HiZeca: A serious game for emotions recognition

Pedro Santos[1], Vinícius Silva[2], João Sena Esteves[1,2], Ana Paula Pereira[3], Filomena Soares[1,2]

[1] Department of Industrial Electronics
[2] R&D Centre Algoritmi
School of Engineering, University of Minho
[3] CIEd, University of Minho
(a74504, a65312)@alunos.uminho.pt, sena@dei.uminho.pt,
appereira@ie.uminho.pt fsoares@dei.uminho.pt

Abstract. To comprehend human behavior can be a very difficult task for children with Autism Spectrum Disorder (ASD). These children have difficulties in social interaction, and they manifest repetitive patterns. Furthermore, they present deficits in imitation which can be directly linked to impairments in social interaction skills. Taking this into account, this paper presents the serious game *HiZeca,* in which a virtual agent (ZECA *Avatar)* is able to interact with a child, in order to promote social interaction and training certain facial movements that will be validated, and that will facilitate imitation and recognition of emotions (content, sad, surprised, among others). In order to validate the system, tests were conducted with typically developing children and children with ASD. The results show that, in general, the game was accepted with a positive feedback from the children.

Keywords: Serious Game, Facial expressions, Emotions, Human-Computer Interaction, Autism Spectrum Disorder.

1 Introduction

The emotional state is mostly defined by the expression of different non-behavioural cues, such as facial prompts. The learning of such facial cues starts from early age when pupils mirror the facial movements of others [1]. Learning by imitation is fundamental to the development of cognitive and social communication behaviours. Furthermore, it directly influences the cognitive empathy that allows a human being to logically comprehend the emotional states of the others, being paramount for successful social interactions. However, children with Autism Spectrum Disorder (ASD) present deficits in imitation which can be directly linked to impairments in social interaction skills [2].

In order to mitigate these impairments, several works in the literature focus on the use of technological tools such as robots with different configurations (from animal-like to humanoid designs) or virtual agents. The humanoid design has been recently more employed by the researchers in different projects in order to promote social inter-

This is an Author Accepted Manuscript version of the following chapter: Santos, P., Silva, V., Sena Esteves, J., Pereira, A.P., Soares, F., HiZeca: A Serious Game for Emotions Recognition, published in Innovations in Mechatronics Engineering. icieng 2021. Lecture Notes in Mechanical Engineering, edited by Machado, J., Soares, F., Trojanowska, J., Yildirim, S., 2022, Springer, reproduced with permission of Springer Nature Switzerland AG. The final authenticated version is available online at: http://dx.doi.org/10.1007/978-3-030-79168-1_35

action with children with ASD. The humanoid design provides a more realistic approach to explore several facets of emotional skills. One example is project ZECA [3] (Portuguese acronym for *Zeno Engaging Children with Autism),* that uses a humanoid robot developed by *Hanson Robotics.* This robot allows to promote social interaction with the children providing the communication and the development of capacities of understanding the emotions. Another more recent example is the *QTrobot* [4] (launched in 2018), created by *LuxAI* to assist children in interpreting facial expressions and promoting social interaction.

However, the technological robots used to support children with ASD are usually expensive and making them widely available in intervention settings may be impractical. A less expensive alternative is using software applications, which design is an area that has been increasing. These applications have several concepts underlined, such as the association of words with images to teach emotions, used by *Sono Flex* [5], or simply allowing the user to choose an item and producing the audio or video that is linked with that item, used by *Livox* [6]. A more advanced concept is proposed by *TippyTalk* [7], which enables an individual with verbal disabilities to communicate by translating pictures into text messages, allowing a child to communicate and express a desire, need, or feeling. Studies conducted with children [8] showed that significant progress has been made in their behaviour regarding relationships with colleagues and teachers, for example.

