Modeling an Empathetic Embodied Conversational Agent

Paulo R. Knob*
Escola Politecnica
PUCRS
Porto Alegre, Brazil
Email: paulo.knob@edu.pucrs.br

Soraia R. Musse
Escola Politecnica
PUCRS
Porto Alegre, Brazil
Email: soraia.musse@pucrs.br

***Abstract*—Empathy can be seen as a complex socio-emotional behavior, which is a result from the interaction between both cognitive and affective devices and is responsible, for instance, for one person being able to identify and mimicry others emotion. Moreover, the human memory is a powerful tool which allows for each person to be able to store and retrieve information about almost everything that happens. Equipping an embodied conversational agent (ECAs) with the ability of empathy, as well other features like memory, can help to make the interaction with humans smoother and more natural. 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. Finally, a few avenues for future work are also presented, elucidating what it is planned to do aiming to improve the quality of this work.**

I. Introduction

Human beings are the only known species that use spoken language to communicate, having a skill developed in communication that uses factors other than speech, such as, for example, body expressions and gazing [1]. A concept that is studied as relevant during communication is empathy, which is the sharing of emotions between individuals, as well as the behavior of adopting another person's point of view [2]. For example, if someone is talking with a person who just lost a beloved relative, he/she can perceive this person is truly sad and also feel sadness as well. Therefore, facial expressions are linked to the content of speech, emotion and personality, as well other behavioral variables, being even able to replace sequences of words, accompany them and be used to help disambiguate what is being said when the acoustic signal is degraded [3]. In fact, facial expressions may be used to communicate and influence other's behavior [4], [5]. Thus, it seems important to detect and understand facial expressions in every interaction between people.

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 [6]–[8]. A fair amount of effort is being directed on ECAs which can help people to have a healthier life [9]–[11], for clinical interviews [12], [13] and the training of some skill [14], [15].

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.

This work contributes in the development of an ECA endowed with an empathy behavior and a human-like memory, among other features. Such features should help this virtual agent to present a more natural interaction behavior, as observed in interactions between human beings. Thus, the main goal consists on the proposal of a model of a multi-purpose empathetic Embodied Conversational Agent (ECA) endowed with several abilities, which are used to improve the communication skills of the virtual agent and provide a smoother and more natural interaction with people.

* Ph.D. thesis

II. Related Work

Zhang et al. [16] 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 [17] 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.

One of the most accepted models concerning human memory cited on literature is known as Autobiographical Memory. As defined by Bluck et al. [18], autobiographical memory is "a system that encodes, stores and guides retrieval of all episodic information related to our person experiences". Also, according to Conway et al. [19], autobiographical memory can be grouped in three levels: lifetime periods, general events and event-specific knowledge. So, such memories can be directly accessed if the cues are specific and relevant to the person. Otherwise, if the cues are too general, a generative retrieval process must be used to produce more specific cues for the retrieval of relevant memories. The authors say that the difference between them is that "the search process is modulated by control processes in generative retrieval but not, or not so extensively, in direct retrieve" [19].

The work of Yalcin [6] 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. [8] 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.

III. 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.

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 [18], [19] 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 [20], represented by the PAD module.

