Evaluating the Communication of Emotion via Expressive Gesture Copying Behaviour in an Embodied Humanoid Agent

Maurizio Mancini[1], Ginevra Castellano[2], Christopher Peters[3], and Peter W. McOwan[2]

[1] InfoMus Lab, DIST, University of Genova, Italy
maurizio.mancini@dist.unige.it
[2] School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom
{ginevra.castellano,Peter.McOwan}@eecs.qmul.ac.uk
[3] Department of Computing and the Digital Environment, Coventry University, United Kingdom
christopher.peters@coventry.ac.uk

Abstract. We present an evaluation of copying behaviour in an embodied agent capable of processing expressivity characteristics of a user's movement and conveying aspects of it in real-time. The agent responds to affective cues from gestures performed by actors, producing synthesised gestures that exhibit similar expressive qualities. Thus, copying is performed only at the expressive level and information about other aspects of the gesture, such as the shape, is not retained. This research is significant to social interaction between agents and humans, for example, in cases where an agent wishes to show empathy with a conversational partner without an exact copying of their motions.

Keywords: Expressivity, gesture, emotion, copying behaviour, ECA.

1 Introduction

Several studies from psychology investigated relationships between emotion and movement qualities [4,21]. Nevertheless, automatic affect recognition from body movement remains an under-explored field [8], limiting the investigation of efficient and natural social interaction loops between agents and humans [18].

In previous work [9], expressive motion cues were proposed for communication in human-computer interaction, including *contraction index*, *velocity*, *acceleration* and *fluidity*. This paper contributes a perceptual experiment (Section 4) based on the automatic analysis, mapping and synthesis (Sections 3.1-3.3) of those cues from actors' upper-body motions and gestures to an embodied agent [17], in order to address the following questions:

- **Q1:** Are the proposed expressive motion cues, reproduced at the expressive level by an embodied agent, effective at communicating the same emotion conveyed by the original movement made by a human?

S. D´Mello et al. (Eds.): ACII 2011, Part I, LNCS 6974, pp. 215–224, 2011.

– **Q2:** How does the human perception of emotion change if specific expressivity parameters are not modulated from their 'neutral' values?

These questions are significant for providing important insights into the role of gestures' characteristics in the communication of affective content, important for informing the design of an agent capable of enabling an affective loop based on movement expressivity. The results of the experiment are described in Section 4.4 and implications for the aforementioned questions discussed in Section 4.5.

2 Related Work

Several studies have addressed affect recognition from the automatic analysis of body movement and postures. Studies on affect recognition from body movement include the work by Gunes and Piccardi [11], who developed a system for the recognition of acted affective states based on analysis of affective body displays and automatic detection of their temporal segments, and by Bernhardt et al. [2], Camurri et al. [7] and Castellano et al. [10], who proposed approaches to affect recognition based on the automatic analysis of movement expressivity. Examples of studies that addressed affect recognition from body posture include those of Sanghvi et al., who proposed a computational model to automatically analyse human postures and body motion to detect engagement of children playing chess with a robotic game companion [20], and of Bianchi-Berthouze and Kleinsmith [3], who proposed a model that can self-organise postural features into affective categories to give robots the ability to incrementally learn to recognise affective human postures through interaction with human partners.

Other studies have used embodied agents, either to investigate perception of emotion in individuals [15], or for reacting automatically to affective expressions of the user and systems providing low-level feedback, such as the generation or mimicry of non-verbal behaviour. Examples include the agent created by Maatman and colleagues [14], capable of generating rapport with human speakers by providing real-time non-verbal listening feedback, by Kopp et al. [13], who endowed their agent Max with the ability to imitate natural gestures performed by humans, and by Reidsma and colleagues [19], who designed a virtual rap dancer that invites users to participate in a dancing activity.

The experiment presented in this paper relates to previous studies on affect recognition from the automatic analysis of movement expressivity [7,10].