As addressed in [1], facial expressions play an important role during a social interaction and children with ASD lack the skills to perceive them. For this reason, the development of applications that are capable of aiding in the learning of emotions and the understanding of states of mind becomes a relevant solution. Furthermore, the design of applications with virtual agents that can infer the user non-behavioural cues may offer a unique opportunity to explore and promote the development of social skills in a wider scale.

This paper presents a serious game called *HiZeca,* used to aid in the learning of emotions and to promote social interaction with children with ASD. The game aims at offering a virtual alternative to the physical robot ZECA for interacting with children with ASD, at a much lower cost. It can be used anywhere with a general purpose computer. The main character of the game is ZECA *Avatar*, a virtual agent that a child interacts with using his facial expressions to imitate the avatar's face. In turn, the avatar mimics the child's expressions. Thus, the child's face will be monitored – using a tool previously presented in [9] – in order to verify the execution of his facial movements and to support their improvement, if they are not executed correctly. ZECA *Avatar* will encourage the improvement of possible incorrect movements. The game has three levels of complexity. It explores and fosters the child's ability to mimic, recognize, and infer the emotional states of ZECA *Avatar.*

This paper is organized in four sections. Section 2 presents the proposed serious game; Section 3 shows the preliminary results obtained so far; and the final remarks and future work are addressed in Section 4.

2 The Serious Game *HiZeca*

The serious game *HiZeca* is devoted to mimicking facial expressions and emotion recognition. The challenges that are presented to the user will allow to understand the type of emotion and interpret emotional states. For this, the user's face will be monitored in order to verify the execution of the facial movements.

HiZeca has three levels with increased difficulty and running in three different scenarios: Movement Training, Expression Training and Emotion Identification (Fig. 1).

Fig. 1. Game Levels

In the first scenario, Movement Training, the purpose is to train some facial movements that will help the user in the execution of emotional expressions. This level also serves to familiarize the user to the main character. The movements made by the user are replicated by ZECA *Avatar* in an imitation context by model [10]. In this case, the user serves as a model for the avatar to copy. For this, the user is asked to perform a set of basic movements (Fig. 2).

Fig. 2. Level 1 - Movement Training (in Portuguese).

In the second scenario, Expression Training, various facial expressions associated with five emotions are trained: happiness, sadness, anger, fear and surprise. The user is asked to execute some expressions where the avatar serves as a model. As the user successfully imitates the expression, the avatar moves away becoming smaller, until it disappears remaining only the label of the emotion (Fig. 3). A control bar has been placed strategically so that the user can control the intensity of his/her expression. This scene is an example of learning by deferred imitation [11].

Fig. 3. Level 2 - Expression Training (in Portuguese).

In the third level, Emotion Identification, the story mode was implemented where the child is invited to listen to a set of 15 different stories and identify the associated emotional state corresponding to one of the previously trained. Fig. 4 presents the menu for the emotion selection. In case the answer given by the user is incorrect, the number of lives (options) is decreased and the answer button disappears.

Fig. 4. Level 3 - Emotion Identification: Menu emotion selection.

3 Results

Laboratory tests involving adults were performed in order to test the robustness of the game and validate the implemented methods. Tests were also performed in a school environment, both with typically developed children and children with ASD, in order to verify the acceptance of the tool by children between 6 and 8 years old. Each child was seated strategically, with the face at the level of the camera or webcam.

3.1 Typically Developed Children

The values shown in table 1 are related to the tests performed for level 1 with three typically developed children. A small flaw in the validation of the mouth opening movement was detected. Only one timeout was recorded, which resulted in a detection rate of 66.67%. For the smile movement, two timeouts were recorded, resulting in a lower validation rate, with a value of 33.33%. For the movement of raising the eyebrows, it was also verified that the validation limits should be high, leading to a timeout for this movement. The movement with the highest average response time was to frown. This movement has a 100% validation rate, but it has a high average time due to the difficulty that the children encountered in executing the movement. As expected, it was also in this movement that a maximum response time was recorded: 39 seconds. The computation of the average times shown in Table 1 did not include the timeouts that occurred. The small number of children being tested implies that the fluctuation in the values corresponding to the responses of each child greatly affects mean values and validation rates.

