Analysis of Emotional Gestures for the Generation of Expressive Copying Behaviour in an Embodied Agent

Ginevra Castellano[1,2] and Maurizio Mancini[2,3]

[1] Department of Computer Science, Queen Mary, University of London, UK
[2] InfoMus Lab, DIST - University of Genova
Viale Causa 13, I-16145, Genova, Italy
ginevra@dcs.qmul.ac.uk
[3] LINC, IUT de Montreuil, University of Paris8, France
maurizio@infomus.org

Abstract. This paper presents a system capable of acquiring input from a video camera, processing information related to the expressivity of human movement and generating expressive copying behaviour of an Embodied Agent. We model a bi-directional communication between user and agent based on real-time analysis of movement expressivity and generation of expressive copying behaviour: while the user is moving, the agent responds with a gesture that exhibits the same expressive characteristics. An evaluation study based on a perceptual experiment with participants showed the effectiveness of the designed interaction.

Keywords: gesture expressivity; emotion; embodied agent.

1 Introduction

In human-computer interaction the ability for systems to understand users' behaviour and to respond with appropriate feedback is an important requirement for generating an affective interaction [1]. Systems must be able to create a bi-directional communication with users: analysing their verbal and non-verbal behaviour to infer their emotional states and using this information to make decisions and/or plan an empathic response.

Virtual agent systems represent a powerful human-computer interface, as they can embody characteristics that a human may identify with [2]. The ability for virtual agents to display a social, affective behaviour may encourage users to engage in a natural interaction and establish bonds with them [3]. In this paper we focus on modelling a bi-directional communication between an embodied conversational agent and human users based on the non-verbal channel of the communication involving movement and gesture. Specifically, our research considers movement and gesture expressivity as a key element both in understanding and responding to users' behaviour.

We present a system capable of acquiring input from a video camera, processing information related to the expressivity of human movement and generating expressive copying behaviour in real-time. Our system has been formed by the integration of two different software platforms: EyesWeb [4] for video tracking and analysis of human

M. Sales Dias et al. (Eds.): GW 2007, LNAI 5085, pp. 193–198, 2009.

movement and the Greta embodied agent for behaviour generation [5]. In this context, we model a bi-directional communication between user and agent: for each gesture performed by the user, the agent responds with a gesture that exhibits the same expressive characteristics. We describe a mapping between the expressive cues analysed in human users and the corresponding expressive parameters of the agent. Moreover, movement expressivity is mapped onto emotions, which are used to select the type of gesture performed by the agent.

In the following sections we report some examples of related work and a description of the system. Finally, we provide an evaluation study based on a perceptual experiment with participants and show results and conclusions.

2 Related Work

Several systems have been proposed in the literature in which virtual agents provide visual feedback or response by analysing some characteristics of users' behaviour. In these systems the input data can be obtained from dedicated hardware (joysticks, hand gloves, etc), audio and video sources. SenToy [6] is a doll with sensors in the arms, legs and body. According to how users manipulate the doll, they can influence the emotions of characters in a virtual game. Kopp et al. [7] designed a virtual agent able to imitate natural gestures performed by humans using motion-tracked data. Reidsma and colleagues [8] designed a virtual rap dancer that invites users to join him in a dancing activity. In a previous study the Greta agent was animated off-line based on the motor behaviour of users extracted from visual data [9].

Previous works [10] introduced the notion of expressivity for virtual agents. The expressivity of the Greta agent is defined over a set of six dimensions, which qualitatively modify the gestures performed by the agent. For the purposes of the presented work, we only focus on four of these parameters: *Spatial Extent*, that changes the amplitude of movements (e.g., expanded versus contracted); *Temporal Extent*, that modifies the duration of movements (e.g., quick versus sustained actions); *Fluidity*, that determines the smoothness and continuity of movement (e.g., smooth, graceful versus sudden, jerky); and *Power*, that alters the dynamic properties of the movement (e.g., weak/relaxed versus strong/tense). In this paper we describe the first application on which the Greta agent responds to users' motor behaviour analysed in real-time.

3 Description of the System

In Figure 1 we present a system that allows for the real-time analysis of human movement and gesture expressivity and the generation of expressive copying behaviour in an agent. The system integrates two different platforms: EyesWeb [4] for video tracking and movement analysis and the Greta agent for behaviour generation [5]. We designed an application in which data is exchanged between these two components through a blackboard structure, implemented with Psyclone [11]. The system modules are connected to the same Psyclone blackboard via a TCP/IP socket. Further details about how the modules work are provided by the following Sections.