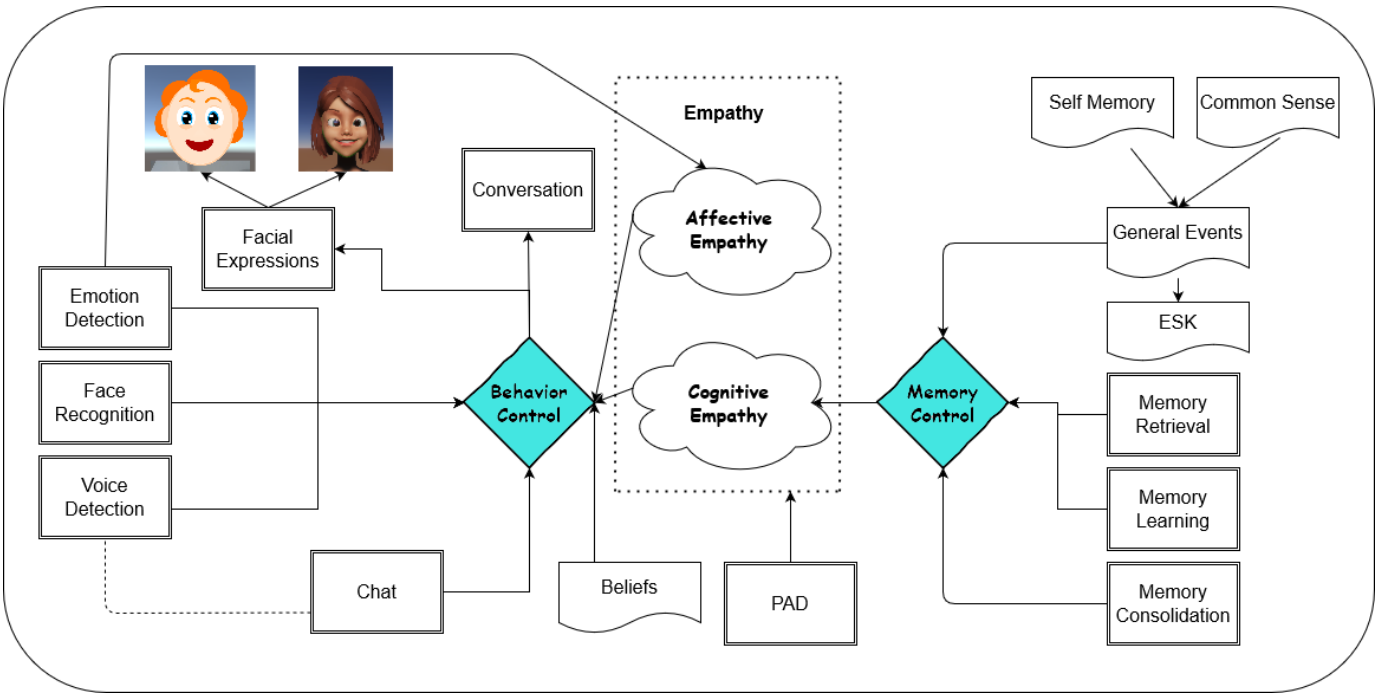

Fig. 1. Overview of the proposed model. In blue, we highlight the two main Controllers. The Behavior Control is responsible to define the appropriate behavior of the virtual agent, while the Memory Control is responsible of store and retrieve memories.

IV. Results

In this section, the results achieved by this work are presented. In order to test our method, we developed some experiments exploring the various features of Arthur and Bella. Mainly, we focus on the main contribution of our work: memory, empathy and the interplay between them. 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 I. Before starting these tasks, all participants read and agreed with the ethics term presented at the beginning of the questionnaire. After that, they were encouraged to download the ECA's executable file and freely interact with it for a short while to get used to it. They were also presented with a brief explanation about emotion and empathy and answered the Toronto empathy questionnaire [21] (TEQ) to measure their empathy level. The mean score computed for men was 43.63, while the mean score computed by for women was 47.12.

TABLE I
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

All tasks presented in Table I are related to some data that is present in the agent's memory, with a respective emotion associated. In T1, T2 and T3 participants should find some information saved in ECA's memory and recognize the expressed ECA's facial emotion. For T4, participants were asked to freely ask about the subject they want. Tasks T1 and T3 are about the ECA's self-memory, while T2 and T4 are related to what the agent knows about the participant. Following one of the definitions of empathy cited by de Wall [22] (the ability to understand and react towards the emotion of others), we believe that such emotional memories can be seen as an empathetic behavior, being able even to trigger such behavior in the participants. Finally, all tasks ask the participants to evaluate the agent's empathy on a Likert scale from 1 (no Empathy) to 5 (Extremely Empathetic).