Table 2 shows the values related to the tests performed for level 2 with three typically developed children. Each emotion was tested five times by each of the three children, resulting in a total of 15 validations per emotion. In this second level, the happiness and fear emotions were those that registered a maximum validation rate: 100%. For these emotions the average response times were exactly the same: 6.8 seconds. The value recorded for the emotion of disgust was the lowest of all: 80%. This is an emotion that, although usual, is not one of the most present in everyday life, hence the difficulty in expressing it. Regarding the average times, the lowest value was registered for the emotion of surprise: 6.56 seconds. On the other hand, the highest value was registered for the emotion of disgust: 8.08 seconds.

Table 3 shows the values related to the tests performed for level 3 (story mode) with three typically developed children. In this third level, some wrong answers were recorded in some stories, as well as some repetitions, which happened when the children felt the need to listen again and see the story to give their final answer. For stories 4 and 11, two wrong answers were recorded, this being the largest number of wrong answers in all tests. The story with the highest average response time was story 3, related to angry emotion, with a value of 13.67 seconds. The maximum response time was recorded for story 10. At this level, the children's reasoning ability to quickly associate emotions with the story in question was noted. Of the three levels tested, the third and last level was the one that showed the best performance and acceptance results by these typically developed children.

3.2 Children with ASD

Tests were performed with two children with ASD, who have very different characteristics (named child A and child B).

Table 1. Summary table with average values of the results of tests in school environment for level 1 with typically developed children.

Movement	Detection	Timeout	Average time (s)	Maximum time (s)
Open the mouth	66,67%	1	5,5	7
Lower the head	100%	0	5	10
Raise the head	100%	0	9,67	13
Tilt head to the left	100%	0	15	21
Tilt head to the right	100%	0	11,33	23
Look to the left	100%	0	4,67	6
Look to the right	100%	0	9,67	16
Raise eyebrows	66,67%	1	8,5	15
Smile	33,33%	2	5	5
Wink	100%	0	10,33	20
Frown	100%	0	16	39
Tilt head to the side	100%	0	5,33	8
Look to the side	100%	0	7,33	11

Table 2. Summary table with average values of the results of tests in school environment for level 2 with typically developed children.

Emotion	Validation	Average time (s)	Maximum time (s)
Happiness	*100%*	*6,80*	*19*
Surprise	*93,33%*	*6,56*	*22*
Fear	*100%*	*6,80*	*28*
Sadness	*80%*	*6,70*	*18*
Anger	*86,87%*	*6,85*	*20*
Disgust	*80%*	*8,08*	*27*

Table 3. Summary table with average values of the results of tests in school environment for level 3 with typically developed children.

Story	Wrong answers	Repetitions	Average time (s)	Maximum time (s)
Story 1	*0*	*0*	*11*	*22*
Story 2	*1*	*1*	*7*	*12*
Story 3	*0*	*0*	*13,67*	*17*
Story 4	*2*	*1*	*11,5*	*14*
Story 5	*0*	*0*	*6,33*	*11*
Story 6	*0*	*1*	*7,33*	*16*
Story 7	*0*	*0*	*7,33*	*15*
Story 8	*0*	*0*	*2,33*	*4*
Story 9	*0*	*1*	*2,67*	*6*
Story 10	*0*	*0*	*8,33*	*18*
Story 11	*2*	*0*	*7*	*9*
Story 12	*0*	*0*	*3*	*5*
Story 13	*1*	*0*	*9,67*	*13*
Story 14	*0*	*1*	*5,67*	*8*
Story 15	*1*	*0*	*4,67*	*7*

Child A has a high degree of autonomy. He can speak fluently and read, which promotes the interaction because, at all levels, in addition to sound there are written elements.

