A Mobile Virtual Character with Emotion-Aware Strategies for Human-Robot Interaction

Caetano M. Ranieri, Roseli A. Francelin Romero
Institute of Mathematical and Computer Sciences
University of São Paulo, USP
São Carlos, Brazil
e-mail: cmranieri@usp.br, rafrance@icmc.usp.br

Humberto Ferasoli Filho
Faculty of Sciences
São Paulo State University, UNESP
Bauru, Brazil
e-mail: ferasoli@fc.unesp.br

***Abstract*— Emotions may play an important role in human-robot interaction, especially with social robots. Although the emotion recognition problem has been massively studied, few research is aimed at investigating interaction strategies produced as response to inferred emotional states. The work described in this paper consists on conceiving and evaluating a dynamic in which, according to the user emotional state inferred through facial expressions analysis, two distinct interaction strategies are associated to a virtual character. An Android app, whose development is in progress, aggregates the user interface and interactive features. We have performed user experiments to evaluate whether the proposed dynamic is effective in producing more natural and empathic interaction.**

Keywords - emotions; mobile devices; human-robot interaction; social robots; virtual assistant.

I. INTRODUCTION

People have traditionally seen the interaction between humans and computer systems as an essentially non-emotional one, in which user emotional reactions do not affect the system behavior [1]. Researchers point, however, that emotions may play an important role on these environments. On the one hand, since emotions have significant influence on human cognition, artificial emotions may provide computer programs with better problem solving or decision-making [2]. On the other hand, Picard [3] suggests that abilities to recognize, understand and show emotions are requisites for these systems to interact naturally with humans. Therefore, the effort to consider emotions in different kinds of computer systems is justified.

Personal assistants for smartphones, endowed with a virtual body, show several features in common with social robots [4]. Although, according to some general definitions, not having a physical body may disqualify them of being an actual robot [5], some definitions of social robots cover virtual agents as well. Hegel *et al.* [6] define a social robot as an autonomous agent that behaves socially in a given context, have specific communicative abilities and shows an explicitly social appearance.

Research in the literature investigate how emotion-aware interaction strategies may influence the feeling of empathy towards robotic agents [7]. However, in these studies, the user's emotional state affects only behavior selection, not the behaviors themselves. In this paper, we investigate not only possible roles of emotion recognition on behavior and content selection during social human-agent interaction, but also whether expressing the same agent's behaviors through different verbal and visual cues, according to an inferred emotional context, may increase the feeling of empathy and improve the quality of the interaction.

It consists on development and evaluation of an Android application provided with a virtual female character, with personal assistant features. The system continuously analyses the user facial expression, obtained by the smartphone frontal camera, and infers its emotional state with an emotion classifier previously developed, as a product of another research. The result of this classification determines one between two possible interaction strategies. One of them, aimed at positive emotional states, is more extroverted, while another, aimed at negative emotional states, is more formal. We have also performed user experiments with the proposed approach.

This paper is organized as it follows. In Section II, some related work is presented. In Section III, we describe the emotion recognizer adopted in the project. The main application is presented, in details, in Section IV. In Section V, the experiments performed are described, jointly with a discussion of the results obtained. Finally, in Section VI, the conclusion and future works are presented.

II. RELATED WORK

Pütten *et al.* [8] investigated whether humans are able to feel empathy towards robots. They conducted an experiment to investigate emotional reactions of participants towards Ugobe's Pleo, a robotic dinosaur, shown in different situations. A group of volunteers watched a video of a friendly interaction between a person and the robot, and another group watched a video in which the robot was tortured. Physiological measurements and self-reported emotional reactions revealed that participants who watched the friendly video experienced soft and positive emotions, while participants who watched the torture video experienced more intense and negative emotions, reporting having felt empathy.

Jo *et al.* [9] investigated whether humans may perceive a robot as a real company, by conducting an experiment with students of the Sungkyunkwan University, South Korea. First, all participants, alone, have watched a presumably funny video, with laughs inserted in determined passages. After that, this time accompanied, they watched another video, with no laughter included. A group watched the video with a human companion, while the other group watched it with a Nao robot, provided with an artificial laugh in certain parts of the video.

