

# Correspondence

## Expressive Copying Behavior for Social Agents: A Perceptual Analysis

Ginevra Castellano, Maurizio Mancini, Christopher Peters, and  
Peter W. McOwan

**Abstract**—Successful human interaction commonly involves prototypical exchanges where interactors are engaged, synchronized, and harmonious in their behaviors. The copying of aspects of the other's behavior, at different levels, seems central to establishing and maintaining such empathic connections. Yet, many questions remain unanswered, particularly how it is possible to reflect the same affective content back to the other when the actual motion itself is not exactly the same as theirs. This paper presents a perceptual study in which emotional gestures conducted by an actor were mapped onto synthesized versions generated by an embodied virtual agent. Copying is at the expressive level, where qualities such as the fluidity or expansiveness of gestures are considered, rather than exact low-level motion matching. Participants were later asked to rate the emotional content of video recordings of both the original and the synthesized gestures. A statistical analysis shows that, in most cases, participants associated the emotional content of the agent's gestures with that intended to be expressed by the original actor. The results suggest that a combination of the type of movement performed and its quality is important for successfully communicating emotions.

**Index Terms**—Copying behavior, movement expressivity, social agents.

### I. INTRODUCTION

Machines capable of displaying social affective behavior are becoming increasingly essential for systems involved in direct interaction with human users across a variety of application domains [1]. Examples include artificial companions, where social perceptive abilities seem necessary for enabling more natural interaction with human users [2], and future multimodal audiovisual content search engines, which should better understand nonverbal expressive emotional cues in order to improve the search process [3].

Although progress has been made toward the design of more sociable machines, current prototypes are still lacking in terms of

Manuscript received December 16, 2009; revised June 24, 2010, December 20, 2010, and June 21, 2011; accepted August 14, 2011. Date of publication November 18, 2011; date of current version April 13, 2012. This work was supported in part by the EU FP7 ICT project LIREC (Living with Robots and IntEactive Companions) (grant no. 215554; 2008-2012) and in part by the EU FP7 ICT project I-SEARCH (A unified framework for multimodal content Search) (DG INFSO Networked Media Unit; grant no. 248296; 2010-2012). This paper was recommended by Associate Editor W. Pedrycz.

G. Castellano is with the School of Electronic, Electrical and Computer Engineering, Centre for Human-Computer Interaction, University of Birmingham, B15 2TT Birmingham, Birmingham, U.K. (e-mail: g.castellano@bham.ac.uk).

M. Mancini is with the InfoMus Laboratory, Department of Communication, Computer and System Sciences, University of Genova, 16145 Genova, Italy (e-mail: maurizio@infomus.org).

C. Peters is with the Department of Computing and the Digital Environment, Coventry University, CV1 5FB Coventry, U.K. (e-mail: christopher.peters@coventry.ac.uk).

P. W. McOwan is with the Department of Computer Science, School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K. (e-mail: Peter.McOwan@eecs.qmul.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCA.2011.2172415

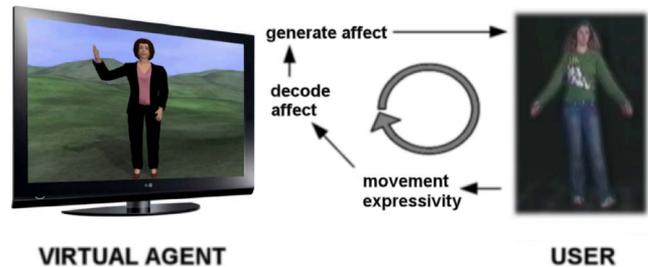


Fig. 1. Establishing an affective loop between (right) a user and (left) an embodied virtual agent requires for the agent to be capable of timely decoding their state in order to generate an appropriate behavioral response. The experiment presented here aims to elucidate some basic aspects toward the challenge of creating such systems, investigating participants' ratings of video recordings of expressive gestures conducted by an actor and synthesized versions conducted by an agent.

important real-time social capabilities and are not yet able to engage in fully natural interaction with the human. Rising to this challenge is a difficult prospect, requiring the machine to be able to adapt to the human, and for the human to reciprocate, creating an *affective loop* [4], [5] for interaction, whereby interactors share a common sense of connectedness, empathizing with each other by displaying congruent emotions [6], [7]. Fundamentally, this facet of social intelligence involves the analysis and generation of appropriate affective cues in a timely manner (see Fig. 1). Additional phenomena may become apparent during such interactions, where interactors seem to become more “engaged,” “synchronized,” and “harmonious” in their exchanges, unintentionally mimicking each other's behaviors, with important prosocial consequences [8].