Table 4 summarizes the performance of the child A at level 1. At this level, he just failed to correctly perform two movements: tilting his head to the side and raising his eyebrows. He was unable to perform the movement of raising his eyebrows due to difficulties in controlling each isolated movement of the face. For this reason, the maximum time for executing the movement has been exceeded. Another difficulty felt was the distinction between the right and the left. For this reason, movements that involve the distinction between right and left have high response times, as can be seen in Table 4. For example, the movement of tilting the head to the right has an answering time of 28 seconds.

Table 5 summarizes the performance of child A level 2. At this level, where the ability to express emotions is tested, the child's performance was surprising. In addition to being able to express all emotions correctly, he performed each expression in acceptable times. For the expression of angry, for example, the response times were low (never exceeding 5 seconds) which translates to an ease in the expression of the emotion. At the last moment only the name of the emotion remains, and the child was able to express each emotion very quickly. The first time the expression associated with disgust was asked, the child had some difficulties, evidenced by the 28 seconds of response time. But in all the identical requests that followed, response times were low. This means that, after a period of learning, the child was able to easily express the emotion.

Table 6 summarizes the performance of child A at level 3 (story mode). In this third level, there was no story in which the maximum response time was reached or the number of lives exhausted. However, from the recorded data it can be seen that there were three moments in which the repetition of the story was necessary for the child to clarify which emotion was right. For example, in story 5 the child initially selected a wrong answer but, after repeating the story, was able to select the right emotion. As with typically developed children, this third level was the one that aroused the most interest for the child. Since this child said everything he was thinking it was possible to realize that, after hearing the story and having watched the face of the avatar, he associated the expression and context of the story with the right emotion, except in stories 1 and 5. The fact that the child knows how to read also helped when doubts arose regarding the expression or image present, as he was able to associate the emotion with the name of the button.

Child B has many difficulties in concentrating and is a very active child. These complications led to greater difficulty in performing the tests. There were also times when the child fixed his attention and concentration on elements external to the game, resulting in total inattention to the game.

At level 1, the validation of movements was affected by the child's constant movements that caused difficulties in detecting his face. For this reason, the level 1 test was not completed, with only a few movements being tested (Table 7).

Table 4. Summary of the performance of child A at level 1.

Movement	Right Answer	Answering Time
Open the mouth	Yes	00:00:08
Lower the head	Yes	00:00:10
Raise the head	Yes	00:00:11
Tilt head to the left	Yes	00:00:28
Tilt head to the right	Yes	00:00:28
Look to the left	Yes	00:00:20
Look to the right	Yes	00:00:06
Raise eyebrows	No	TimeOut
Smile	Yes	00:00:06
Wink	Yes	00:00:03
Frown	Yes	00:00:02
Tilt head to the side	No	TimeOut
Look to the side	Yes	00:00:04

Table 5. Summary of the performance of child A at level 2.

Emotion	Right Answer	Answering Time
Happiness 1	Yes	00:00:05
Happiness 2	Yes	00:00:10
Happiness 3	Yes	00:00:06
Happiness 4	Yes	00:00:03
Happiness 5	Yes	00:00:03
Sadness 1	Yes	00:00:05
Sadness 2	Yes	00:00:02
Sadness 3	Yes	00:00:05
Sadness 4	Yes	00:00:02
Sadness 5	Yes	00:00:02
Disgust 1	Yes	00:00:28
Disgust 2	Yes	00:00:02
Disgust 3	Yes	00:00:03
Disgust 4	Yes	00:00:03
Disgust 5	Yes	00:00:03
Fear 1	Yes	00:00:04
Fear 2	Yes	00:00:04
Fear 3	Yes	00:00:06
Fear 4	Yes	00:00:11
Fear 5	Yes	00:00:08
Surprise 1	Yes	00:00:04
Surprise 2	Yes	00:00:14
Surprise 3	Yes	00:00:03
Surprise 4	Yes	00:00:02
Surprise 5	Yes	00:00:03
Anger 1	Yes	00:00:04
Anger 2	Yes	00:00:05
Anger 3	Yes	00:00:02
Anger 4	Yes	00:00:02
Anger 5	Yes	00:00:02

Table 6. Summary of the performance of child A at level 3.