 ISBN: 978-1-61208-462-6

Results revealed that the participants had more fun when the watched the video accompanied, no matter if this companion was a person or a robot. The authors concluded that this might mean that participants interacted empathically with the robot, sharing positive emotions with it.

The above-mentioned works have shown that humans are capable of feeling empathy toward robots. Pütten *et al.* [8] evaluated emotional reactions, whereas Jo *et al.* [9] investigated the pleasure caused by a robotic companion. However, these researches did not investigate how changes in the robot behaviors would affect that feeling of empathy, and how such features may improve the social human-robot interaction. To provide researches on this direction, more dynamic environments would be needed.

Some research deals with automatic interaction strategies to produce empathic interactions between human and robot. Leite *et al.* [7] have created an environment in which the iCat robot played chess with 8 to 12 year-old children. The system inferred the child emotional state through her facial expressions, caught by a webcam. The robot displayed some empathic strategies, such as making encouraging comments, offering help or deliberatively playing a bad move. A reinforcement-learning algorithm selected the interaction strategy. According to its results, the referred work successfully provided a set of adaptive behaviors and applied emotion recognition to train the learning algorithm responsible for the arbitration process. However, the described system used emotional response to determine which behaviors it would execute, but not how. In other words, the behaviors were the same, despite the user emotion.

The Smartphone Intuitive Likeness and Expression (SMILE) app, described by Russel *et al.* [10], is an interface to produce animated emotions, synthetize speech and learn vocabulary. The smartphone, in which the app runs, was placed on a UM-L8 robot, dedicated to children. The smartphone screen was displaying a pair of eyes, which blinked intermittently, used to modulate artificial emotions. Although the system was capable of synthetizing emotions, it had no component to analyze users' emotions. The authors conducted only exploratory studies. Although this paper relates few relevant experiments, the emotional interface is interesting. Besides, placing a mobile device as part of a social robot is a viable approach to endow a virtual character with a physical body.

III. THE EMOTION RECOGNIZER

Concerning emotion recognition, several systems have been proposed and evaluated. The available approaches may consider different cues of the emotion that a person is experiencing, such as neurological responses, autonomic activity, facial expressions or speech [1]. The available emotional models may be classified in two categories: discrete (i.e., a finite set of well-defined emotions) [11] or continuous (i.e., a dimensional space with continuous variables attributed to different emotional properties) [12].

The emotion recognizer adopted in our application is a product of a previous research performed in our lab, the Robots Learning Laboratory (LAR) of the Institute of Mathematical and Computer Sciences at University of São Paulo (ICMC/USP) [13]. It takes frontal images of human faces and classifies them as one between seven discrete emotions. Six of them are the basic emotions proposed by Ekman [11]: happiness, surprise, fear, anger, disgust and sadness. The other one is the neutral emotion.

FaceTracker library [14], applied in that project as a feature extractor, obtains interest points (regions of the nose, the mouth, the eyebrows and the chin, for example) on images of human faces. The recognizer computes the ratio between distances and angles of pairs of these points, as illustrated in Figure 1, and stores them in a feature vector. Then, it applies the generated vector as input of a classification technique.

The system provides six different feature vectors and three different machine-learning techniques for emotion classification: multilayer perceptron (MLP), support vector machines (SVM) and the C4.5 algorithm. A detailed discussion on how the system prepares each feature vector, with comparisons between all combinations of feature vectors and classification techniques, is found in Libralon [13]. The machine learning techniques were trained with two datasets: the Radboud Faces Database (RaFD) [15] and the Extended Cohn-Kanade (CK+) [16].

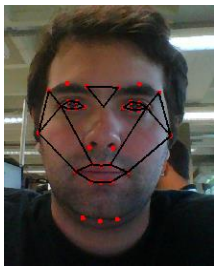

Figure 1 - Points, distances and angles considered by the emotion recognizer.