The task of creating empathizing and theorizing systems [9], [10] for a machine therefore seems desirable but necessarily requires a deeper understanding and elaboration of the fundamentals of human copying behavior: since an exact copying of another's low-level motion does not seem conducive to natural interaction, one may expect only a subset of characteristics or generalities to be conveyed while nonetheless retaining the affective content of the gesture. An important question arising from this consideration is exactly what characteristics of the motion may convey affective content: if the gestures performed by the copier are associated with the same emotion expressed by the original gestures, the method of mapping could be considered valid. Machines capable of achieving this have the possibility to copy the human in more subtle and natural ways, supporting their credibility over longer term interactions.

This paper presents an experiment conducted to investigate the perception of synthesized emotional gestures performed by an embodied virtual agent based on modified originals from an actor. The experiment concerns copying behavior at the expressive level: video recordings of expressive gestures conducted by an actor are analyzed according to a number of lightweight motion cues. These are mapped (see Section III) to expressivity parameters relating to characteristics of the quality of motion, its fluidity or expansiveness, for example, to generate a set of synthesized gestures conducted by the agent. During the experiment (Section IV), participants later rated video recordings of both the original and the synthesized gestures. A statistical analysis shows that, in most cases, participants associated the emotional content of the agent's gestures with that intended to be expressed by the

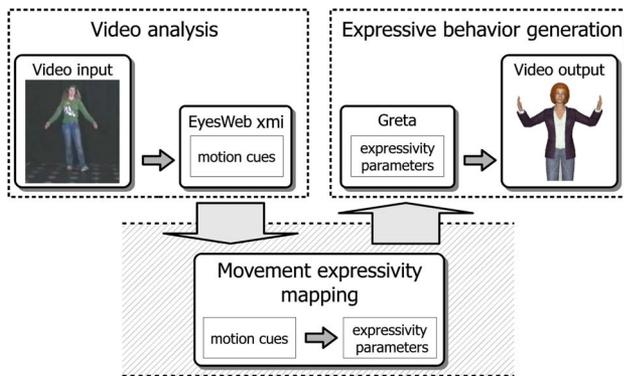


Fig. 2. Overview of the modules and phases involved in the generation of expressive copying behavior. Video recordings of expressive gestures conducted by an actor are analyzed according to a number of lightweight motion cues. These are mapped onto expressivity parameters relating to characteristics of the quality of motion, its fluidity or expansiveness, for example, to generate a set of synthesized gestures conducted by the agent.

original actor. The results (Section V) suggest that a combination of the type of movement performed and its quality is important for successfully communicating emotions.

## II. BACKGROUND

Research from multiple domains is relevant to this paper, particularly automatic analysis and generation of movement, affect and movement expressivity, and social agents, which are as follows.

### A. From Analysis to Generation: Copying Behavior

In terms of analysis, researchers have been addressing the issue of affect recognition, from affective expressions to prototypical emotions and complex mental states [11]. Affect recognition based on the processing of immediately available affective cues displayed by the user has received intensive scrutiny: examples include systems based on automatic analysis of facial expressions [12], [13], expressive body movements [14], and speech [15]. Efforts toward combining multiple modalities of expression for affect recognition have also been reported in the literature [16], as those exploiting contextual information related to task or affective state [17].

Here, we focus on the analysis of affective cues in order to generate copying behavior (see Fig. 2). We use this term to refer to a “loose” categorization of concepts related to copying and synchronization during interaction, such as mirroring [18], imitation [19], mimicry [8], and establishing rapport [20]. Copying may take place at multiple levels of sophistication and speed. At basic levels, for example, one may attempt to directly mirror another’s verbal or nonverbal behavior. At higher levels, the semantics of motions may instead be mapped, so that although gesture shape can change, similar content or meaning is conveyed. For example, in [21], imitation is described at two possible levels: an action level, in which detailed specification of the various motor sequences composing a complex action is imitated, and a program level, in which only the broader more highly structured component of a complex skill is retained. In this paper, we consider copying at the *expressive level*, relating to characteristics of the quality of the motion, as described next.