Story	Right Answer	Answering Time
Story 3	Repetition	
Story 3	Yes	00:00:33
Story 6	Yes	00:00:17
Story 9	Repetition	
Story 9	Yes	00:00:04
Story 12	Yes	00:00:10
Story 15	Yes	00:00:11
Story 14	Yes	00:00:32
Story 11	Yes	00:00:09
Story 8	Yes	00:00:07
Story 5	No	
Story 5	Repetition	
Story 5	Yes	00:00:13
Story 2	Yes	00:00:11
Story 1	No	
Story 1	Yes	00:00:59
Story 4	Yes	00:00:07
Story 7	Yes	00:00:15
Story 10	Yes	00:00:12
Story 13	Yes	00:00:25

Table 7. Summary of the performance of the child B at level 1.

Movement	Right Answer	Answering Time
Lower the head	Yes	00:00:23
Raise the head	Yes	00:00:26
Tilt head to the left	No	TimeOut
Tilt head to the right	No	TimeOut
Smile	Yes	00:00:39
Frown	Yes	00:00:04
Tilt head to the side	No	TimeOut

Level 2 tests achieved better results. At this level, the child was initially able to express some emotions (Table 8). For the surprise emotion, the child associated the movement of opening the mouth to this emotion having obtained acceptable response times.

At the third and final level this child's behavior changed significantly (Table 9). For the first story presented, two wrong answers were recorded. Then the child was encouraged to listen and see the story again, after which he was able to answer correctly. In the following stories the child's attention and concentration increased significantly. It was necessary to repeat some stories. In these cases, the child could understand the emotion. Response times were better than expected. The acceptance of this third level was also noted and the child's performance was surprising.

Table 8. Summary of the performance of child B at level 2.

Emotion	Right Answer	Answering Time
Happiness 1	No	TimeOut
Happiness 2	No	TimeOut
Happiness 3	No	TimeOut
Happiness 4	No	TimeOut
Happiness 5	No	TimeOut
Sadness 1	Yes	00:00:15
Sadness 2	Yes	00:00:16
Sadness 3	Yes	00:00:02
Sadness 4	Yes	00:00:05
Sadness 5	Yes	00:00:05
Disgust 1	Yes	00:00:08
Disgust 2	Yes	00:00:03
Disgust 3	Yes	00:00:15
Disgust 4	Yes	00:00:12
Disgust 5	Yes	00:00:20
Fear 1	Yes	00:00:06
Fear 2	Yes	00:00:03
Fear 3	Yes	00:00:05
Fear 4	Yes	00:00:07
Fear 5	Yes	00:00:07
Surprise 1	Yes	00:00:05
Surprise 2	Yes	00:00:02
Surprise 3	Yes	00:00:04
Surprise 4	Yes	00:00:16
Surprise 5	Yes	00:00:03
Anger 1	Yes	00:00:28
Anger 2	No	TimeOut
Anger 3	Yes	00:00:19
Anger 4	Yes	00:00:19
Anger 5	Yes	00:00:30

After completing all tests with typically developed children and children with ASD, it can be said that the feedback is positive. The acceptance *HiZeca* of was good at all levels. Level 3 was the level that attracted the most interest from all children.

4 Final Remarks

In order to understand the other's emotional states, it is important to be able to interpret their non behavioural cues, for example facial expressions. This ability allows human beings to adjust their behaviour and react suitably. Consequently, understanding facial expressions correctly and extracting the pertinent social information from them is important for social interactions and communication. Individuals with ASD present deficits in imitation which can be directly linked to impairments in social interaction skills.