IV. PROPOSED SYSTEM

The proposed application consists on an embodied virtual agent for smartphones, whose interaction strategy adapts itself to the user's emotional state inferred by his facial expression analysis. The character always displays one between two interaction strategies: friendly or normal. Besides influencing content selection, the current interaction strategy determines how the agent must express each of its behaviors. For now, it is done by, depending on the interaction strategy, using a different sentence for the same verbal behavior or showing slightly different general visual cues. The behavior of the system is the following, according to each strategy:

- **Friendly:** the environment is more proactive than in normal interaction. Besides, it shows a more personal language, using first person nouns and verbs. For example, "I would be very happy to help, in case you need something. May I show you the available commands?"
- **Normal:** the environment is less proactive than in friendly strategy, and its language is more objective. For example, "Would you like to check the available commands?"

 ISBN: 978-1-61208-462-6

The system selects the current interaction strategy based on the last outputs of the emotion recognizer. If the inferred emotion is positive, the system changes the interaction strategy to friendly. If the emotion is negative, it changes the interaction strategy to normal. If the emotion is neutral, the interaction strategy is not changed.

The user may interact with the system through speech, applying the Android native speech recognition and synthesis Application Programming Interfaces (API). A text input was also included. The spoken language is Brazilian Portuguese, given we are performing user studies in Brazil.

As Figure 2 illustrates, the interaction strategy determines slight changes on the visual representation of the character. For now, the only difference is that, when performing the friendly strategy, the character shows a slight smile, which does not happen when it performs the normal strategy.

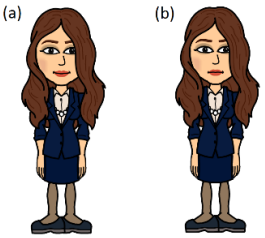

Figure 2. Virtual character, with strategy: (a) friendly; (b) normal.

The application development consists of three modules. One of them is the emotion recognition module, already described in Section III. The two others are the interaction motor and the content motor, presented further in this section. Connections between these three modules within themselves and with the human user are shown in Figure 3.

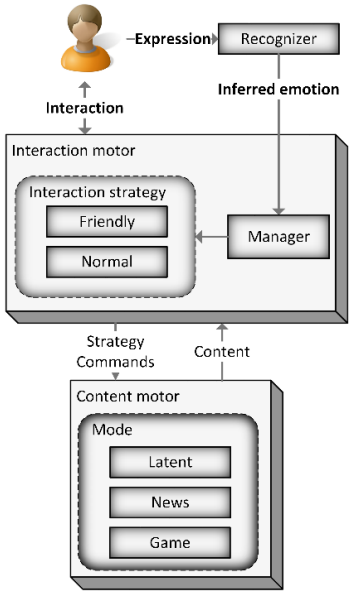

Figure 3. Proposed system architecture, comprised by an emotion classifier, an interaction motor and a content motor.

The emotion recognizer takes, as input, images of the user's face, acquired in real time from the smartphone frontal camera. Every time it finishes the classification process, the system sends the result to the interaction motor. This process repeats indefinitely, whenever the application is running.

The interaction motor aggregates the speech recognizer, the speech synthesizer, the text interface and the visual representation of the character. Therefore, besides selecting the interaction strategy based on the recognized emotion, it is responsible for all interaction features: receiving the user inputs, recognizing and sending commands to the content motor, manifesting interactive behaviors and providing content to the user. It expresses all behaviors according to the current interaction strategy, by providing verbal and visual cues that allows the user to perceive the current interaction strategy.

The content motor, which has access to the current interaction strategy, is responsible for selecting the behaviors and the specific content shown to the user. It also processes the user commands. The content motor may operate in three distinct *modes*, each related to a specific functionality, comprising a subsystem: latent, game and news. In the latent mode, the environment keeps with no action most of the time, and eventually, it manifests proactivity and suggests an activity. The other modes are described on the following subsections.

A. Game Mode

In this mode, the user plays a classical words-based game called Doublets, which does not require a dedicated graphical interface, endowed with some proactive behaviors. As the experiment described in Section V deals solely with the game, it is appropriate to give a more detailed description of it.