### B. Affect and Movement Expressivity

Psychological models describe motor expression, which consists of different forms of behavioral patterns including gesture and posture, as one of the components of an emotional process [22]. In human–human

communication, a major role is played by actions that are purely expressive, that is, actions, conscious or unconscious, conducted in order to display feelings and emotion.

Gestures that convey information related to the emotional affective domain are herein referred to as *expressive gestures*. The movement qualities (e.g., velocity, acceleration, energy, fluidity, expansiveness, etc.) responsible for the communication of emotional affective content are referred to as *expressive motion cues* and *movement expressivity* [23] and can be described with what Boone and Cunningham [24] called *propositional aspects of movement*. These do not necessarily convey a specific meaning and are embodied in the direct and natural emotional expression of body movement.

Several studies from psychology investigated expressive body movements. While some studies have found evidence for characteristic body movements accompanying specific emotions [25], others argue that movements may be only indicative of the intensity of emotion but not of its quality [26]. Other studies have mainly investigated the relationship between emotions and movement qualities [24], [27]. Recently, body movement has also started to attract the attention of the affective neuroscience [28] and the automatic affect recognition communities [16], [29]–[33].

### C. Social Agents

Embodied virtual agents [34] are virtual entities with graphical representations that often possess a humanoid 3-D virtual embodiment and are capable of varying the degrees of behavioral autonomy. An important specialization of such agents, when referred to as *embodied conversational agents* (ECAs), is interaction; here, the generation of appropriate gestures, facial expressions, and verbal outputs for the agent provide the potential for a potent and natural means of human–machine interaction, whereby the machine adapts to the human’s natural “interface.” Virtual agents represent a valuable resource in several contexts: they can act as personal assistants, companions, teachers, guides, and entertainers. Empathic agents, endowed with the abilities to perceive and decode affective cues and states of the user (see [35] for example), can use this knowledge to generate more appropriate affective responses for interaction. In the context of the perceptual analysis conducted here, virtual agents are particularly useful as they allow for replicable and exact modifications to be made for mapping analyzed motions onto synthesized gestures, greatly aiding experimentation.

Examples of agents that can react to affective expressions of the user include those providing low-level feedback, such as the generation or mimicry of nonverbal behavior. Maatman *et al.* [36], for example, designed an agent capable of creating a sense of rapport in human speakers by providing real-time nonverbal listening feedback (e.g., head nods and shakes, changes in posture, etc.), including mimicry in response to the speaker’s voice and body movements. Kopp *et al.* [37] endowed their agent Max with the ability to imitate natural gestures performed by humans. Imitation is conducted on two levels: when mimicking, the agent extracts and reproduces the essential form features of the gesture stroke; the second level is a meaning-based imitation level that extracts the semantic content of gestures in order to re-express them with different movements. Reidsma *et al.* [38] designed a virtual rap dancer that invites users to join him in a dancing activity. Users’ dancing movements are tracked by a video camera and guide the virtual rap dancer in his own dance movements and gestures. In the social robotics field, Riek and Robinson [39] designed an expression-mimicking robot to support empathy in human–robot interaction.

Other virtual agent systems infer affective information by perceiving and reasoning about higher level input provided by the user, such as the context of the interaction or the state of a game [40], [41].

### III. METHOD

In this section, the software modules and phases involved in the generation of the video corpus for the experiment (Section IV) are described.

Two software platforms are used in this process: EyesWeb XMI, [42] for motion tracking and processing of movement expressivity, and Greta, an ECA capable of generating expressive gestures [43]. These two platforms, combined with a third module that performs the mapping between the actor's and the agent's movement expressivity [44], are involved in the generation of expressive copying behavior. These three components communicate via Psyclone, a distributed artificial intelligence framework [45] that is used to allow data exchange through a blackboard structure over TCP/IP.

While the system is capable of running in real-time, allowing immediate copying of a user's behavior by the agent, for the purposes of this study, it ran in an offline mode: here, the input to the system consisted of a set of video recordings from the GENEVA Multimodal Emotion Portrayals (GEMEP) corpus [46], [47], featuring an actor performing emotional gestures. The output was a set of video recordings containing the synthesized gestures by an embodied virtual agent for later use in the perceptual study.

The copying process was accomplished in three main phases (Fig. 2).