This paper presented the serious game *HiZeca,* used to aid in the learning of emotions and promote social interaction with children with ASD. It aims at offering a virtual alternative to the physical robot ZECA for interacting with these children, at a much lower cost. It can be used anywhere with a general purpose computer.

Table 9. Summary of the performance of child B at level 3.

Story	Right Answer	Answering Time
Story 1	No	
Story 1	No	
Story 1	Repetition	
Story 1	Yes	00:00:17
Story 4	No	
Story 4	Repetition	
Story 4	No	
Story 4	Yes	00:00:51
Story 13	Yes	00:00:09
Story 2	Yes	00:00:12
Story 5	Repetition	
Story 5	Yes	00:00:17
Story 3	Yes	00:00:12
Story 6	Yes	00:00:07
Story 14	Repetition	
Story 14	Yes	00:00:02
Story 7	Yes	00:00:05
Story 10	Yes	00:00:05
Story 8	Repetition	
Story 8	Yes	00:00:12
Story 15	Yes	00:00:03
Story 11	Yes	00:00:04
Story 9	Repetition	
Story 9	Yes	00:00:23
Story 12	Yes	00:00:06

The game, with three levels of complexity, explores and fosters the child's ability to mimic, recognize, and infer the emotional states of virtual agent ZECA *Avatar,* the main character.

Preliminary tests were conducted in a school environment, both with typically developed children and children with ASD, in order to verify the acceptance of the tool by children between 6 and 8 years old.

After completing all tests with typically developed children and children with ASD, it can be said that the feedback is positive. The acceptance of *HiZeca* was good at all levels. Level 3 was the level that attracted the most interest from all children.

Future work considers optimizing further the system by improving the automatic recognition of facial expressions, and conducting tests in a school environment with a larger sample of children with ASD, in order to find out if the game can actually be a complementary intervention tool. With this game, children are expected to promote their cognitive abilities and improve their ability to communicate and interact with peers.

Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. Vinicius Silva also thanks FCT for the PhD scholarship SFRH/BD/ SFRH/BD/133314/2017.

References

1. Piaget: Piaget and His School. Berlin, Springer Berlin Heidelberg, 1976.
2. Ingersoll, B.: The Social Role of Imitation in Autism. Infants Young Child. vol. 21, no. 2, pp. 107–119, 2008.
3. Silva, V., Soares, F., Sena Esteves, J.: Mirroring Emotion System – On-line Synthesizing Facial Expressions on a Robot Face. 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Lisbon, Portugal, 18-20 October, 2016.
4. LuxAI, QTrobot - LuxAI.com, http://luxai.com/qtrobot/, last accessed December 6, 2018.
5. Tobii Dynavox logo, Sono Flex for Communicator 5 - Tobii Dynavox, https://www.tobiidynavox.com/software/content/sono-flex- for-communicator-5/, last accessed December 7, 2018.
6. Livox, Livox: para pessoas com deficiências e transtornos de aprendizagem, http://www.livox.com.br/pt/quem-somos/#sobre, last accessed December 6, 2018 (in Portuguese).
7. TippyTalk - Comunicação Instantânea para Pessoas Não Verbais. (n.d.), http://www.tippytalk.com/, last accessed January 16, 2019 (in Portuguese).
8. Ferreira, V.: Utilização das tecnologias em crianças com perturbações do espectro autista em contexto da Prática de Ensino Supervisionada. Universidade Católica Portuguesa, Centro Regional de Braga, Faculdade de Ciências Sociais, Braga, 2013 (in Portuguese).
9. Santos, P., Silva, V., Soares, F., Simões, A.: Facial Virtual Tracking: A System to Mirror Emotions. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
10. McLeod, S.: Bandura - Social Learning Theory. 2016.
11. Moura, M.L., Ribas, A.: Imitação e desenvolvimento inicial: evidências empíricas, explicações e implicações teóricas. Estudos de Psicologia, 2002 (in Portuguese).