Multimodal Empathic Feedback Through a Virtual Character

Eleonora Chitti[1[0000−0003−4369−4615]], Manuel Pezzera[1[0000−0002−7649−7987]], and N. Alberto Borghese[1[0000−0002−0925−3448]]

Department of Computer Science, University of Milan, Milan, Italy
{eleonora.chitti,manuel.pezzera,alberto.borghese}@unimi.it

Abstract. The development and application of empathic virtual agents is rising fast in many fields, from rehabilitation to education, from mental health to personal wellbeing. The empathic agents should be developed to appropriately react to the user's affective state, with the aim of establishing an emotional connection with him/her. We propose a position paper to shape the design of an Empathic Virtual Character to be included in an existing platform with exer-games supporting postural rehabilitation. The character will express emotions to the user with facial animations and speech statements. The character's emotion to express will be based on the current user's affective state, inferred by the input data. Finally, we propose a possible improvement of the developed interaction framework.

Keywords: Empathic agent · Conversational agent · Affective computing.

1 Introduction

In recent years, research on affective computing has risen fast. In particular, the application of empathic virtual characters and robotic agents has increased in different fields, from physical and cognitive rehabilitation support [2] to remote mental health intervention [10], from educational purposes [17] to students' wellbeing [11].

Social agents change the way they act in reaction to the user's perceived affective state [2]. This change can be shown though different channels and modalities. For instance it can be shown changing the facial expression or the vocal speech content or pitch profile, or, in case of robots, changing the action to perform in the real world.

New technologies, as RGB cameras and smartwatches that provide data as the heart rate and gyroscope, support the development of algorithms to infer of the user's affective state across time. For instance, smartwatch's gyroscopes and heart rate data are used in [13] along with the images of the user's face collected through an RGB camera and the user's voice audio, to infer the user's affective state. More recently, skin conductance devices and user's face image analysis have been exploited with the same goal [2].

Usually the affective state is described either according to FACS (Facial Action Coding Systems) classification [7] or as a point in the Russell's circumplex model [14] shown in Fig. 1 [10]. More recently, the PAD model, an extension of the Russell's circumplex model with the Pleasure on the x-axis Arousal on the y-axis and the Dominance on the z-axis, has been proposed and used [13].

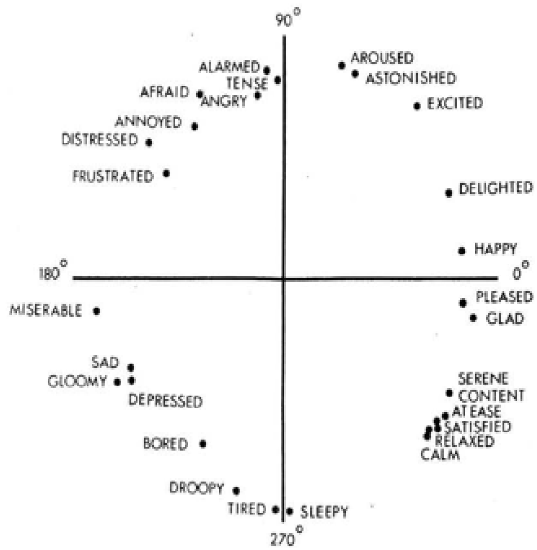

Fig. 1. Russell's model, circular scaling for 28 affect word. Image from [14]

Due to the subtle changes that different emotions provoke, the current reports on emotions identification usually consider a reduced set of emotional states: [13] takes into account the subset of the possible affective states that correspond to the 6 basic emotions exposed by Ekman [7], while [10] takes into account an even smaller subset that the basic emotions.

The current affective state identified is then used to select the most adequate agent's reaction. For instance, in [10] a set of agent's reactions for each possible affective state has been defined, or in [2] the information on affective state is used to match set of rules to pick-up the appropriate output.

2 Methods

In this work we propose the design of a Multimodal Empathic Virtual Character to be included into a platform that provides autonomous rehabilitation at home [5,12]. In this platform rehabilitation exercises are guided through fully adaptable exer-games and rehabilitation sessions are supervised by a virtual therapist that supervises the correct execution of the exercises and the safety of the patient.

