Modeling an Empathetic Embodied Conversational Agent

Paulo Ricardo Knob[1], **Soraia Raupp Musse**[1]

[1]Virtual Humans Lab – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
90619-900 – Porto Alegre – RS – Brazil

paulo.knob@edu.pucrs.br, soraia.musse@pucrs.br

Abstract. *This work aims to propose an empathetic conversational agent endowed with a human-like memory and the ability to show some extent of empathy. Also, this model endow the virtual agent with a few other abilities, like recognizing the person it is talking to by its face and store/retrieve information based on a human memory model. Some experiments were conducted to gather both quantitative and qualitative information, which show that the proposed model works as intended.*

Resumo. *Este trabalho tem como objetivo propor um agente conversacional empático dotado de uma memória similar à humana e capaz de mostrar certa extensão de empatia. Além disso, este modelo dota o agente virtual com algumas outras habilidades, como reconhecer a pessoa com quem está conversando pela face e armazenar/recuperar informações com base em um modelo de memória humana. Alguns experimentos foram realizados para coletar informações quantitativas e qualitativas, as quais mostram que o modelo proposto funciona como pretendido.*

1. Introduction

Embodied Conversational Agents (ECAs) are virtual agents which are able to interact and talk with humans in a natural way. In the last years, many research was made to improve the quality of the communication abilities of such ECAs, both verbal and non-verbal [Yalçın 2020, Biancardi et al. 2019, Sajjadi et al. 2019]. A fair amount of effort is being directed on ECAs which can help people to have a healthier life [Kramer et al. 2019, Spitale et al. 2020, Das et al. 2019], for clinical interviews [Philip et al. 2020, Martínez-Miranda et al. 2019] and the training of some skill [Chetty and White 2019, Ayedoun et al. 2019].

Following this line of research, in this work it is aimed to propose an empathetic Embodied Conversational Agent (ECA) with general purpose endowed with many abilities. We developed both a 2D and a 3D model for our ECA, called Arthur (2D) and Bella (3D). We did so both to give more options for the user to choose and to investigate the difference in perception by the users. Besides a conversational module, using text and voice, this ECA is able to recognize the person he/she is talking to, as well to assess the user emotional state through his/her facial expressions. Also, Arthur/Bella is able to demonstrate different levels of emotion through his/her facial expressions, being also endowed with an Empathy Module. Lastly, it is equipped with a memory module, which tries to replicate the behavior of human memory and, thus, allows for Arthur/Bella to learn information with and from the user, while interacting; and to remember it later in the conversation or, even, in a different interaction. The empathy is built in the communication

with the user in mainly three parts of our model: firstly, through a pre-defined module of communication where Arthur/Bella asks questions about the user (demonstrating interest in the conversation); then, in the module of memory once the user feels that Arthur/Bella remember him/her; and finally with simple facial expressions that Arthur/Bella applies as a result of detecting facial expression module of Arthur/Bella.

2. Related Work

Zhang et al. [Zhang et al. 2019] propose to solve the problem of consistency on chatbot responses, concerning both context and personas (casual speaker). In their work, they present a self-supervised approach that uses the natural structure of conversational data to learn and leverage both topic and persona features. The results achieved indicate that the proposed model is able to capture meaningful topics and personas features. Also, the incorporation of the learned features helps to significantly improve the quality of generated responses on both data sets, even when comparing with models which explicit persona information.

A recent work conducted by Croes et al [Croes and Antheunis 2021] aimed to discover if a human being can build a relationship with a chatbot, as well which set of traits can help in such interactions. In order to conduct their research, they used the chatbot Mitsuku (https://www.pandorabots.com/mitsuku/). The set of traits measured were: social attraction, self-disclosure, intimacy, interaction quality, empathy, communication competence and feelings of friendship. The results achieved show that all these social processes diminish as time passes by, but intimacy. It suggests that the more people interacted with the chatbot, the worse the evaluation was. It seems to be reinforced by another discovery of the authors: after multiple interactions, people did not consider the chatbot as their friend.

The work of Yalcin [Yalçın 2020] aims to model empathetic behavior on Embodied Conversational Agents (ECAs). The ECA built by Yalcin has three stages: listening, where the agent captures input from the person it is talking to; thinking, where the agent process the information; and speaking, where the agent gives a proper response, both with words and gestural behavior. Since an empathetic behavior relies on the emotion of the subject, an emotion recognition module is used alongside the video input for the agent. Concerning this emotion, the audio of the person speaking is also used to help determine the overall emotion.