3 System Overview

The system (see [9] for more details) consists of two integrated software platforms: *EyesWeb XMI* [5] for motion tracking and movement expressivity processing (Section 3.1); and *Greta* [17], an Embodied Conversational Agent (ECA) with a humanoid appearance, capable of generating expressive behaviours (Section 3.2). The system architecture (Figure 1) integrates these two platforms with a module for mapping movement expressivity between a human and agent (Section 3.3).

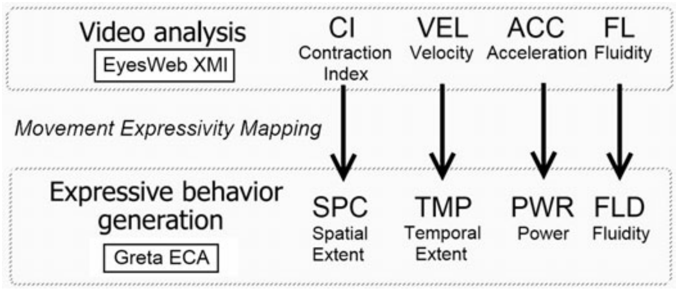

Fig. 1. The system's modules implementing the expressivity copying process

3.1 Video Analysis

Tracking of the full-body and the hands of the user is performed by the system using the EyesWeb Expressive Gesture Processing Library [6] to automatically extract the following low-level expressive motion cues:

Contraction Index - CI is a measure of the degree of contraction and expansion of the body, computed as the minimum bounding rectangle around it.

Velocity - VEL is calculated using the asymmetric backward numeric derivative, given the coordinates in a 2D plane of sampled points in a motion trajectory (here the barycentre of the coordinates of the user's right or left hand's).

Acceleration - ACC is calculated in the same way as Velocity. The asymmetric backward numeric derivative is applied to Velocity samples' x and y components.

Fluidity - FL is applied to the trajectory of the velocity of the hand's barycentre in a 2D plane. It is related the directness, or *Directness Index*, of the trajectory, indicating how direct or flexible it is [6].

3.2 Expressive Behaviour Generation

Hartmann et al. [12] have defined the expressivity of the Greta agent's behaviour over 6 dimensions. Here, 4 out of 6 dimensions varying in $[-1, 1]$ are modulated when expressive gestures are generated:

Spatial Extent - SPC determines the amplitude of gestures, contracting or expanding the agent's wrist end-effector positions according to the sectors illustrated in McNeill's diagram [16].

Temporal Extent - TMP refers to the global duration of gestures. It modifies the speed of execution of all gesture phases: preparation, hold, stroke and retraction.

Fluidity - FLD refers to the smoothness and continuity of movement (e.g., smooth/graceful versus sudden/jerky). It varies the continuity parameter of splines interpolating the agent's hands position.

Power - PWR makes the gesture look more strong/weak or tense/relaxed. It acts on stroke phase duration and agent's hand interpolation curves.

3.3 Movement Expressivity Mapping

The generation of expressive copying behaviour by the Greta agent is based on a movement expressivity mapping which is performed starting from the expressive motion cues automatically extracted during the video analysis. The Contraction Index is mapped onto the Spatial Extent, since they both provide a measure of the amplitude of movements; the Velocity onto the Temporal Extent, as they both refer to the speed of movements; the Acceleration onto the Power, as both are indicators of the acceleration of the movements; the Fluidity onto the Fluidity, as they both refer to the degree of the smoothness of movements.

Note, copying is performed only at the expressive level and information about other aspects of the gesture, such as the shape, is not retained.

4 Experiment

4.1 Overview

An experiment was performed to investigate the extent to which people can recognise the affective content expressed by the agent through the alteration of movement expressivity. It was designed to explore if and how the perception of emotion changes when the agent's expressivity parameters are altered while others remain static. The expressive copying behaviour was generated for the agent according to the process described in Section 3, to address the following:

- **Q1:** Are the proposed expressive motion cues, reproduced at the expressive level by the agent, effective at communicating the same emotion conveyed by the original movement?
- **Q2:** How does the perception of emotion change if some of the agent's expressivity parameters are not modulated from their 'neutral' values?