In order to conduct the evaluation, we raise one main

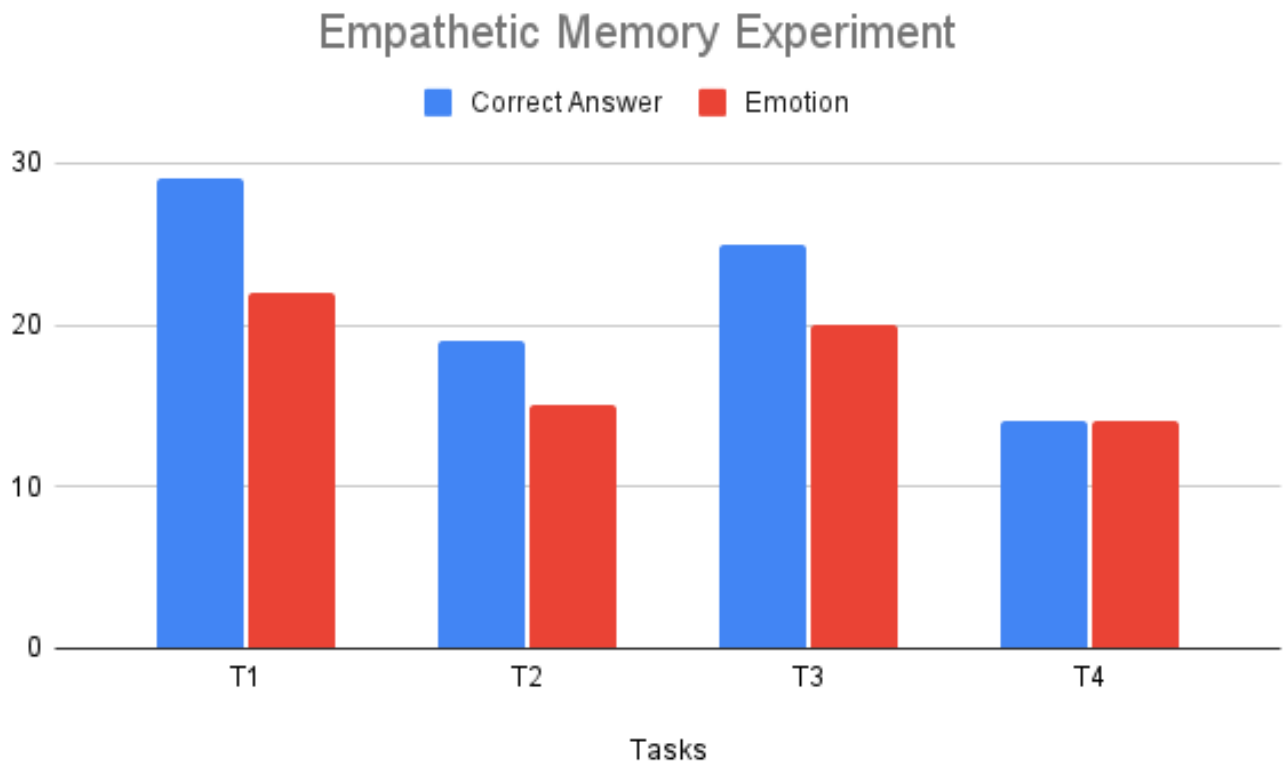

Fig. 2. Scores of the thirty participants from the experiment. "Correct Answer", in blue, refers to the amount of people who answered as expected. "Emotion", in red, refers to the amount of people who correctly identified the agent's conveyed emotion.

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. "Correct Answer", in blue, refers to the amount of people who answered as expected. For instance, in **T1** (Table I), it was expected that the participants were able to discover that the virtual agent enjoys playing video games. "Emotion", in red, refers to the amount of people who correctly identified the agent's conveyed emotion. 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 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**.