Sajjadi et al. [Sajjadi et al. 2019] conducted an experiment which aimed to investigate the effect of a person interacting with a personality-driven ECA. To test they hypothesis, the authors built a prototype of an ECA with a personality-driven model. An experiment was conducted with 41 participants in order to evaluate the initial hypothesis. The results achieved seem to validate them. As the authors comment, it was observed that an emotionally-personified ECA with an extrovert-based personality generates a higher sense of behavioral involvement in human users, when compared to a less emotionally-personified agent with no non-verbal behavior.

3. Proposed Model

The overview of our model is illustrated in Figure 1. As it can be seen in Figure 1, our model is divided into several modules. In blue we highlighted the two main Controllers,

which are responsible of controlling the interplay between many modules. The Behavior Control is responsible to define the appropriate behavior of the virtual agent, according all data available (i.e., person who is talking to it, agent memory, emotion detected, and so on). In other words, it allows the virtual agent to react to a given input provided by the user. Therefore, it is connected with all other modules and controllers. While the Chat module deals with what Arthur/Bella can tell to the user, the Conversation module is the output of such decision. The Facial Expressions module represent both the two different embodiment (i.e. Arthur and Bella) as well as the facial expressions modeled for both of them. The Emotion Detection module is responsible of identify the emotion expressed by the user, while the Face Recognition module can identify the user itself (e.g., Paul, Maria, etc.). The Beliefs module aims to endow Arthur/Bella with some level of reasoning regarding different pieces of information. Finally, the Voice Detection module allows the virtual agent to receive voice as input and transform it to text, so Arthur/Bella can understand what is being said.

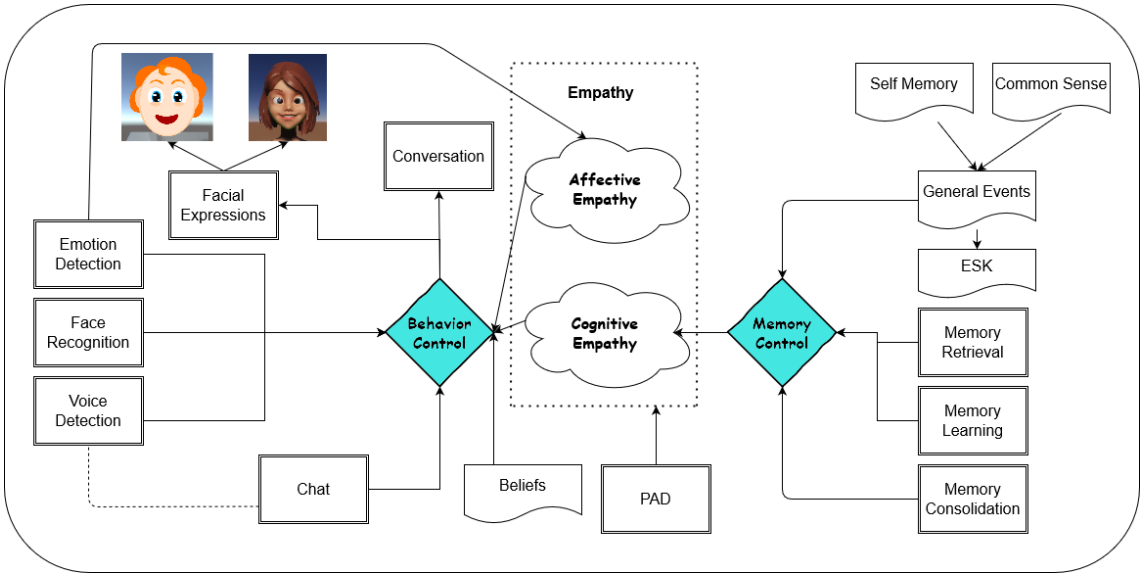

Figura 1. Overview of the proposed model. In blue, we highlight the two main Controllers. The Behavior Control is responsible to define the appropriate behavior of the virtual agent, while the Memory Control is responsible of store and retrieve memories.

The Memory Control is responsible for managing the memory of the virtual agent and is linked with all the memory features. The Self Memory and the Common Sense modules give our virtual agent some previous knowledge about many things before it can start to interact with people. While the Self Memory module gives the agent knowledge about himself/herself, the Common Sense module gives the agent knowledge about several things about the world and the environment. The General Events and ESK modules are related with the Autobiographical Memory model [Bluck and Levine 1998, Conway and Pleydell-Pearce 2000] and are used to build the memory of Arthur and Bella. The Memory Retrieval module is used to recover pieces of information from the memory of the virtual agent, while the Memory Learning module is responsible for storing new information into the memory of Arthur/Bella. Finally, the Memory Consolidation module is responsible for deciding which pieces of information

should stay inside the agent's memory and which ones should be deleted.

