Embodied Conversational Agent for Emotional Recognition Training

Karl Daher*, Zeno Bardelli†, Jacky Casas*, Elena Mugellini*, Omar Abou Khaled* and Denis Lalanne†
*University of Applied Sciences and Arts Western Switzerland, Fribourg, Switzerland
Email: firstname.lastname@hes-so.ch
†University of Fribourg, Switzerland
Email: firstname.lastname@unifr.ch

***Abstract*—Avatars are known in the world of video games, where heroes with specific characters, attributes and powers are assigned to players. However, avatars are evolving and reaching domains like companions, assistants and tutors. These avatars now use speech, facial expression, body language or text to interact with humans. When we say interaction, we say emotional expression and empathy. Avatars are still short in the emotional and empathic world; they cannot express nor share emotions. In this paper, we research the emotional avatar world, and we present the Anthropomorphic Chatbot for Emotion Recognition (ACER), an empathic friend companion designed for children. The goal of ACER is to teach children about emotions by expressing them through facial expressions and body language while texting through a chat. An experiment was held to test the avatar effect. Qualitative and quantitative results show users positive emotions tending towards having a chat with ACER with facial and body expressions instead of only ACERs chatbot.**

Keywords–HCI; ECA; Conversational agent; Avatar; Emotions.

I. INTRODUCTION

Humanising technology, having a human-computer interaction similar to human-human interaction, is driving researchers and companies in recent years to develop Artificial Intelligence (AI). We can mention AI-driven conversational agents, known as chatbots, avatars, assistants, and humanoid robots with human-like functionality and purpose. Chatbots started in 1950 with Alan Turing wondering if a computer system can communicate in an equivalent way as a human [1] which later led to the Turing test used to test AI-driven conversational agents. In practice, chatbots are programs that allow the user to interact with the machine using natural language. Chatbots development is a growing field, especially with intelligent personal assistants like Siri and Cortana, which are well known to most. The applications of those agents are manifold: they can act as virtual assistants to help the users within an online store by answering their questions, or by booking him a flight online to checking their balance of a bank account [2]. There are also great possibilities in the areas of customer service, but also health [3] and coaching [4].

The most common category of chatbots involves interaction via keyboard, through an interface similar in every way to that of a chat program, and the conversation consists of an exchange of orders. However, those characteristics undermine the naturalness of the conversation itself. In a conversation between two human beings, there is much more than just a conversational expression between the two parties. There are numerous communicative behaviours complementary to the meaning of words, divided into two categories, verbal and non-verbal communication. We can cite, for example, the tone, frequency, and amplitude of the voice or pauses between words from the verbal category. On the other hand, the non-verbal communication category beholds other ways of interaction, where it is generally defined as the aspect of communication that is not expressed in words [5]. For example, facial expression, hand gestures, head movements, gaze, and body posture are involuntary and voluntary behaviours that are an integral part of a conversation between two humans.

Chatbots using text as a means of communication convey emotions in a way that is not practicable, albeit the evolution made in chatbots to support emojis and Graphics Interchange Format (GIFs). Nevertheless, these chatbots still fall short of human-human communication; the result of chatbot-human communication is still direct, cold, impersonal, and unrealistic. To overcome these limitations, many people have explored the possibilities provided by Embodied Conversational Agents (in short ECAs). ECAs employ gestures as the body, hands, and legs movements, mimics as the facial expressions macro-micro expressions and speech to communicate with the users [6]. These features make the interaction more realistic and more humanised. However, researchers and developers are still testing prototypes to find the ideal avatar that can be perceived and treated as human by the user. In this vision, the work on including emotions and empathy within machines started to evolve, but yet many drawbacks and dilemmas still exist and need more deepen exploration and examination. Machines should be able to show empathic capabilities and understand its users and their needs [7].

Within the spectrum of avatars, emotions, and chatbots, we present Anthropomorphic Chatbot for Emotion Recognition (ACER), an embodied conversational agent to teach emotion recognition. ACER's long term goal is to become a friend and a tutor. He is designed for kids and people facing difficulties in understanding and expressing emotions. Alexithymia is "conceptualised as a cluster of cognitive traits which include difficulty identifying feelings and difficulty describing feeling to others" [8] and it is present in "approximately 10% of the population with significantly higher incidence levels within autistic populations 50%" [9]. The goal of ACER is to teach how emotions are expressed using facial expressions and body language while using a chat as a communication tool. In this article, we will present the related work for conversational agents and avatar in general, then briefly present empathy and emotions. Moreover, we present avatars that include and use empathy and emotions, and we end with an analysis and synthesis. In Section 3, we present the prototype, where we develop the architecture and its usage. Further, we give the details of the experiment in Section 4, followed by the qualitative and quantitative results in Section 5. We wrap up this article in Section 6 presenting the future work, and the conclusion in Section 7.