V. 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. [23] 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.

Another interesting avenue to explore is the dialog system. Besides building and integrating a generic chatbot model, we would like to take a deeper look into small talks. Our Embodied Conversational Agent is endowed with a small talk module, allowing it to start a conversation with the user about a pre-defined topic. The main problem is that such topics and dialogues need to be manually defined: each dialog tree needs to be manually written and added to the virtual agent. One possible future work would be to find a way to automatize this process. For instance, it would be interesting to investigate if is possible to use previous interactions as material to create new dialog trees. If so, it would be possible to build a script which can automatically gather previous interactions between Arthur/Bella and the users and build new topics and dialogues for the small talk module.

VI. Published Research

Generating background NPCs motion and grouping behavior based on real video sequences.
Paulo Knob, *Marlon Alcântara, Estêvão Testa, Rodolfo Favaretto, Gabriel Lima, Leandro Dihl and Soraia Raupp Musse.*
Journal Entertainment Computing (Elsevier). Volume 27, pages 179–187. 2018a.
DOI: https://doi.org/10.1016/j.entcom.2018.06.003.

Visualization of interactions in crowd simulation and video sequences.
Paulo Knob, *Victor Flavio de Andrade Araujo, Rodolfo Migon Favaretto, Soraia Raupp Musse.*
2018 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames).
DOI: https://doi.org/10.1109/SBGAMES.2018.00037.

Simulating crowds with ocean personality traits.
Paulo Knob, *Marcio Balotin, Soraia Raupp Musse.*
IVA '18: Proceedings of the 18th International Conference on Intelligent Virtual Agents, November 2018, Pages 233–238.
DOI: https://doi.org/10.1145/3267851.3267871.

Detecting personality and emotion traits in crowds from video sequences.
*Rodolfo Migon Favaretto, **Paulo Knob**, Soraia Raupp Musse, Felipe Vilanova, Ângelo Brandelli Costa.*
Machine Vision and Applications, volume 30, pages 999–1012 (2019).
DOI: https://doi.org/10.1007/s00138-018-0979-y.

Urban walkability design using virtual population simulation.
*CD Tharindu Mathew, **Paulo R Knob**, Soraia Raupp Musse, Daniel G Aliaga.*
Computer Graphics Forum, volume 38, pages 455-469 (2019).
DOI: https://doi.org/10.1111/cgf.13585.

Bioclouds: A multi-level model to simulate and visualize large crowds.
*Andre Da Silva Antonitsch, Diogo Hartmann Muller Schaffer, Gabriel Wetzel Rockenbach, **Paulo Knob**, Soraia Raupp Musse.*
Computer Graphics International Conference, CGI 2019: Advances in Computer Graphics pp 15-27.
DOI: https://doi.org/10.1007/978-3-030-22514-8_2.

How much do you perceive this? an analysis on perceptions of geometric features, personalities and emotions in virtual humans.
*Victor Araujo, Rodolfo Migon Favaretto, **Paulo Knob**, Soraia Raupp Musse, Felipe Vilanova, Angelo Brandelli Costa.*
IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, July 2019, Pages 179–181.
DOI: https://doi.org/10.1145/3308532.3329454.

Optimal Group Distribution based on Thermal and Psycho-Social Aspects.
***Paulo Knob**, Gabriel Wetzel Rockenbach, Claudio Rosito Jung, Soraia Raupp Musse.*
CASA '19: Proceedings of the 32nd International Conference on Computer Animation and Social Agents, July 2019, Pages 59–64.
DOI: https://doi.org/10.1145/3328756.3328765.

Moving virtual agents forward in space and time.
*Gabriel F Silva, **Paulo Knob**, Douglas A Schlatter, Carlos G Johansson, Soraia R Musse.*
2020 19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames).
DOI: 10.1109/SBGames51465.2020.00026.

Arthur: a new ECA that uses Memory to improve Communication.
***Paulo Knob**, Willian S Dias, Natanael Kuniechick, Joao Moraes, Soraia Raupp Musse.*
2021 IEEE 15th International Conference on Semantic Computing (ICSC).
DOI: 10.1109/ICSC50631.2021.00036.