The Empathy module endows Arthur and Bella to demonstrate an empathetic behavior towards the person it is talking with. The emotional states used to express such empathy are defined inside a PAD (Pleasure-Arousal-Dominance) space [Mehrabian and Epstein 1972], represented by the PAD module.

4. Results

In this section, the results achieved by this work are presented. For the experiment discussed in this section, the personality of the agent is set as the following OCEAN values: O = 0.9; C = 0.5; E = 0.9; A = 0.7; N = 0.5. The initial PAD value is, thus, set as follows: P = 0.8; A = 0.5; D = 1. This experiment was conducted with only Short-term Interactions (STIs). Participants were recruited to interact with Arthur or Bella and answer an online questionnaire, summing up 30 people (22 Men and 8 Women). Of these 30 volunteers, 13 are Undergraduates, another 13 are Graduated, 3 completed High School and 1 person is a high school student. Concerning their past experience interacting with virtual agents, 6 participants answered as Very Low, 9 as Low, 9 as being Regular, 4 as being High and 2 as being Very High. The average age of the participants was 27.43, with a standard deviation of 11.84. Each participant was asked to accomplish a set of tasks to complete, as presented in Table 1.

Tabela 1. Tasks of the empathetic memory experiment.

Task	Description	Emotion
T1	Discover if the virtual agent likes video games and if it has a favorite game.	Happiness
T2	Discover if the virtual agent remembers about the participant's study and work.	Happiness
T3	Discover if the virtual agent has any pets, as well as more information about it.	Sadness
T4	Discover if the virtual agent remembers about any other subject that the participant already spoke with it.	Varied

In order to conduct the evaluation, we raise one main hypothesis: *H1:* We expect that participants can trigger ECA's memories and identify the associated emotion. Figure 2 presents the scores of the thirty participants from the experiment. Concerning **T1**, from 30, 29 participants were able to find out that the virtual agent likes video games. Also, 23 participants were able to identify the agent's favorite game, while 22 participants correctly identified the emotion conveyed by Arthur or Bella (i.e., Happiness). Concerning **T2**, from 30, 19 participants reported that the ECA was able to remember information about their study/work, and 15 of them correctly identify the emotion conveyed (i.e., Happiness). Regarding **T3**, 25 of 30 participants were able to answer that the virtual agent had a pet, and 24 were also able to identify the pet's name. Moreover, 20 participants could correctly identify the emotion conveyed by Arthur or Bella (i.e., Sadness). Concerning **T4**, 14 participants reported that Arthur or Bella was able to remember about some other subject that they chose to speak about and conveyed an appropriate emotion.

The results presented suggest that the participants were, in general, able to trigger the expected memories from Arthur or Bella and correctly identify the emotion associated with it, thus validating *H1*. It is also possible to notice that the worst results were found

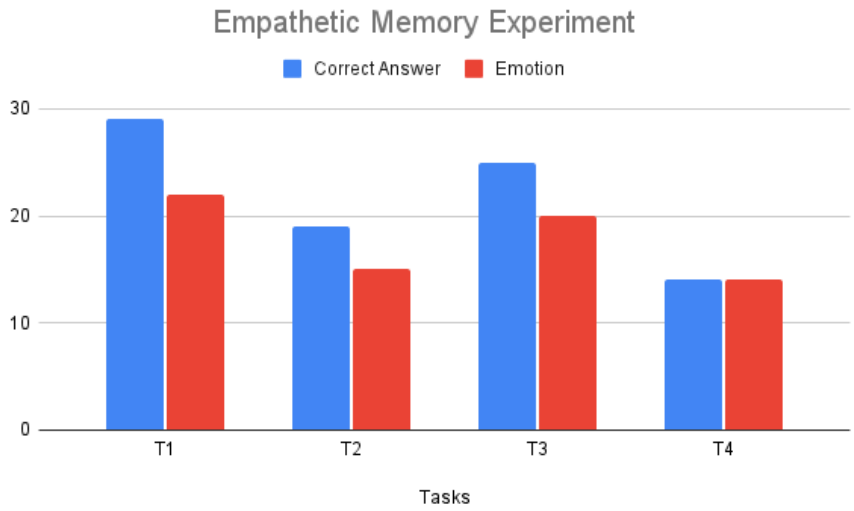

Figura 2. Scores of the thirty participants from the experiment. "Correct Answer", in blue, refers to the amount of people who answered as expected. "Emotion", in red, refers to the amount of people who correctly identified the agent's conveyed emotion.

when the 30 participants had to retrieve a memory about him/herself (19 participants answered correctly in **T2** and 14 in **T4**), when compared with memories about the agent itself, i.e., 29 participants correctly answer about video games in **T1**, and 25 concerning pets in **T3**. In this case, we hypothesize that **T1** and **T3** are more straight-forward tasks than **T2** and **T4**.