4.2 Materials

A set of gestures from an extract of videos of the GEMEP (GEneva Multimodal Emotion Portrayals) corpus, a corpus of acted emotional expressions [1], was analysed with EyesWeb XMI. Six videos of the corpus were considered, with three different emotions (*anger*, *joy*, *sadness*) expressed by two different actors observed by a frontal camera. The actors were free to choose the gestures they wished to conduct in each case.

As previously described in Section 3, the actor's expressive motion cues were extracted from the videos and reproduced by the Greta agent. Mapped values used in the experiment are reported in Table 1. Since the selected indicators of movement expressivity and the agent's expressivity parameters vary in different ranges, a rescaling of the former is performed based on maximum and minimum values of the indicators obtained by empirical observation.

For each video of the GEMEP corpus different types of gestures of the Greta agent were synthesised and recorded in videos according to the following specifications:

Table 1. EyesWeb motion cues and Greta's expressivity parameters

		EyesWeb motion cues				Greta's expressivity parameters			
	emotion	CI	VEL	ACC	FL	SPC	TMP	PWR	FLD
actor 1	anger	0.42	1.62	42.00	0.55	-0.04	0.18	1.00	0.55
	joy	0.27	2.73	22.41	0.53	1.00	1.00	0.06	0.00
	sadness	0.51	0.17	4.20	0.43	-0.50	-0.88	-0.80	-0.10
actor 2	anger	0.56	3.17	59.21	0.66	-0.41	1.00	1.00	0.66
	joy	0.30	0.60	13.92	0.49	1.00	-0.62	-0.53	0.50
	sadness	0.65	0.35	5.09	0.36	-0.83	-0.77	-0.83	0.36

Fig. 2. The *beat* performed by the Greta agent. Face has been obscured to ensure that it does not influence the perception of emotions from the participants.

1. Greta performs a *beat* gesture by modulating all the four expressivity parameters from their original neutral values. A beat gesture is a conversational gesture whose shape does not appear to convey any obvious emotional expression or meaning (see Figure 2). Six videos were created in this phase;
2. Greta performs a beat gesture by modulating three out of four expressivity parameters from their neutral values, disregarding at each step one of the four expressivity parameters (i.e., no expressivity control for one of the parameters); twenty-four videos were created in this phase;

Table 2 summarises the gestures considered in the experiments. All videos were modified to obscure the face of the agent in order to avoid influencing participants' judgements.

4.3 Procedure

Twelve students and researchers in computer science (10M:2F, average age: 29 years old) participated in the experiment. Each participant was asked to observe thirty-six videos over six conditions (C1 - C6; see Table 2). These consisted of six videos for each of the conditions C1 to C6, respectively.

A computer was used to show the participants the videos in a random order, different for each participant. Participants were told that they were participating in a study aiming to investigate the relationship between emotions and movement expressivity in an expressive virtual agent. Each participant was presented

Table 2. A total of 36 videos were used, 6 for each condition C1-C6. Gestures performed by actors featured in 6 videos (C1), while the remaining 30 videos contained synthesised gestures performed by the embodied agent (C2-C6).

Condition	Type of Movement	Performer
C1	Freely performed gestures	Actor
C2	Beat gesture: All expressivity parameters modulated	Embodied Agent
C3	Beat gesture: Fluidity not modulated	Embodied Agent
C4	Beat gesture: Power not modulated	Embodied Agent
C5	Beat gesture: Spatial Extent not modulated	Embodied Agent
C6	Beat gesture: Temporal not modulated	Embodied Agent

with the following instructions: "*You will be shown a set of videos in which a real person or a virtual agent performs one gesture. For each video you will be required to observe the body movements of the person or agent, and to evaluate which emotion(s) is/are being expressed*". Participants were asked to observe the gestures in the videos and to associate an emotion label (*anger*, *joy* or *sadness*) with each gesture using a slider: each emotion could be rated in a continuum scale of 1 to 100. Participants were allowed to watch each video as many times as they wanted and could select and rate all of the emotions they wanted or none.