The virtual therapist is embodied into a 3D virtual character, the Virtual Therapist (VT) Hannah (Fig. 2). This plays the role of guiding the patient through the exer-games, to explain the play rules, advice him/her and provide feed-back on any hazard or incorrect motion.

Besides these functionalities, a critical functionality for at home rehabilitation is motivation. Rehabilitation may easily require a long period of training, and the virtual therapist has to engage the patient motivating him/her to keep-on with exercising. In this framework, the VT receives the motion data of the patient, and the data of interaction with the game element, it analyzes them in real-time and provides a feed-back to the patient at the right time.

Fig. 2. The virtual therapist Hannah provides feed-back to the patient at the end of an exer-game session

However, the VT can go one step further and provide also a supporting feed-back to increase motivation of the patient. To this aim, data that can be used to infer the user's current affective state have to be provided (e.g. through "Affectiva" software libraries). Typical input sources here are facial expressions and natural speech processing [15,1]. These are indeed the main channels also for emotional human-human communication. In the present work, we use the devices already used to track patient's motion and that are used to animate the avatar inside the game [12]. In particular the video images produced by the Kinect RGB-D camera used to track the skeleton and the audio acquired through the same camera.

Such inputs are sent to an Emotional Intelligence module that processes them to produce the most adequate output. This module is aimed to process the video images to infer the current affective state of the patient. A set of rules is exploited to this aim, evaluating the Arousal and Valence, and then identifying the affective state as one of the subset of 4 of the basic emotions by Ekman [7,8]: enjoyment, sadness, anger, surprise. To provide variability and consistency in the time course of emotional response, the approach based on stochastic finite state machines proposed in [6] will be explored and further developed. Since the existing platform [5] is mainly used for rehabilitation, the data about the performance of the player in the exer-games are collected and they are combined, through a set of rules defined with the therapists, with the current emotional state identified to select the most appropriate output.

The output is the portray of a given emotion by the VT. To this aim we resort to facial animation and the speech production, in which the emotional content is represented through specific pitch time course [2,13] In particular, the Emotional intelligence module will provide a set of reactions for each one of the possible user's affective states (Fig. 3).

The virtual character data processing procedure is shown in Fig. 3.

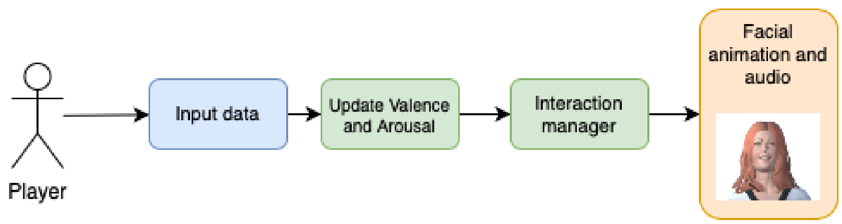

Fig. 3. Data processing procedure

The VT reaction will be represented as facial animation and vocal response. The avatar's facial animations is based a simplified version of the Facial Action Coding System (FACS) [9] and it shows the 4 basic emotions [7]: enjoyment, sadness, anger, surprise. For example, the virtual character shows enjoyment with lips pulled back and teeth exposed in the smile, it shows anger with eyebrows pulled down together and lips pressed tight. To provide a more realistic interaction, the amount of emotion produced can be regulated. To this aim, for each expression produced an intensity value is also produced that indicates the intensity of the agent's facial expression while communicating a specific emotion. This can be implemented through a fuzzy system similar to that used to estimate the degree of hazard in patient's movement [12].

The feed-back produced through voice by the VT as well as the content of feed-back sentences will also portray its emotional state. To this aim statements are divided into sets with homogenous meaning: e.g. "Good job", "Very good", "Good performance".... According to the emotional state the adequate voice pitch is selected along with the most adequate feed-back class. The output statement is synthetized through a text to speech module.

To this aim, the statements will be synthetized through Google [1] text to speech module and then the pitch of the agent's voice will also be changed based on the emotion to express. Some specific words in the speech need to be emphasized to better express a specific emotion, therefore the text, to be elaborated by the text to speech module, includes the instructions, supported by Google APIs[2] , to give emphasis on specific words.