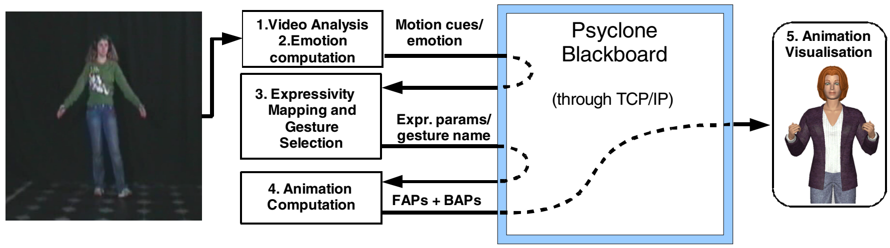

Fig. 1. Overview of the system architecture

3.1 Video Analysis

The automatic extraction of the expressive motion cues is done with EyesWeb [4] and the EyesWeb Expressive Gesture Processing Library. Based on previous studies on emotion and bodily expressions [12]:

- *Contraction Index* (CI) is a measure of the degree of contraction or expansion of the user's body. CI can be computed using a technique related to the bounding region, i.e., the minimum rectangle surrounding the body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette.
- *Velocity* and *Acceleration* are computed on the trajectory of the user's left or right hand in the 2D plane of the video image provided as input to EyesWeb. We extract the 2D coordinates of the hand and we compute the module of velocity and acceleration.
- *Fluidity* is considered as a measure of the uniformity of motion, so that fluidity is maximal when, in the movement between two specific points in the space, the acceleration is equal to zero.

3.1.1 Time Window

The expressive motion cues extracted by EyesWeb in real-time are *continuous*: that is, we do not segment the user's movement by looking for example for the occurrence of a gesture or posture change. Instead, the output of our system is an agent that performs a sequence of *discrete* gestures, each of them with its own expressivity and shape. So, in order to use continuous values (the movement cues) to drive discrete events (the agent's gestures) we defined a *time window*, an interval of time at the end of which we send the computed motion cues from EyesWeb to the rest of the system.
The duration of the time window corresponds to the duration of a gesture performed by the agent. Inside the time window we continuously extract the motion cues from the user's movement, but we consider the following *local* indicators: the *maximum* for Velocity and Acceleration, the *minimum* of the Contraction and the *average* value of the Fluidity. At the end of the time window we send these indicators to the rest of the system.

3.2 Emotion Computation

After extracting expressive motion cues in real-time, the system associates them with one of the following emotion states: *anger*, *joy*, *sadness*. Based on previous studies on emotion expressivity [12] we defined correspondences between expressive motion cues and emotions: for example anger is chosen when movements are expanded, very fast and accelerates, and so on.

3.3 Expressivity Mapping and Gesture Selection

We defined the one-to-one linear mapping of Figure 2 between the motion cues automatically extracted by EyesWeb and the Greta expressivity parameters: the *Contraction Index* is mapped onto the *Spatial Extent*, since they provide a measure on the amplitude of movements; the *Velocity* onto the *Temporal Extent*, as they refer to the velocity of movements; the *Acceleration* onto the *Power*, as both are indicators of the acceleration of the movements; the *Fluidity* onto the *Fluidity*, as they refer to the degree of the smoothness of movements.

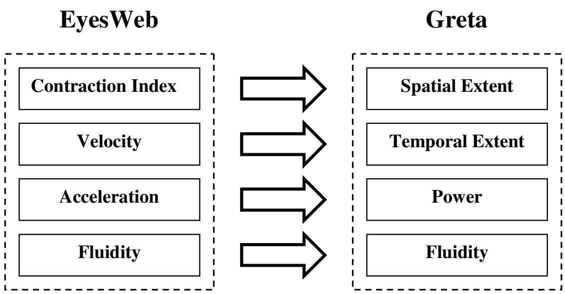

Fig. 2. One-to-one mapping between EyesWeb motion cues and Greta expressivity parameters

Moreover, we select of the gesture to be performed by the agent depending on the emotion determined by our system, as explained in 3.2. If *anger* is detected, the agent performs a deictic gesture; if the detected emotion is *joy*, the agent performs a gesture by opening its arms; if the user's expresses *sadness*, the agent raises and lowers its arms.

3.4 Animation Computation and Animation Visualization

These modules receive the expressivity parameters and the information about the gesture to be performed from the blackboard and compute the animation data needed to animate the agent. Then the Greta graphical engine, given the animation data in input, creates a graphical representation of the virtual agent.

4 Evaluation

In order to evaluate the mapping between the expressive motion cues extracted with EyesWeb and the Greta expressivity parameters, we performed an evaluation study. We selected a set of videos from the GEMEP corpus of acted emotional expressions [13] and we provided these videos as input to our system. More specifically, we focused on six videos of the corpus, with three different emotions (anger, joy, sadness) expressed by two different actors, observed by a frontal camera. For each video provided as input we recorded the video synthesized by our system. That is, we obtained videos in which the Greta agent performs emotion-specific gestures (see the end of Section 3.3) by modulating expressivity. We performed a perceptual experiment to verify how participants perceived the emotions communicated by the agent. Both the actors and the generated videos were modified so as to hide the face of the actor/agent in order not to influence the participants' perception of emotions by the facial expression of the actor/agent.