 ISBN: 978-1-61208-761-0

II. State of the art

In the early 2000s, the problem of Embodied Conversational Agents (ECAs) was studied by researchers. Many articles have been written proposing scenarios for the use of such virtual entities and what features they need to have in order to be as credible as possible. In the next sub-sections, we will present research over conversational agents and avatars, following an introduction about empathy and emotions, adding avatars that use empathy and emotions, and ending this section with an analysis and synthesis.

A. Conversational agents and avatars

One of the first conversational agents with an avatar was Gandalf [10]. Gandalf is a virtual humanoid who allows simple conversations. Equipped with a face and a hand, it integrates expressions and gestures in its dialogues. Besides, it reacts adequately to misunderstandings, showing uncertainty and hesitation. Gandalf uses a microphone and sensors to perceive the movement of the user's body and eyes. In this way, it is also able to read non-verbal aspects of communication.

Another early example is Real Estate Agent (Rea) [11][12]. It has a database containing data and images about homes and apartments in Boston. It can then share this information with the user, acting just like a virtual real estate agent. Its creators are committed to providing it with various features to make a conversation with her as natural as possible. It is capable of superficial small talk. It can take a turn during a conversation and also to both use and react to non-verbal communicative behaviours. For example, it can understand when the user raises a hand to ask a question.

A way to design a chatbot with an avatar, speech synthesis, and speech recognition is described by Angga et al. [13]. A conversation takes place as a cycle of separate operations. The user speaks into the microphone, and the program translates the audio into text. At this point, the chatbot API generates the appropriate response, always in the form of text. The text is used both to generate the spoken response and the 3D avatar's behaviour.

The search for realistic behaviour has been the subject of study by Cassell and Vilhàlmsson [14]. In this case, it was not for conversational agents, but rather avatars of users in a three-dimensional virtual world. Important details are the movement of the mouth associated with what is said, the ability to speak and continuous, and involuntary movements like raising eyebrows, head inclination, and blinking eyes. Emotions can also be expressed both by facial expression and body movements. Bringing all these elements together is essential to achieve a credible avatar that can communicate more than just words.

B. Empathy and emotions

Empathy is an essential factor in everyday life; it fosters strong relationships and collaborations between individuals [15]. Machines, robots, chatbots, and avatars that adopt the concept of empathy earn more trust towards the human [16]. The concept of empathy is defined in many ways in research, Omdah divided empathy into two parts, affective empathy, and cognitive empathy [17]. Cognitive empathy is the understanding of other's emotional states, while affective empathy is the response to other emotional states. Another psychological definition of empathy is "putting yourself in the shoes of others", where it is elaborated as taking the position of the other mentally, trying to feel the emotional states he is going through based on personal experience [18]. While Davis defined empathy as the following: "Empathy is a set of constructs having to do with the responses of one individual, to the experiences of another. These constructs specifically include the processes taking place within the observer and the affective and non-affective outcomes, which result from those processes" [19]. When talking about empathy, communication has to be considered, where interaction is included. Roa-Seïler and Craig mention that empathy is an interaction between two individuals who share each other's experiences and feelings [20]. An interaction can cause a continuous development of emotions, thanks to the relationship between the interactants [21] at the same time emotions can affect our behaviours, choices, mood, and our well being in every-days life [22]. Paul Ekman pointed out the six basic categories of emotions that consist of anger, disgust, fear, happiness, sadness, and surprise. These emotions are shared among all humankind around the world and are universal across many cultures [23]. The emotional state of the individual is usually expressed by the face and the body language. However, it becomes biased toward the emotion expressed by the body when both convey conflicting emotional information [24]. Empathy and emotions are two essential research subjects to have an empathic machine.

C. Avatars, empathy and emotions

Modern avatars are starting to include emotions, human-like capabilities, empathy, and emotional behaviours, from the text expressed to the speech tonality, moving towards the gesture shown by the body or the facial expressions that are programmed. However, all still have many drawbacks and major challenges, like shortage of quality training data, the balance between emotion level and content level responses, a fully end-to-end experience, or even modelling emotions throughout conversations [2].

Poggi et al. [25] wrote about Greta, a virtual talking head. It is capable of conducting social conversations. When Greta expresses something, the process of generating behaviour has three phases. First, it generates the sentence with which to respond, based on factors like personality, culture, emotions and age. Then it uses a sort of tag system to identify the right non-verbal behaviour, and finally, it expresses it through her avatar.