3 Discussion and future works

The purpose of this work is to illustrate the design of the interaction framework of an empathic virtual character. This character has been developed to support an exer-game-based platform developed to support autonomous at-home rehabilitation [5,4]. From the operational point of view, the result is a 3D agent that reacts empathically, through speech and animation, to the user's affective state inferred from biometric data. The agent will such capable to guide the player through the exer-games, explain the playing rules and advise him/her in an empathic way. This is expected to increase the overall motivation of the patient, to continue training with the platform, increasing fidalization and motivation.

A first version of the emotional engine has been implemented and evaluated. The framework is currently under refinement along several directions. The aspect of the avatar and the virtual handles on the model that allow precise motivation can be improved to provide also the most subtle changes of expression that can make emotions expression more realistic. The library currently evaluated for identifying emotions is "Affectiva" [3], that shows a high level of sensitivity.

Such an approach has been developed mainly for elder population. However, due to Covid pandemic, several platforms to support training and learning in young population have been developed. Such platforms can largely benefit from such approach. This will be for instance explored by the authors in the H2020 project ESSENCE that will start in November 2020. For this particular population we are designing a 2D comic-like character, since it is preferable to design comic-like characters in case of agents specifically developed for children, as [16] that developed a virtual agent with a comic superhero aspect to engage children to discover tourist information explained by the agent, or a colored personal robot with soft fur, as [18].

To give a cartoonish look to the agent, when expressing anger the tone of red of the agent's skin color rises, in case of disgust the tone of green, in case of fear the purple. Some of the agent's expressions are shown in Fig. 4. In addition, to make the user more comfortable and engaged, we will include a section to personalize the character, changing the sex, the skin color, the haircut and color, the dress and the accessories.

All this is related to the interaction with a single character. Especially for children for which the variety is fundamental, we are exploring the possibility of

[1] https://cloud.google.com/text-to-speech/
[2] https://cloud.google.com/text-to-speech/docs/ssml
[3] https://www.affectiva.com/

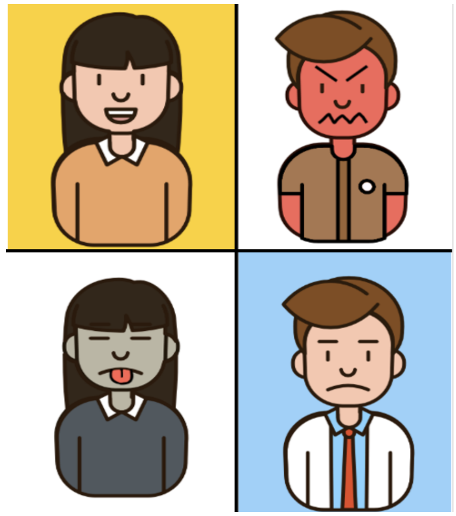

Fig. 4. Examples of agent's expressions, from top left to bottom right: enjoyment, anger, disgust, sadness

having several character and letting the emotional intelligence to choose which character can be the most effective to convey the emotional content feed-back, dynamically. In this way more variety is added to the system along with a surprise element that is a key ingredient to make graphical interaction long lasting.

4 Conclusion

We have exposed the design and development of a 3D empathic virtual character and the integration on the already available exer-gaming platform for postural home rehabilitation. The user's affective state is inferred exploiting Kinect and Wii balance board data, that are the controllers of the exer-gaming platform. The agent's reaction is then appropriately selected based on the user's affective state. The agent's reaction includes facial animation and speech statements. Finally, we provided an overview on possible future improvements of the existing agent.