5. Final Remarks

This work presented a model of an empathetic Embodied Conversational Agent (ECA) endowed with many abilities, like face recognition, emotion detection, expressiveness, empathy and memory modeling. The main contribution of this work lies on the memory model, the empathy model and on the interplay between them. Some experiments were conducted in order to test the proposed model and collect both quantitative and qualitative information. The results achieved seem to confirm that Arthur/Bella presented the expected behavior.

This work has some limitations. Firstly, the number of users is certainly an issue that we want to work in a future. Having more participants is going to allow us to explore other hypotheses, such as the perception of people concerning Arthur and Bella. For future work, there are many avenues to follow. For instance, we want to invest more time in the visual behavior and facial animation of Arthur and Bella. Besides the modeling of different emotions, we would like to make this experience more personal, as it would be with a friend. In this topic, Melgare et al. [Melgare et al. 2019] suggested the existence of emotion styles, where each person would have their own way to demonstrate an emotion. In this sense, one interesting future work would be to endow Arthur/Bella with the ability to identify such style on the face of the user and mimic it. This way, we believe that the user would feel more comfortable with the facial expressions of Arthur/Bella.

Referências

Ayedoun, E., Hayashi, Y., and Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29(1):29–57.

Biancardi, B., Wang, C., Mancini, M., Cafaro, A., Chanel, G., and Pelachaud, C. (2019). A computational model for managing impressions of an embodied conversational agent in real-time. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7. IEEE.

Bluck, S. and Levine, L. J. (1998). Reminiscence as autobiographical memory: A catalyst for reminiscence theory development. *Ageing & Society*, 18(2):185–208.

Chetty, G. and White, M. (2019). Embodied conversational agents and interactive virtual humans for training simulators. In *Proc. The 15th International Conference on Auditory-Visual Speech Processing*, pages 73–77.

Conway, M. A. and Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, 107(2):261.

Croes, E. A. and Antheunis, M. L. (2021). Can we be friends with mitsuku? a longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1):279–300.

Das, K. S. J., Beinema, T., Op Den Akker, H., and Hermens, H. (2019). Generation of multi-party dialogues among embodied conversational agents to promote active living and healthy diet for subjects suffering from type 2 diabetes. In *5th International Conference on Information and Communication Technologies for Ageing Well and e-Health, ICT4AWE 2019*, pages 297–304. SCITEPRESS.

Kramer, L., ter Stal, S., Mulder, B., de Vet, E., and van Velsen, L. (2019). Developing embodied conversational agents for healthy lifestyles: a scoping review. In *ARPH*.

Martínez-Miranda, J., Martínez, A., Ramos, R., Aguilar, H., Jiménez, L., Arias, H., Rosales, G., and Valencia, E. (2019). Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *Journal of medical systems*, 43(8):246.

Mehrabian, A. and Epstein, N. (1972). A measure of emotional empathy. *Journal of personality*.

Melgare, J. K., Musse, S. R., Schneider, N. R., and Queiroz, R. B. (2019). Investigating emotion style in human faces and avatars. In *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 115–124. IEEE.

Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A., and Micoulaud-Franchi, J.-A. (2020). Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *npj Digital Medicine*, 3(1):1–7.

Sajjadi, P., Hoffmann, L., Cimiano, P., and Kopp, S. (2019). A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users. *Entertainment Computing*, 32:100313.

Spitale, M., Catania, F., Crovari, P., and Garzotto, F. (2020). Multicriteria decision analysis and conversational agents for children with autism. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Yalçın, Ö. N. (2020). Empathy framework for embodied conversational agents. *Cognitive Systems Research*, 59:123–132.

Zhang, Y., Gao, X., Lee, S., Brockett, C., Galley, M., Gao, J., and Dolan, B. (2019). Consistent dialogue generation with self-supervised feature learning. *arXiv preprint arXiv:1903.05759*.