Is my agent good enough? Evaluating Embodied Conversational Agents with Long and Short-term

interactions.
*Juliane Santos, **Paulo Ricardo Knob**, Victor Putrich Scherer, Soraia Raupp Musse.*
Proceedings of Simpósio Brasileiro de Jogos e Entretenimento Digital, 2021, Brasil.
DOI: 10.5753/sbgames_estendido.2021.19661

Arthur and Bella: multi-purpose empathetic AI assistants for daily conversations.
***Paulo Ricardo Knob**, Natalia Pizzol, Soraia Raupp Musse, Catherine Pelachaud.*
The Virtual Computer Journal, 2023.
DOI: 10.1007/s00371-023-02994-9.

WebCrowds: An Authoring Tool for Crowd Simulation.
*Gabriel Silva, **Paulo Ricardo Knob**, Rubens Montanha, Soraia Raupp Musse.*
2022 Brazilian Symposium on Computer Games and Digital Entertainment (SBGames).
DOI: 10.1109/SBGAMES56371.2022.9961084.

REFERENCES

[1] J. Cassell, J. Sullivan, E. Churchill, and S. Prevost, *Embodied conversational agents*. MIT press, 2000.

[2] F. B. de Waal and S. D. Preston, "Mammalian empathy: behavioural manifestations and neural basis," *Nature Reviews Neuroscience*, vol. 18, no. 8, pp. 498–509, 2017.

[3] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, 1994, pp. 413–420.

[4] M. Argyle and M. Cook, "Gaze and mutual gaze." 1976.

[5] G. Collier and G. J. Collier, *Emotional expression*. Psychology Press, 2014.

[6] Ö. N. Yalçın, "Empathy framework for embodied conversational agents," *Cognitive Systems Research*, vol. 59, pp. 123–132, 2020.

[7] B. Biancardi, C. Wang, M. Mancini, A. Cafaro, G. Chanel, and C. Pelachaud, "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)*. IEEE, 2019, pp. 1–7.

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

[9] L. Kramer, S. ter Stal, B. Mulder, E. de Vet, and L. van Velsen, "Developing embodied conversational agents for healthy lifestyles: a scoping review," in *ARPH*, 2019.

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

[11] K. S. J. Das, T. Beinema, H. Op Den Akker, and H. Hermens, "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*. SCITEPRESS, 2019, pp. 297–304.

[12] P. Philip, L. Dupuy, M. Auriacombe, F. Serre, E. de Sevin, A. Sauteraud, and J.-A. Micoulaud-Franchi, "Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020.

[13] J. Martínez-Miranda, A. Martínez, R. Ramos, H. Aguilar, L. Jiménez, H. Arias, G. Rosales, and E. Valencia, "Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour," *Journal of medical systems*, vol. 43, no. 8, p. 246, 2019.

[14] G. Chetty and M. White, "Embodied conversational agents and interactive virtual humans for training simulators," in *Proc. The 15th International Conference on Auditory-Visual Speech Processing*, 2019, pp. 73–77.

[15] E. Ayedoun, Y. Hayashi, and K. Seta, "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*, vol. 29, no. 1, pp. 29–57, 2019.

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

[17] E. A. Croes and M. L. Antheunis, "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*, vol. 38, no. 1, pp. 279–300, 2021.

[18] S. Bluck and L. J. Levine, "Reminiscence as autobiographical memory: A catalyst for reminiscence theory development," *Ageing & Society*, vol. 18, no. 2, pp. 185–208, 1998.

[19] M. A. Conway and C. W. Pleydell-Pearce, "The construction of autobiographical memories in the self-memory system." *Psychological review*, vol. 107, no. 2, p. 261, 2000.

[20] A. Mehrabian and N. Epstein, "A measure of emotional empathy." *Journal of personality*, 1972.

[21] R. N. Spreng*, M. C. McKinnon*, R. A. Mar, and B. Levine, "The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures," *Journal of personality assessment*, vol. 91, no. 1, pp. 62–71, 2009.

[22] F. B. De Waal, "Putting the altruism back into altruism: the evolution of empathy," *Annu. Rev. Psychol.*, vol. 59, pp. 279–300, 2008.

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