It is believed that the display of a conversational agent improves interaction for the user. The reason is that an anthropomorphisation of the chatbot takes place. In order to verify the correctness of the information, an experiment was conducted in which a conversational agent acted as a tutor, instructing users on how to use an interface [26]. In particular, the agent was presented in three formats: with a realistic human avatar, with a monkey cartoon-like avatar, and without an avatar. In both cases with the avatar, the agent proved to be more efficient.

The usefulness of ECAs in Clinical Psychology is being studied [27]. The field is in an early stage, so there are mostly prototypes not ready for evaluation. Most of the proposed treatments deal with autism, particularly for training social skills. It is not yet clear whether ECAs are effective, but there

 ISBN: 978-1-61208-761-0

Figure 1. Facial expressions of the avatar [32]

is certainly a great interest in developing new technologies of this kind.

However, the display of a virtual agent is not always considered positive. A very realistic appearance can influence users' expectations upwards. The more advanced it seems, the more complex and realistic the interaction is expected to be. When these expectations are regularly disappointed, the impression on the agent is very negative [28]. Moreover, according to another study, it turns out that users can better receive a chatbot without an avatar because it does not cause the famous uncanny valley effect [29].

Samuela, another avatar developed by Roa-Seïler [30], has the purpose of being part of the "home of the future", it lives with the owner and can deliver comfort and encouragement. A cooking companion application was developed using Samuela allowing people to choose between different dishes. Samuela has a female face and style. It has a pretty character, a female voice. It is emotionally expressive, can respond to questions, is useful, helpful in every-days life, and acts like a companion. Samuela is excepted to have human behaviour [20].

An empathic companion was developed at the National Institute of Informatics in Tokyo, Japan [31]. This companion consists of a character-based interface that is used to accompany the user in the setting of a virtual job interview. The user physiological signals are taken into consideration in real-time, and the response of this interface is the states that the human is going through. These states are analysed from the physiological signals.

In 2012, a pedagogical agent was developed [32]. This agent has characteristics like facial expressions and body gestures. The users used the mouse-clicking to interact with the avatar while reading, and the avatar motivates them accordingly. The purpose of the avatar was to motivate and encourage the user to make more reading effort. The facial expressions of the avatar can be seen in Figure 1.

The technology aims towards solving practical problems we face. The research was developed for the people who have schizophrenia, who have difficulties in recognising emotions in other facial expressions. These difficulties decrease their abilities for social interaction and thus their integration. For this purpose, the authors created a virtual realistic-looking avatar for assessment of emotion recognition deficit. The experiment held took into consideration the avatar and static images into the recognition of the set of facial expressions [33].

Another important domain where researchers are trying to find solutions is eldercare. The goal is to automate and facilitate elderly lives and at the same time, keep emotional bonding and empathic interactions around them. A research was conducted to create an autonomous conversational agent system that can simulate human-like affective behaviour and act as a daily companion for adults at home. This avatar includes speech recognition, text to speech, and a graphical touch interface. The primary purpose of the companion is to support older adults with many functionalities like locating objects, creating reminders, and orientations with household activities [34].

These days with the technology evolving and the ability to create applications on the Apple App Store or Google Play Store, we find many applications that were developed to help children with difficulty in recognising emotions. In particular, there are several applications for smartphones and tablets that set themselves precisely this goal. Sung et al. [35] did research work testing the various iOS and Android applications available, in particular about facial emotion recognition. Here are a few examples:

- Autimo (developed by Auticiel) offers three different mini-games: associate photos of people with the same expressions, find a different expression among many and guess the emotion shown in a photo.
- CopyMe (developed by Games Studio) uses the web-cam to asks the user to mimic the expression shown in a simple picture.
- Emotions 2 (developed by I Can Do App) contains images of people with different expressions. With these images, it offers five types of exercises, including associating the photos to a scenario or a label based on a scenario.

The essential features for these applications are ease of use and immediate feedback, to keep the user's interest and degree of gratification high.

D. Analysis and Synthesis

In this state of the art, a review was created on many conversational agents and avatars, targeting mainly the avatars developed with emotion and empathic interaction. Each of the systems presented has its advantages as well as disadvantages. With the advancement of the technology, these avatars will continue to develop and become more complex, especially in their graphical side as well as their conversational and understanding capabilities, moving towards human-like companions. The research conducted led that the world of avatars is advancing since its early days. Now, this advancement will keep evolving; we can see that some of the oldest companions use text as a way of communication and interaction. On the other, we find others avatars that use static images to represent facial expressions.