The goal of Doublets is, given two words, A and B, to start from word A and change it one letter at a time, always resulting in an existent word, until to obtain the word B. As input interface, the user can always choose between speech and a keyboard interface, which has been included to prevent the user of being stuck in case the speech interface is inaccurate in a given situation. In implemented version, the supported language is Brazilian Portuguese.

At any time, the virtual character may take initiative by exhibiting three verbal behaviors: making encouraging comments, giving information on how many words the user has reached or suggesting a viable word. Although the system randomly decides whether to take initiative or not, a reinforcement-learning algorithm arbitrates on which behavior it will show. The same behavior may cause different sentences to be spoken, according to the interaction strategy.

The arbitration process has been modeled as a multi-armed bandit problem, which consists in, given a set of possible strategies, called gambling machines, choosing the next machine so that the reward is maximized in the long run [17].

After having chosen each behavior (gambling machine) at least once, the implemented algorithm selects the behavior that maximizes (1), where $\bar{x}$ is the average reward obtained from behavior i, n_i is the number of times behavior i was selected and n is the number of behaviors selected so far.

$$\overline{x_i} + \sqrt{\frac{2 \ln n}{n_i}}\,. \qquad (1)$$

 ISBN: 978-1-61208-462-6

We applied a reinforcement rule based on emotional response. It considers the mean value of the first two emotions recognized immediately after the behavior exhibition (the greater the value, the more positive the emotion; zero means neutral) and subtracts the mean value of the last two emotions recognized before the behavior exhibition. The result of this computation is the behavior's reward.

B. News Mode

In this mode, the system downloads a collection of recent news and recommend some of them, based on their category. As we did with the game behaviors, we have modeled this recommender as a multi-armed bandit problem, in which each news category is a gambling machine.

The interaction happens as follows: after having chosen each news category at least once, the system chooses the news category that maximizes (1). Then, the character reads the selected news title and allows the user to access its description. If the user requests the description, the referred news category receives a reward. Otherwise, the character asks the user to rate his interest on that news in a 1 to 3 scale, and obtains a reward value based on it.

To get the news, we have chosen Folha de São Paulo, a traditional Brazilian newspaper that provides separate RSS files for each news category. In the implemented system, we have included the following news categories: politics, sports, tech and science.

V. Experiments and Results

As already mentioned, we have performed user experiments only with the game mode, which is more interactive and allows the character to take initiative in more situations. Thus, for the experiment, we have modified the application to operate only on the game mode.

A. Experimental Setup

The participants evaluated two versions of the application. For convenience, we are going to assign arbitrary names to each version: Claire and Rachel. Claire is the full app, as described above. Rachel is a modified version, which shows only the friendly interaction strategy and selects its proactive behavior randomly. The experiment aims to test whether users feel more empathy towards Claire than towards Rachel, and whether they perceive a more interesting behavior in the former than in the latter.

We performed the experiment on 14th and 15th January 2016, at the Integration of Systems and Intelligent Devices Laboratory (LISDI) of São Paulo State University (UNESP), Bauru campus. All 11 participants were 18 to 21 year-old Computer Science freshman students. We asked them to play with both versions and fill a subjective questioner after each session. Participants were informed that the system would analyze their facial expression, but they were not told the differences between the two versions of the app, neither which version they were experiencing at each time. The experiment was counterbalanced, it is, half of the participants played first with Claire, and the other half played first with Rachel.

Each answer of the questioner had to be a 1 to 5 rate. The participants answered whether they felt empathy towards the virtual character and evaluated some desirable characteristics, such as realism, kindness, pleasantness and competence. Given the within-subject approach of the experiment, the means of the differences of both user evaluations were analyzed and it was applied a paired t-test to check whether there was statistical significance.

B. Results and Discussion

The experiments have shown interesting results concerning the empathy feeling of the users, although the differences on the perception of the other characteristics has shown no statistical significance.

Concerning the question about having felt empathy, the mean of the differences of the two approaches was 0.36, with standard deviation 0.67. The t-test obtained $p = 0.052$, which shows a strong tendency towards statistical significance (given that, by convention, it is desirable that $p < 0.05$).

This result points that the described approach is likely to increase the feeling of empathy of a human user towards an artificial agent. Stronger evidence may be provided by improving the emotion recognizer, expanding the application, which is still a prototype, and running further experiments to achieve actual statistical significance.

