technologies for autism. Syn. Lect. Assistive Rehabil. Health-Preserving Technol. **2**(2), 1–177 (2013)
13. Lee, J., Takehashi, H., Nagai, C., Obinata, G.: Design of a therapeutic robot for interacting with autistic children through interpersonal touch. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, pp. 712–717 (2012). https://doi.org/10.1109/ROMAN.2012.6343835
14. Lee, J., Takehashi, H., Nagai, C., Obinata, G., Stefanov, D.: Which robot features can stimulate better responses from children with autism in robot-assisted therapy? Int. J. Adv. Robot. Syst. **9** (2012). https://doi.org/10.5772/51128
15. Marangunić, N., Granić, A.: Technology acceptance model: a literature review from 1986 to 2013. Univ. Access Inf. Soc. **14**(1), 81–95 (2014). https://doi.org/10.1007/s10209-014-0348-1
16. Mitjans, A.A.: Affective computation in human-robot interaction. Master thesis, Centro de Investigación CientÍfica y de Educación Superior de Ensenada, Baja California (2020). http://cicese.repositorioinstitucional.mx/jspui/handle/1007/3283
17. Oei, A.C., Patterson, M.D.: Enhancing cognition with video games: a multiple game training study. PLoS ONE **8**(3) (2013). https://doi.org/10.1371/journal.pone.0058546
18. Olazarán, J., et al.: Nonpharmacological therapies in Alzheimer's disease: a systematic review of efficacy. Dement. Geriatr. Cogn. Disord. **30**(2), 161–178 (2010). https://doi.org/10.1159/000316119
19. Pares, N., Masri, P., Van Wolferen, G., Creed, C.: Achieving dialogue with children with severe autism in an adaptive multisensory interaction: the "mediate" project. IEEE Trans. Visual Comput. Graph. **11**(6), 734–743 (2005)
20. de Paula, G., Valentim, P., Seixas, F., Santana, R., Muchaluat-Saade, D.: Sensory effects in cognitive exercises for elderly users: stroop game. In: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 132–137 (2020). https://doi.org/10.1109/CBMS49503.2020.00032
21. Robins, B., et al.: Human-centred design methods: developing scenarios for robot assisted play informed by user panels and field trials. Int. J. Human Comput. Stud. **68**(12), 873–898 (2010)
22. Salleh, M.H.K., Miskam, M.A., Yussof, H., Omar, A.R.: HRI assessment of ASKNAO intervention framework via typically developed child. Procedia Comput. Sci. **105**, 333–339 (2017)
23. Samonte, M.J.C., Guelos, C.M.C., Madarang, D.K.L., Mercado, M.A.P.: Tap-to-talk: Filipino mobile based learning augmentative and alternative through picture exchange communication intervention for children with autism. In: Proceedings of the 2020 The 6th International Conference on Frontiers of Educational Technologies, pp. 25–29 (2020)
24. Sandoval Bringas, J.A., Carreño León, M.A., Cota, I.E., Carrillo, A.L.: Development of a videogame to improve communication in children with autism. In: 2016 XI Latin American Conference on Learning Objects and Technology (LACLO), pp. 1–6 (2016). https://doi.org/10.1109/LACLO.2016.7751751
25. Santatiwongchai, S., Kaewkamnerdpong, B., Jutharee, W., Ounjai, K.: Bliss: using robot in learning intervention to promote social skills for autism therapy. In: Proceedings of the International Convention on Rehabilitation Engineering & Assistive

Technology. i-CREATe 2016, Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, Midview City, SGP (2016)
26. Santos, L., Geminiani, A., Schydlo, P., Olivieri, I., Santos-Victor, J., Pedrocchi, A.: Design of a robotic coach for motor, social and cognitive skills training toward applications with ASD children. IEEE Trans. Neural Syst. Rehabil. Eng. **29**, 1223–1232 (2021)
27. Shibata, T.: An overview of human interactive robots for psychological enrichment. Proc. IEEE **92**(11), 1749–1758 (2004). https://doi.org/10.1109/JPROC.2004.835383
28. Shibata, T.: Therapeutic seal robot as biofeedback medical device: qualitative and quantitative evaluations of robot therapy in dementia care. Proc. IEEE **100**(8), 2527–2538 (2012). https://doi.org/10.1109/JPROC.2012.2200559
29. Smith, T.: Discrete trial training in the treatment of autism. Focus Autism Other Dev. Disabil. **16**(2), 86–92 (2001)
30. So, W.C., et al.: Who is a better teacher for children with autism? Comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition. Res. Dev. Disabil. **86**, 62–75 (2019)
31. Tan, C.T., Harrold, N., Rosser, D.: Can you copyme? An expression mimicking serious game. In: SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications, SA '13. Association for Computing Machinery, New York (2013). https://doi.org/10.1145/2543651.2543657
32. Tentori, M., Ziviani, A., Muchaluat-Saade, D.C., Favela, J.: Digital healthcare in Latin America: the case of Brazil and Mexico. Commun. ACM **63**(11), 72–77 (2020)
33. Valentim, P.A., Barreto, F., Muchaluat-Saade, D.C.: Possibilitando o reconhecimento de expressões faciais em aplicações Ginga-NCL. In: Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web, pp. 53–56. SBC (2020)
34. Zubrycki, I., Granosik, G.: Designing an interactive device for sensory therapy. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 545–546. IEEE (2016)