Moreover, videos have been used to support facial or body movements. Lately, we found agents that use body gestures and facial expressions combined. The aim is to combine all interaction techniques to make avatars as affective and human-like as possible. For that, in the next section, we will present ACER, an avatar designed for kids, that fusion the communication and interaction techniques used by the human. ACER utilise facial and body language as well as text chats to interact with the user. ACER aims to become a tutor to teach kids about emotions and how is it expressed in the face and the body language while interacting using text. ACER is designed to have different interaction methods combining the multiple methods seen in state of the art. ACER is designed to become a friend able to handle a conversation about specific topics and expressing it through emotional responses using his empathic behaviour.

 ISBN: 978-1-61208-761-0

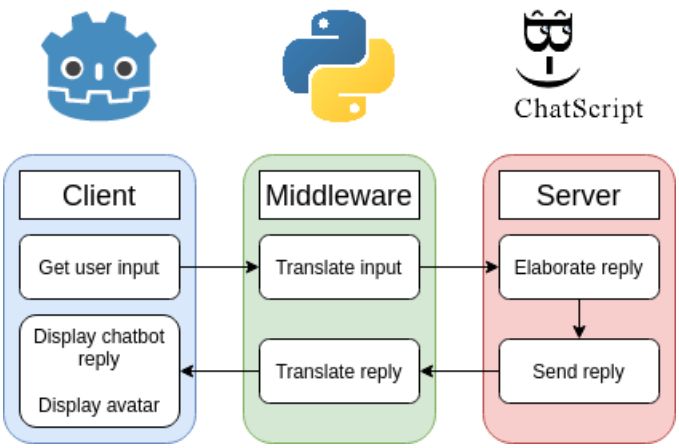

Figure 2. Architecture of ACER, role of its components and message exchange via TCP tunnel. The logos from left to right represent the Godot Game Engine, the language Python and the ChatScript Engine.

III. The Prototype ACER

ACER is an embodied conversational agent whose task is to help children to train themselves to recognize emotions. Its name is an acronym for Anthropomorphic Chatbot for Emotion Recognition. It can show six emotions: calm, happiness, anger, sadness, fear, and disgust. ACER is still a prototype; its purpose is to show how such software could work and to provide a complete framework from which to start. In the following sections, we will first describe the architecture of the prototype and then explain how it works in practice.

A. Architecture

ACER runs on Linux operating systems and consists of three software running at the same time. They are a local server that hosts the chatbot, a client and middleware to communicate between the two. Communication between the parties is done via Transmission Control Protocol (TCP). Figure 2 summarises the architecture and the exchange of messages. It also indicates the language or engine used to develop the components.

The task of the server is to host the chatbot. In practice, it is the place where user input is examined, and answers are processed. To write the server for ACER, ChatScript, a powerful chatbot engine, was used. ChatScript is a free and open-source chatbot engine created and maintained by Bruce Wilcox. Among its salient features are natural language processing, relatively compact and clean syntax, word generalisation, topic encapsulation, and pattern matching. The server manages both ACER's replies and emotions. The emotion model is still basic. In practice, each response that the chatbot can give is tagged with an emotion corresponding to it. Since the bot communicates in plain text, the convention is that each response begins with a 3-letter tag that indicates the emotion. In addition, the 4th character is a hyphen to improve readability while browsing the chatbot script file.

To develop the client, we used Godot, a free and open-source game engine. The reasons for using a game engine are the availability of means both to create an interface and to display an avatar and the ease of interaction between the different components. The client basically takes care of receiving the user's input to send it to the chatbot and show the response. It can be divided into three components: the chat,

Figure 3. The chatbot client showing a calm ACER conversating with the user.

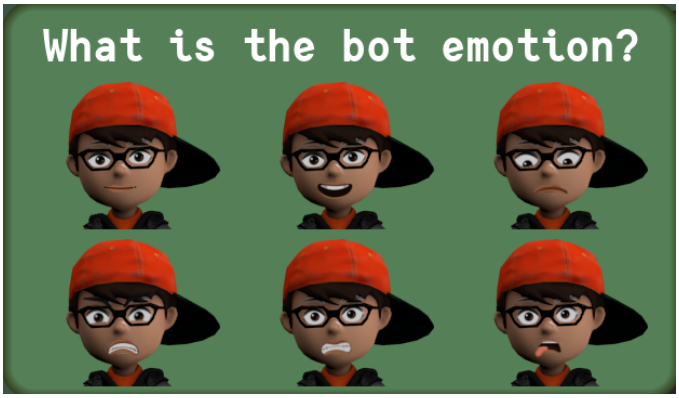

Figure 4. Guess the emotion? The facial expressions of ACER from left to right, top to bottom: calm, happiness, sadness, anger, fear, disgust.

the avatar, and the window that is prompted when asking the user to match the facial expression with the emotion of ACER. The first two of those components are shown in Figure 3.