4.4 Results

In order to investigate the effect of the type of movement on the participants' ratings, a one-way analysis of variance (ANOVA) with repeated measures was performed for each rated emotion (the dependent variable) when that specific emotion is expressed and in correspondence with different conditions of type of movement (the independent variable, with six levels: C1 to C6). Means and standard deviations for the one-way ANOVAs are reported in Table 3. Note that, in each one-way ANOVA, the focus is on the participants' ratings corresponding to the same emotion originally expressed by the actors.

Moreover, in order to identify specific differences among the ratings of emotions in correspondence with different conditions of type of movement, pairwise comparisons (Bonferroni corrected) were performed.

Anger: The one-way ANOVA for ratings of anger when anger is expressed by the actors showed a significant main effect of the type of movement [$F(5, 55) = 23.04$, $p < 0.001$][1]. Pairwise comparisons showed that ratings of anger when anger is expressed in correspondence with condition C1 are significantly higher than ratings of anger in correspondence with C2 ($MD = 43.13$, $p < 0.01$), C3

[1] In the case of sphericity violation, effects were Greenhouse-Geisser corrected (for readability reasons, the original degrees of freedom are reported throughout the results).

Table 3. Mean values and standard deviations of ratings of anger, joy and sadness for the different conditions (N = 12 in each condition, all ratings out of 100)

Condition		**Ratings of Anger**	**Ratings of Joy**	**Ratings of Sadness**
C1	Mean	82.00	30.63	46.29
	S.D.	13.40	17.46	26.69
C2	Mean	38.88	66.83	32.96
	S.D.	24.40	24.37	18.81
C3	Mean	41.92	58.75	28.13
	S.D.	23.75	24.94	15.64
C4	Mean	29.38	62.21	43.08
	S.D.	14.54	18.03	19.10
C5	Mean	45.83	39.50	37.17
	S.D.	23.90	17.43	15.24
C6	Mean	38.08	64.33	33.42
	S.D.	20.03	20.01	19.88

(MD = 40.08, $p < 0.01$), C4 (MD = 52.63, $p < 0.001$), C5 (MD = 36.17, $p < 0.01$), and C6 (MD = 43.92, $p < 0.001$).

Joy: The one-way ANOVA for ratings of joy when joy is expressed showed a significant main effect of the type of movement [$F(5, 55) = 11.40$, $p < 0.001$]. Pairwise comparisons highlighted significant differences between the ratings of joy in correspondence of different conditions for type of movement: C1-C2 (MD = -36.21, $p < 0.01$); C1-C4 (MD = -31.58, $p < 0.01$); C1-C6 (MD = -33.71, $p < 0.05$); C2-C5 (MD = 27.33, $p < 0.05$); C4-C5 (MD = 22.71, $p < 0.01$); C5-C6 (MD = -24.83, $p < 0.05$).

Sadness: The one-way ANOVA for ratings of sadness when sadness is expressed showed a significant main effect of the type of movement [$F(5, 55) = 3.34$, $p < 0.05$]. Pairwise comparisons highlighted no significant difference for ratings of sadness among all conditions for type of movement.

4.5 Discussion

The main objective of the experiments was to address two main questions. In the following, the results are discussed in relation to each of them.

Q1: Are the proposed expressive motion cues, reproduced at the expressive level by the agent, effective at communicating the same emotion conveyed by the original movement?

This question addresses the issue of evaluating whether the gestures performed by the Greta agent are associated with the same emotion expressed by the correspondent gestures performed by the actors.