References

1. Alva, M.Y., Nachamai, M., Paulose, J.: A comprehensive survey on features and methods for speech emotion detection. In: 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp 1–6. (2015) https://doi.org/10.1109/ICECCT.2015.7226047
2. Alves-Oliveira, P., Janarthanam, S., Candeias, A., Deshmukh, A., Ribeiro, T., Hastie, H., Paiva, A., Aylett, R.: Towards dialogue dimensions for a robotic tutor in collaborative learning scenarios. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp 862–867. (2014) https://doi.org/10.1109/ROMAN.2014.6926361

3. Berner, U., Rieger, T.: A Scalable Avatar for Conversational User Interfaces. In: Carbonell N, Stephanidis C (eds) Universal Access Theoretical Perspectives, Practice, and Experience, pp 350–359, Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
4. Boccignone, G., de'Sperati, C., Granato, M., Grossi, G., Lanzarotti, R., Noceti, N., Odone, F.: Stairway to Elders: Bridging Space, Time and Emotions in Their Social Environment for Wellbeing, Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods - Volume 1, ICPRAM, pp 548-554. (2020) https://doi.org/10.5220/0009106605480554
5. Borghese, N.A., Pezzera, M., Mainetti, R., Essenziale, J., Cazzaniga, R., Reggiori, B., Mercurio, S., Confalonieri, P.: A cloud-based platform for effective supervision of autonomous home rehabilitation through exer-games, In: 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH), (2018) https://doi.org/10.1109/SeGAH.2018.8401383
6. Cattinelli, I., Goldwurm, M., Borghese, N.A.: Interacting with an artificial partner: modeling the role of emotional aspects, Biological Cybernetics **99**, 473-–489 (2008) https://doi.org/10.1007/s00422-008-0254-9
7. Ekman, P.: Basic Emotions. In: Handbook of Cognition and Emotion. John Wiley & Sons, Ltd, pp 45–60. (1999)
8. Ekman, P., Friesen, W.V.: Measuring facial movement, Environmental psychology and nonverbal behavior **1**, 56—75 (1976) https://doi.org/10.1007/BF01115465
9. Ekman, P., Friesen, W.V.: Facial action coding system, (1978) https://www.paulekman.com/facial-action-coding-system/
10. Ghandeharioun, A., McDuff, D., Czerwinski, M., Rowan, K.: EMMA: An Emotion-Aware Wellbeing Chatbot. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp 1–7. (2019) https://doi.org/10.1109/ACII.2019.8925455
11. Jeong, S., Alghowinem, S., Aymerich-Franch, L., Arias, K., Lapedriza, A., Picard, R., Park, H.W., Breazeal, C.: A Robotic Positive Psychology Coach to Improve College Students' Wellbeing, In: The 29rd IEEE International Conference on Robot and Human Interactive Communication, (2020) [ONLINE] Available: http://arxiv.org/abs/2009.03829
12. Pirovano, M., Mainetti, R., Baud-Bovy, G., Lanzi, P.L., Borghese, N.A.: IGER - Intelligent Game Engine for Rehabilitation, IEEE Trans. on CIAIG. Vol.8(1), March 2016, pp. 43-55. (2016)
13. Rudovic, O., Lee, J., Dai, M., Schuller, B., Picard, R.W.: Personalized machine learning for robot perception of affect and engagement in autism therapy, Sci Robot 3:eaao6760. (2018) https://doi.org/10.1126/scirobotics.aao6760
14. Russell, J.A.: A circumplex model of affect. J Pers Soc Psychol 39:1161–1178. (1980) https://doi.org/10.1037/h0077714
15. Sariyanidi, E., Gunes, H., Cavallaro, A.: Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence **37**, 1113—1133 (2015) https://doi.org/10.1109/TPAMI.2014.2366127
16. Sorrentino, F., Spano, L.D., Scateni, R.: SuperAvatar: Children and mobile tourist guides become friends using superpowered avatars, Proceedings of 2015 International Conference on Interactive Mobile Communication Technologies and Learning, IMCL 2015, pp 222–226. (2015) https://doi.org/10.1109/IMCTL.2015.7359591

17. Spaulding, S., Chen, H., Ali, S., Kulinski, M., Breazeal, C.: A Social Robot System for Modeling Children's Word Pronunciation: Socially Interactive Agents Track. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, pp 1658–1666. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2018)
18. Westlund, J.M.K., Park, H.W., Williams, R., Breazeal, C.: Measuring Young Children's Long-Term Relationships with Social Robots, In: Proceedings of the 17th ACM Conference on Interaction Design and Children. Association for Computing Machinery, pp 207–218,New York, NY, USA (2018)