The chat looks like a classic client. At the bottom, there is space to input text messages and send it, while above a scrollable window for messages history. Each time a message is sent or received the scrollbar reaches the bottom. The bot messages have a different colour.

The avatar is a kid with a hat and glasses. Its cartoon look is meant to appeal to children. Depending on the emotion it simulates, it adopts a different animation for the expression of the emotions. In addition to animations related to emotions, it has one to greet the user when he opens the program or is about to leave. The emotion and animation switching is triggered when the client receives a message from the server. We check the tag contained in the reply, and if it corresponds to an emotion different than the current one, the switch happens. The 3D model and its animations have been downloaded from Mixamo, a store of 3D models. Adobe Incorporated created them. However, due to compatibility issues, the avatar that the user sees in the client window is not a three-dimensional model. These are pre-rendered images, so it is actually a two-dimensional animation.

The last component of the client is the window mentioned above (see Figure 4). When it appears, the user must click on the face with the expression associated with the emotion simulated by the avatar. The window appears when the bot

 ISBN: 978-1-61208-761-0

changes emotion and is positioned so that neither the last text received nor the avatar is hidden. After the user clicks, a symbol appears to immediately give feedback on whether the answer was correct (green check) or not (red cross).

Finally, there is the middleware written in Python. Since ChatScript only receives messages in a specific format that it is not clear if achievable using GDScript, Godot's scripting language, this program receives messages from the client and adapts them for the server. It also takes care of restarting the server, selecting the right chatbot, and building it.

B. Usage

In order to run ACER, it is necessary to run the three software separately in the following order: server, middleware, and client. From here on, the user needs to watch only the client. After a few seconds of loading, the client opens, showing ACER greeting the user with both gestures and text message. At this point, the conversation can begin. Every time ACER changes emotion based on the text, the avatar will show a different animation, and a window will appear asking the user to compare a facial expression with the current body language expression of the bot in addition to the discussion context. The top left corner shows the results of the session, indicating how many correct answers and the number of questions there were.

The interface is designed to be intuitive and straightforward so that a child can use it. The only interactive elements are the window to enter text and faces of the bot with expressions when it pops-up. Regarding the recognition of non-verbal communication, ACER works on three levels.

The idea is to have a conversation as natural as possible, in any subject possible, yet for this prototype, the subjects were limited for more accuracy during the experiment. ACER tries to lead the conversation so that it is clear to the user what kind of answers they can give. For example, it gives multiple options or asks questions that can only be answered by yes or no. Below is a small example of the discussion that can be held with ACER.

IV. Experiment

In order to subject ACER to an initial test, we conducted an experiment with 20 users. This preliminary study aimed at testing the user experience, the prototype testing was conducted with people aged between 21 and 30 years old that were recruited from the authors' circle of acquaintance. This group of people was chosen for a mature evaluation of the first prototype. In particular, we were interested in the empathic and emotional reactions of the users. The goal of this test was to verify the added value of the avatar by analysing the user experience. Due to COVID-19 lockdown circumstances in Switzerland, test sessions were organised remotely via desktop sharing and video conferencing software. Future testing will be conducted on children to evaluate the effectiveness of the system.

The experiment consists of using ACER for five minutes with two different modes. The first mode includes the usage of ACER chatbot only, by that we mean the user will have a chat with ACER without having any facial or body language responses. ACER will be changing his emotions based on the discussion and the user will have to guess which emotions ACER is having. The chats were aimed at certain subjects to reduce the scope, and not have a haphazard discussion.

The second mode includes ACER's facial expressions as well as its body language. The user will be able to chat with ACER, but this time ACER will share its emotions through facial and body language expressions. Whenever ACER changes emotion, the user will have to know which emotion ACER is expressing. In this way, ACER will be able to tutor the user using it about how emotions are expressed. The chat is part of the prototype, as described in Section III. Thus, in this experiment, the facial and body expressions were removed. Since the situation is quite unusual during the COVID19 crisis, the tests were carried out remotely using a videoconferencing tool.

After each mode, each of the users was asked to fill a survey which evaluates the qualitative and quantitative aspects of the emotions felt towards ACER. Positive and negative emotions are to be evaluated in the next section. For the quantitative survey, meCUE 2.0 [36] questionnaire was used to evaluate the key components of the user experience. On the other hand, another survey was prepared to understand the qualitative emotional effects of the embodied ACER chatbot compared to the text-only ACER chatbot.