First of all, in the case of anger, there is a significant difference between conditions C1 and C2 when anger is expressed by the actors (higher values in correspondence with C1 and greater than 50%, see Table 3). Values in correspondence with C2 are less than this, approaching 40%: We hypothesise that the type of gesture performed (i.e., the shape, in addition to the expressivity) plays an important role in emotion communication, although this is not the primary focus of this work.

For ratings of joy, there is a significant difference between C1 and C2 when joy is expressed by the actors (higher values for C2). Moreover, Table 3 shows that joy is given ratings greater than 50% for Condition 2 (66.83%), while ratings of joy for Condition 1 is 30.63%. These results suggest that, while joy is not recognised by the participants when the actors perform the gestures, the expressivity of the original movement reproduced in the Greta agent allows them to associate the original affective content with the synthesised gesture.

Finally, in the case of sadness, no significant difference was found when sadness is expressed by the actors. Ratings of sadness are higher for C1 (46.29 %) but they do not reach the 50% (see Table 3). This result suggests that the gesture and related expressivity chosen by the actors to express sadness might have been misleading for the participants. This may also be the reason why ratings of sadness for C2 are not very high (32.96%), but it may also be possible that the expressivity reproduced on the beat gesture performed by the agent did not evoke sadness in the participants.

Q2: How does the perception of emotion change if some of the agent's expressivity parameters are not modulated from their 'neutral' values?

This question refers to whether the percentage of the recognition of emotion by the participants decreases when some expressive motion cues are not modulated from their neutral values by the agent (conditions C3, C4, C5 and C6) in comparison with the case in which the agent performs the *beat* gesture modulating all expressivity parameters.

Results from the one-way ANOVAs for each emotion are of particular note in the case of joy: while for anger and sadness no significant difference between C2 and C3, C4, C5 and C6 was found, when joy is expressed by the actors pairwise comparisons highlighted a significant difference in the ratings of joy between C2 and C5. When the Spatial Extent is not modulated from its neutral value the recognition of joy by the participants decreases significantly. This result suggests that the degree of contraction and expansion is a relevant cue in the expression and recognition of joy, at least in the case of the sample of original and synthesised gestures that was considered in this study. Moreover, Table 3 shows that ratings of joy for C2 are also higher than those for C3, C4 and C6, suggesting that in the case of joy the choice of not modulating from their neutral values all motion cues results in an overall lower percentage of recognition of joy.

Overall, the results provide support for the proposed motion cues in the communication and expression of emotion in human-agent interaction.

5 Conclusion

In this paper we presented an evaluation study of a system for establishing an affective loop between a human user and an ECA: when the user performs an emotional gesture the embodied agent responds by copying the user's gesture expressivity. Results show that expressivity could convey the emotional content of the user's behaviour, i.e., if the agent's expressivity is not altered then emotional content cannot be conveyed effectively.

In the near future we aim to further enrich the agent's response: for example, by analysing the user's gesture expressivity, the agent could perform facial expressions that reflect the user's emotional state. We also aim to reverse the direction of our affective loop. For example, we could implement an embodied agent capable of influencing the user's quality of movement. If the agent perceives that the user's response is not the expected one it could perform gestures that exhibit, for example, faster or slower speeds.

Acknowledgements. We thank Tanja Bänziger and Klaus Scherer for the use of the GEMEP corpus, and Catherine Pelachaud and Antonio Camurri for their support. This work is partly supported by the EU FP7 project LIREC (LIving with Robots and IntEractive Companions; ICT-215554 2008-2012) and the EU FP7 project SIEMPRE (the project SIEMPRE acknowledges the financial support of the Future and Emerging Technologies FET programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 250026-2).