4.1 Perceptual Experiment

Twelve students and researchers in computer science and computer engineering (8 male and 4 female, from 26 to 38 years old, average age: 31.4) participated in the experiment. Twelve videos were evaluated by each participant: the six videos of the real actors and the six videos of Greta. Participants were asked to observe the actor/agent movements and to choose an emotional label (anger, joy or sadness) using a slider: each emotion could be rated from a minimum of 0 to a maximum of 10. Participants were allowed to watch each video as many times as they wanted and to select and rate all the emotions they wanted. Table 1 shows the recognition rates of each emotion. Results show that emotions expressed by the actors are always recognized with a high percentage, except joy, which is confused with anger. As far as it concerns the gestures performed by the Greta agent, emotion recognition rates are high for sadness (70.8%) and reach the 83.37% for anger and joy.

Table 1. From left to right: recognition rates for anger, joy and sadness when gestures are performed by actors and by Greta

	Recognition rates for ANGER		Recognition rates for JOY		Recognition rates for SADNESS	
	Actors	Greta	Actors	Greta	Actors	Greta
Anger	**95.8%**	**83.3%**	**58.3%**	8.3%	20.8%	20.8%
Joy	0%	16.7%	33.3%	**83.3%**	4.2%	8.3%
Sadness	4.2%	0%	8.3%	8.3%	**70.8%**	**70.8%**
None	0%	0%	0%	0%	4.2%	0%

5 Conclusion

In this paper we modelled a bi-directional communication between human users and an agent based on movement and gesture. We designed a system capable of processing information related to the expressivity of the user's movement, associating an emotion with the expressivity communicated by it and generating the agent's expressive behaviour in real-time. This system represents a first step towards the design of an empathic agent. Movement expressivity can increase the effectiveness of the affective interaction between user and agent by generating an affective loop: it can be used to recognize the user's emotions and to increase the user's sense of engagement when the agent generates a copying behaviour. The evaluation study showed the effectiveness of the designed interaction between user and agent, demonstrating that the analysis and synthesis processes were successful, as well as the choice of the indicators of movement expressivity to communicate emotions.

Acknowledgements. This research was supported by the EU-IST Project HUMAINE (Human-Machine Interaction Network on Emotion) (IST-2002-507422). We thank Tanja Bänziger and Klaus Scherer for the concession of videos of the GEMEP corpus.

References

1. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
2. Reeves, B., Nass, C.: The media equation: How people treat computers, television and new media like real people and places. CLSI Publications, Stanford (1996)
3. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating Rapport with Virtual Agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS, vol. 4722, pp. 125–138. Springer, Heidelberg (2007)
4. Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A., Volpe, G.: Toward real-time multimodal processing: EyesWeb 4.0. In: Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, UK (March 2004)
5. Pelachaud, C.: Multimodal expressive embodied conversational agents. In: MULTIMEDIA 2005: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 683–689. ACM Press, New York (2005)
6. Paiva, A., Chaves, R., Piedade, M., Bullock, A., Andersson, G., Höök, K.: Sentoy: a tangible interface to control the emotions of a synthetic character. In: AAMAS 2003: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pp. 1088–1089. ACM Press, New York (2003)
7. Kopp, S., Sowa, T., Wachsmuth, I.: Imitation games with an artificial agent: From mimicking to understanding shape- related iconic gestures. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS, vol. 2915, pp. 436–447. Springer, Heidelberg (2004)
8. Reidsma, D., Nijholt, A., Poppe, R.W., Rienks, R.J., Hondorp, G.H.W.: Virtual Rap Dancer: Invitation to Dance. In: Olson, G., Jeffries, R. (eds.) CHI 2006, Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, April 25, 2006, pp. 263–266. ACM, New York (2006)
9. Mancini, M., Castellano, G., Bevacqua, E., Peters, C.: Copying behaviour of espressive motion. In: Proc. of the the Mirage Conference (2007)
10. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing expressive gesture synthesis for embodied conversational agents. In: Gibet, S., Courty, N., Kamp, J.-F. (eds.) GW 2005. LNCS, vol. 3881, pp. 188–199. Springer, Heidelberg (2006)
11. Thórisson, K.R., List, T., Pennock, C., DiPirro, J.: Whiteboards: Scheduling Blackboards for Interactive Robots. In: The AAAI 2005 Workshop On Modular Construction of Human-Like Intelligence, AAAI 2005, Pittsburgh, PA, July 9-13 (2005)
12. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology 28, 879–896 (1998)
13. Bänziger, T., Pirker, H., Scherer, K.: GEMEP – GEneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. In: 5th International Conference on Language Resources and Evaluation (LREC 2006), Genova, Italy (2006)