V. Results

In this section, we present the results of the experience which hold the meCUE 2.0 [36] questionnaire for the quantitative analysis and the qualitative survey where users are asked about their personal interaction with ACER. Details can be found in Table I. The questionnaire meCUE 2.0 is dedicated to the user experience of interactive technical products, in our case ACER. The questionnaire is divided into multiple modules; we are interested in module III User emotions to study the emotional effect of the body language and facial expressions addition to ACERs chatbot.

A. Qualitative Analysis

A questionnaire about how ACER is perceived was submitted to the subjects. They had to state if they agree with several sentences with respect to the chatbot without the avatar and the chatbot with it. The questions and the distribution of answers are summarised in Table I.

In general, it can be seen how the avatar version has generated better impressions for each of the sentences. 90% of the users thought that learning from the avatar has a positive effect, while only 35% stated the same for the text-based chatbot. More than half of the users said that the interaction with ACER is natural in both cases, showing appreciation for how easy it is to converse with a chatbot. Still, the percentages scored are 55% without the avatar and 80% with the avatar. The 95%, respectively the 75% of the users said that the chatbot with avatar has a personality, respectively looks clever and competent. Those numbers are 35%, respectively 35% for the chatbot without avatar. To 10 people over 20, the chatbot without avatar is boring, but only one person said the same about the chatbot with the avatar. Finally, the text-based conversational agent looked emotionless to 75% of the users, while the embodied conversational agent only made that impression in 20% of the subjects.

B. Quantitative Analysis

According to the measurements made with module III of the meCUE 2.0 questionnaire, the chatbot with the avatar

 ISBN: 978-1-61208-761-0

TABLE I. SUBJECTS WERE ASKED IF THEY AGREE WITH THOSE SENTENCES WITH RESPECT TO THE CHATBOT WITHOUT AVATAR AND THE CHATBOT WITH AVATAR.

Sentence to agree with	No Avatar	Avatar
Learning from ACER has a positive effect.	35%	90%
The interaction with ACER is natural.	55%	80%
ACER has personality.	35%	95%
ACER looks clever and competent.	35%	75%
ACER is boring.	50%	5%
ACER looks emotionless.	75%	20%

causes more positive emotions and less negative emotions in users. In terms of mean and standard deviation, over a scale of 7, the results were the following: positive emotions without avatar $\bar{x} = 3.31$, SD$= 1.26$, with avatar $\bar{x} = 3.37$, SD$= 1.25$; negative emotions without avatar $\bar{x} = 2.14$, SD$= 0.84$, with avatar $\bar{x} = 1.95$, SD$= 1.10$. We also performed two-tailed t-test with $p = 0.05$ to verify if the difference is statistically significant. For positive emotions, we obtained that the t-value is 0.20283, while the p-value is 0.840349. For negative emotions, we obtained that the t-value is 0.45388, while the p-value is 0.652497. There is no statistical difference. Even if the avatar with the body language and facial expressions achieved a better result, we could not state that it is better than the text-based chat. One of the reasons for not achieving a statistical result might be the small population size.

Nevertheless, the overall analysis shows a mean result of 2.0 over 5 for the text-based avatar, while embodied ACER had an overall score of 2.9 over 5. This difference is presented in different metrics and features received from the questionnaire, first in terms of usability we can see a clear difference between a 4.68 and a 5.63 score over 7 for the usability of the embodied agent over the text-based. Another feature, commitment shows a 4.08 for text-based and 4.62 for embodied ACER.

The results of the meCUE questionnaire show that all the features of the embodied ACER received a score higher than the text-based ACER. More participants will be recruited in the future to have a larger dataset and more precise result.

VI. CONCLUSION AND FUTURE WORK

In this article, we researched avatars that use emotions and empathy for defined purposes. From state of the art, we were able to synthesise that it is still a growing field and research is yet improving and developing. Avatars, among all other technologies, still fall short when compared to humans. With the aim of humanising technology, these avatars need to have human abilities each in their way based on its purpose. Avatars in state of the art use many ways of communication like text, speech, facial expression or body language sometimes these techniques are combined.

In the aim of having empathic avatars, avatars that can express its emotions based on the context of a conversation, we designed ACER, a friend and a tutor. ACER, the Anthropomorphic Chatbot for Emotion Recognition training, a companion with a tutoring purpose is designed to train humans to understand the emotional expressivity of the facial and body language while chatting with a chatbot. The design of ACER was made to be easy to use and easy to understand. An experiment was conducted where the user is asked to connect body language and facial expressions emotions according to the conversation being held. Quantitative and qualitative results show that the users tend to have a conversation with ACER with its body and facial representation rather than only ACER text-based chat. Although still in the prototype stage, ACER has all the bases to grow.