References

1. Banziger, T., Scherer, K.: Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In: Scherer, K.R., Bänziger, T., Roesch, E.B. (eds.) Blueprint for Affective Computing: A Sourcebook. Oxford University Press, Oxford (2010)
2. Bernhardt, D., Robinson, P.: Detecting affect from non-stylised body motions. In: Paiva, A., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 59–70. Springer, Heidelberg (2007)
3. Bianchi-Berthouze, N., Kleinsmith, A.: A categorical approach to affective gesture recognition. Connection Science 15(4), 259–269 (2003)
4. Boone, R.T., Cunningham, J.G.: Children's Decoding of Emotion in Expressive Body Movement: The Development of Cue Attunement. Developmental Psychology 34(5), 1007–1016 (1998)
5. Camurri, A., Coletta, P., Varni, G., Ghisio, S.: Developing multimodal interactive systems with EyesWeb XMI. In: Proceedings of the 2007 Conference on New Interfaces for Musical Expression, pp. 305–308 (2007)
6. Camurri, A., Mazzarino, B., Volpe, G.: Analysis of expressive gesture: The Eyesweb Expressive Gesture Processing Library. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 460–467. Springer, Heidelberg (2004)
7. Camurri, A., Lagerlof, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. International Journal of Human-Computer Studies 59(1-2), 213–225 (2003)

8. Castellano, G., Caridakis, G., Camurri, A., Karpouzis, K., Volpe, G., Kollias, S.: Body gesture and facial expression analysis for automatic affect recognition. In: Scherer, K.R., Bänziger, T., Roesch, E.B. (eds.) Blueprint for Affective Computing: A Sourcebook. Oxford University Press, Oxford (2010)
9. Castellano, G., Mancini, M.: Analysis of emotional gestures for the generation of expressive copying behaviour in an embodied agent. In: Dias, M.S., Gibet, S., Wanderley, M., Bastos, R. (eds.) GW 2007. LNCS (LNAI), vol. 5085, pp. 193–198. Springer, Heidelberg (2009)
10. Kleinsmith, A., Bianchi-Berthouze, N.: Recognising Human Emotions from Body Movement and Gesture Dynamics. In: Paiva, A., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 48–58. Springer, Heidelberg (2007)
11. Gunes, H., Piccardi, M.: Automatic temporal segment detection and affect recognition from face and body display. IEEE Transactions on Systems, Man and Cybernetics - Part B 39(1), 64–84 (2009)
12. Hartmann, B., Mancini, M., Buisine, S., Pelachaud, C.: Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: Third International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS), Utretch (July 2005)
13. 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 (LNAI), vol. 2915, pp. 436–447. Springer, Heidelberg (2004)
14. Maatman, R.M., Gratch, J., Marsella, S.: Natural behavior of a listening agent. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 25–36. Springer, Heidelberg (2005)
15. McDonnell, R., Joerg, S., McHugh, J., Newell, F., O'Sullivan, C.: Investigating the role of body shape on the perception of emotion. ACM Transactions on Applied Perception 6(3), 14–25 (2009)
16. McNeill, D.: Hand and mind: what gestures reveal about thought. University of Chicago Press, Chicago (1992)
17. Niewiadomski, R., Mancini, M., Hyniewska, S., Pelachaud, C.: Communicating emotional states with the Greta agent. In: Scherer, K.R., Bänziger, T., Roesch, E.B. (eds.) Blueprint for Affective Computing: A Sourcebook. Oxford University Press, Oxford (2010)
18. Petta, P., Pelachaud, C., Cowie, R. (eds.): Emotion-oriented systems: The Humaine handbook. Cognitive Technologies Series. Springer, Heidelberg (2011)
19. Reidsma, D., Nijholt, A., Poppe, R., Rienks, R., Hondorp, H.: Virtual rap dancer: Invitation to dance. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 263–266. ACM, New York (2006)
20. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: ACM/IEEE International Conference on Human-Robot Interaction. ACM, Lausanne (2011)
21. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology 28(6), 879–896 (1998)