The evaluation of other desirable features led to p-values farther from the conventional 0.05 threshold. TABLE I shows the means of the differences between Claire and Rachel for each feature, the respective standard deviations and the obtained p-values.

TABLE I. Evaluation Results

Feature	Mean	Standard deviation	p-value
Empathy	0.36	0.67	0.05
Realism	0.18	0.98	0.28
Pleasantness	0.18	1.17	0.31
Kindness	0.09	0.70	0.34
Competence	0.09	0.54	0.29

VI. Conclusion and Future Works

In this article, it has been proposed and developed an environment for human-robot interaction based on emotion recognition, which produces adaptive interaction strategies. The environment suggests, proactively, behaviors and content that may be interesting for the user, during the activities provided. The inferred emotional state of the user might determine the interaction strategy and influence the behavior selection.

User studies were performed for evaluating whether users perceive the proposed dynamic, with two interaction strategies combined on an emotion-aware approach, as more interesting than a static strategy, and how this may affect their feeling of empathy towards the virtual agent. The results were promising on conceiving a more empathic interaction, but improvements must be made to achieve stronger evidence.

Future works will include, besides improving the emotion recognizer and finishing the application, conducting studies with physical robots and investigating the differences when applying the described approach. To do so, we are going to

 ISBN: 978-1-61208-462-6

connect mobile devices to small robots and introduce some non-verbal behaviors.

ACKNOWLEDGMENT

FAPESP (São Paulo State Research Support Foundation) supports this work, under grant 2014/16862-4.

REFERENCES

[1] S. Brave and C. Nass, "Emotion in human-computer interaction," in The human-computer interraction handbook, 2002, pp. 53-58.

[2] A. Damasio, Descartes' error: Emotion, reason, and the human brain, Penguin Books, 2005.

[3] R. W. Picard, Affective computing, MIT Press, 2000.

[4] T. Holz, M. Dragone, and G. M. O'Hare, "Where robots and virtual agents meet," International Journal of Social Robotics, vol. 1, no. 1, 2009, pp. 83-93.

[5] M. J. Mataric, The Robotics Primer, MIT Press, 2007.

[6] F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer, "Understanding social robots," in ACHI'09. Second International Conferences on Advances in Computer-Human Interactions, 2009, pp. 169-174.

[7] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings," in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, 2012, pp. 367-374.

[8] A. M. Rosenthal-von der Pütten, N. C. Kramer, L. Hoffmann, S. Sobieraj, and S. C. Eimler, "An experimental study on emotional reactions towards a robot," International Journal of Social Robotics, vol. 5, no. 1, 2013, pp. 17-34.

[9] D. Jo, J. Han, K. Chung, and S. Lee, "Empathy between human and robot?," in Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction, 2013, pp. 151-152.

[10] E. Russell, A. Stroud, J. Christian, D. Ramgoolam, and A. B. Williams, "SMILE: A Portable Humanoid Robot Emotion Interface," in 9th ACM/IEEE International Conference on Human-Robot Interaction, Workshop on Applications for Emotional Robots, HRI14, Bielefeld University, Germany, 2014, pp. 1-5.

[11] P. Ekman, "Basic emotions," Handbook of cognition and emotion, vol. 4, 1999, pp. 5-60.

[12] J. A. Russell, M. Lewicka, and T. Niit, "A cross-cultural study of a circumplex model of affect.," Journal of personality and social psychology, vol. 57, no. 5, 1989, pp. 848-856.

[13] G. Libralon and R. A. Romero, "Geometrical facial modeling for emotion recognition," in The 2013 International Joint Conference on Neural Networks (IJCNN), 2013, pp. 1-8.

[14] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," International Journal of Computer Vision, vol. 91, no. 2, 2011, pp. 200-215.

[15] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," Cognition and Emotion, vol. 24, no. 8, 2010, pp. 1377-1388.

[16] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94-101.

[17] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Machine learning, vol. 47, no. 2-3, 2002, pp. 235-256.

 ISBN: 978-1-61208-462-6