Inspired by several smartphone applications, ACER is proposed as a new tool to help people, and children in particular, with difficulty in reading non-verbal language. ACER allows them to relate expression, body language, and context of a discussion. As a next step towards improving user experience and testing, ACER will be tested with kids and people facing emotion recognition problems to improve the quality of interaction. ACER provides a solid foundation; many technical improvements are set for development. First, we will start with the chatbot, where the server will have more fluency and broader topic scope discussions. More into chat, ACER will be designed to have an empathic behaviour when discussing with the user, being able to analyse the emotion that the user is feeling through the text provided. Another improvement that we are aiming for is to make ACER personalised. For example, it will save the discussions for each user, analyse and learn from them. First, the chatbot brain will develop over time, and the second analysis will be made to see the progression of the user. Design-wise, we would like to improve the ACERs presentation by personalising the shape of ACER and the design of the application. Quizzes and games will be integrated for long term interaction purposes. In this way, the user will be able to get rewarded as well as test his abilities. Levels and complexity of the game and quiz will be adapted to the learning process and capabilities of the user. Another fundamental tool that we would like to add is speech recognition, where the user will be able to interact with ACER either through text or using speech. From a more technical point of view, it would be interesting to port the program to other operating systems and devices, in particular on smartphones, since they are the most accessible devices by kids. A design for ACER on a smartphone will be needed because of privacy concerns and performance matters of smartphones.

Another point that is set for the future is to replace the pre-rendered animations with a real three-dimensional model. It would allow a smooth scrolling between the animations thanks to interpolation, and above all, it would make the program smaller in terms of size on disk. Indeed, many images are needed to maintain high-quality animations. The loading time would also be faster. Finally, we can imagine these three-dimensional models to be rendered in the virtual world using Virtual Reality (VR) where ACER will be having more specifications as height and size and having more real interaction with the user.

REFERENCES

[1] A. M. Turing, "Computing machinery and intelligence," in Parsing the turing test. Springer, 2009, pp. 23–65.

[2] T. Spring, J. Casas, K. Daher, E. Mugellini, and O. Abou Khaled, "Empathic response generation in chatbots." in SwissText, 2019.

[3] K. Daher, J. Casas, O. Abou Khaled, and E. Mugellini, "Empathic chatbot response for medical assistance," in Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, 2020, unpublished.

[4] J. Casas, E. Mugellini, and O. A. Khaled, "Food diary coaching chatbot," in Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2018, pp. 1676–1680.

[5] U. Hess, "Nonverbal communication," Encyclopedia of Mental Health, 12 2016.

 ISBN: 978-1-61208-761-0

[6] E. Andre and C. Pelachaud, Interacting with Embodied Conversational Agents, 07 2010, pp. 123–149.

[7] K. Daher, M. Fuchs, E. Mugellini, D. Lalanne, and O. Abou Khaled, "Reduce stress through empathic machine to improve hci," in International Conference on Human Interaction and Emerging Technologies. Springer, 2020, pp. 232–237.

[8] L. Ricciardi, B. Demartini, A. Fotopoulou, and M. Edwards, "Alexithymia in neurological disease: A review," The Journal of neuropsychiatry and clinical neurosciences, vol. 27, 02 2015, p. appineuropsych14070169.

[9] J. R. Absher and J. Cloutier, Neuroimaging personality, social cognition, and character. Academic Press, 2016.

[10] K. R. Thórisson, "Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people," in First ACM International Conference on Autonomous Agents, 1997, pp. 536–537.

[11] J. Cassell, "More than just another pretty face: Embodied conversational interface agents," Communications of the ACM, vol. 43, no. 4, 2000, pp. 70–78.

[12] T. Bickmore and J. Cassell, "Social dialogue with embodied conversational agents," in Advances in natural multimodal dialogue systems. Springer, 2005, pp. 23–54.

[13] P. A. Angga, W. E. Fachri, A. Elevanita, R. D. Agushinta et al., "Design of chatbot with 3d avatar, voice interface, and facial expression," in 2015 International Conference on Science in Information Technology (ICSITech). IEEE, 2015, pp. 326–330.

[14] J. Cassell and H. Vilhjálmsson, "Fully embodied conversational avatars: Making communicative behaviors autonomous," Autonomous agents and multi-agent systems, vol. 2, no. 1, 1999, pp. 45–64.

[15] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth, "Empathy in virtual agents and robots: a survey," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 7, no. 3, 2017, p. 11.

[16] S. Brave, C. Nass, and K. Hutchinson, "Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent," International journal of human-computer studies, vol. 62, no. 2, 2005, pp. 161–178.

[17] B. L. Omdahl, Cognitive appraisal, emotion, and empathy. Psychology Press, 2014.

[18] L. Rameson and M. Lieberman, "Empathy: A social cognitive neuroscience approach," Social and Personality Psychology Compass, vol. 3, 01 2009, pp. 94 – 110.

[19] M. H. Davis, Empathy: A social psychological approach. Routledge, 2018.

[20] N. R. Seïler and P. Craig, "Empathetic technology," in Emotions, Technology, and Design. Elsevier, 2016, pp. 55–81.

[21] C. Marinetti, P. Moore, P. Lucas, and B. Parkinson, Emotions in Social Interactions: Unfolding Emotional Experience, 10 2011, pp. 31–46.

[22] R. J. Dolan, "Emotion, cognition, and behavior," science, vol. 298, no. 5596, 2002, pp. 1191–1194.

[23] P. Ekmann, "Universal facial expressions in emotion," Studia Psychologica, vol. 15, no. 2, 1973, p. 140.

[24] H. Meeren, C. Heijnsbergen, and B. Gelder, "Rapid perceptual integration of facial expression and emotional body language," Proceedings of the National Academy of Sciences of the United States of America, vol. 102, 12 2005, pp. 16 518–23.

[25] I. Poggi, C. Pelachaud, F. de Rosis, V. Carofiglio, and B. De Carolis, "Greta. a believable embodied conversational agent," in Multimodal intelligent information presentation. Springer, 2005, pp. 3–25.

[26] R.-J. Beun, E. De Vos, and C. Witteman, "Embodied conversational agents: effects on memory performance and anthropomorphisation," in International Workshop on Intelligent Virtual Agents. Springer, 2003, pp. 315–319.

[27] S. Provoost, H. M. Lau, J. Ruwaard, and H. Riper, "Embodied conversational agents in clinical psychology: a scoping review," Journal of medical Internet research, vol. 19, no. 5, 2017, p. e151.

[28] M. S. B. Mimoun, I. Poncin, and M. Garnier, "Case study—embodied virtual agents: An analysis on reasons for failure," Journal of Retailing and Consumer services, vol. 19, no. 6, 2012, pp. 605–612.

[29] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, "In the shades of the uncanny valley: An experimental study of human–chatbot interaction," Future Generation Computer Systems, vol. 92, 2019, pp. 539–548.

[30] N. Roa-Seïler, P. Craig, J. A. Arias, A. B. Saucedo, M. M. Díaz, and F. L. Rosano, "Defining a childs conceptualization of a virtual learning companion," in INTED2014 Proceedings. IATED, 2014, pp. 2992–2996.

[31] H. Prendinger and M. Ishizuka, "The empathic companion: A character-based interface that addresses users' affective states." Applied Artificial Intelligence, vol. 19, 03 2005, pp. 267–285.

[32] G.-D. Chen, J.-H. Lee, C.-Y. Wang, P.-Y. Chao, L.-Y. Li, and T.-Y. Lee, "An empathic avatar in a computer-aided learning program to encourage and persuade learners," Journal of Educational Technology & Society, vol. 15, no. 2, 2012, pp. 62–72.

[33] S. Marcos-Pablos, E. González-Pablos, C. Martín-Lorenzo, L. A. Flores, J. Gómez-García-Bermejo, and E. Zalama, "Virtual avatar for emotion recognition in patients with schizophrenia: A pilot study," Frontiers in human neuroscience, vol. 10, 2016, p. 421.

[34] C. Tsiourti, M. B. Moussa, J. Quintas, B. Loke, I. Jochem, J. A. Lopes, and D. Konstantas, "A virtual assistive companion for older adults: design implications for a real-world application," in Proceedings of SAI Intelligent Systems Conference. Springer, 2016, pp. 1014–1033.

[35] A. Sung, A. Bai, J. Bowen, B. Xu, L. Bartlett, J. Sanchez, M. Chin, L. Poirier, M. Blinkhorn, A. Campbell et al., "From the small screen to the big world: mobile apps for teaching real-world face recognition to children with autism," Advanced Health Care Technologies, vol. 1, 2015, pp. 37–45.

[36] M. Minge, M. Thüring, I. Wagner, and C. V. Kuhr, "The mecue questionnaire: a modular tool for measuring user experience," in Advances in Ergonomics Modeling, Usability & Special Populations. Springer, 2017, pp. 115–128.

 ISBN: 978-1-61208-761-0