- 1) *Video analysis*: Expressive motion cues are automatically extracted from the video source by the EyesWeb XMI platform; further details are provided in Section III-A.
- 2) *Expressive behavior generation*: The engine of the Greta module receives the expressivity parameters and gesture data and creates a graphical representation of the agent, which generates a copying of the actor's movement expressivity while performing a gesture, as described in Section III-B.
- 3) *Movement expressivity mapping*: The extracted expressive motion cues are mapped onto the agent's expressivity parameters, and a selection of the gesture to be performed by the agent is made; further description of these steps is provided in Section III-C.

The three phases are described next in further detail.

#### A. Video Analysis

Based on visual input, tracking of the full body and the hands of the actor is performed by the system using EyesWeb XMI. Low-level expressive motion cues are automatically extracted using the EyesWeb Expressive Gesture Processing Library [48], which provides as output continuous temporal series. Global indicators of the full-body movement and the dynamics of the right hand's barycentre are considered. Specifically, the system extracts the following expressive motion cues, derived from psychological studies on the role of movement qualities in emotion expression [24], [25], [27], [49].

- 1) *Contraction index (CI)*: This is a measure, ranging from zero to one, of the degree of contraction and expansion of the body. CI is calculated using a technique related to the bounding region, i.e., the minimum rectangle surrounding the body (see Fig. 3): the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette [50]. If the body is contracted and the limbs remain close to it, the CI is high. Alternatively, if the limbs are fully stretched out, the CI is low.
- 2) *Velocity (VEL)*: This is calculated given the coordinates in a 2-D plane ( $x, y$ ) of sampled points in a motion trajectory (here, the coordinates of the actor's right hand's barycentre) [50]. The module of the velocity is computed starting from its components along the  $x$  and  $y$  axes.

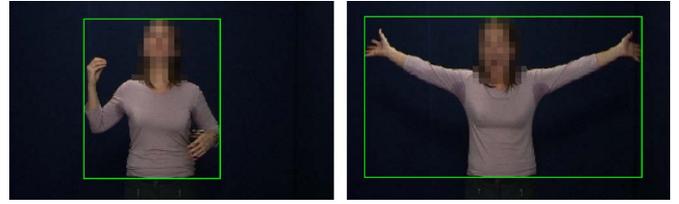


Fig. 3. (Left) High and (right) low values of CI of a posed gesture extracted from the GEMEP corpus [46], [47].

- 3) *Acceleration (ACC)*: This is calculated in the same way as the velocity, given the  $x$  and  $y$  components of the velocity samples. The module of acceleration is obtained from its  $x$  and  $y$  components.
- 4) *Fluidity (FL)*: This is a concept applied to the trajectory of a specific point (e.g., the right hand's barycentre) in a 2-D plane ( $x, y$ ). Starting from the definition of *Directness Index* (a measure of how much a trajectory is direct or flexible; see [50]), fluidity is computed as the directness index of the trajectory followed by the velocity of a point in the 2-D plane. Fluidity provides a measure of the uniformity of motion, so that fluidity is maximum when, in the movement between two specific points in the space, the acceleration is equal to zero.

#### B. Expressive Behavior Generation

Hartmann *et al.* [51] have defined the expressivity of the Greta agent's behavior over six dimensions: Overall Activation, Spatial Extent, Temporal Extent, Fluidity, Power and Repetitivity. Their names intuitively suggest the visual effects that are possibly perceived by the human user when the parameters' values are changed. For example, the Spatial Extent value influences the agent's gesture space volume, the Temporal Extent varies the gesture duration, and so on. These parameters affect the quality of the agent's gestures: expressivity control, described by a single value for each gesture, is applied to the gesture's stroke. Joint angles and timing details required for the low-level specification of the agent's gestures are calculated using the IKAN inverse kinematics package [52]. The expressivity parameters primarily alter the positions and timings of the wrist end-effectors of the agent's arms during inverse kinematics planning; full details are provided in [53].

Here, four parameters are considered. All of them alter the agent's movement compared to its *neutral* movement. In our system, neutral refers to the behavior of the agent without any expressivity control: the agent performs gestures with the amplitude, speed, energy, and fluidity that were defined by the system designer. These four parameters are the following.

- 1) *Spatial Extent (SPC)*: This determines the amplitude of gestures (e.g., expanded versus contracted). It is a real number defined in the interval  $[-1, 1]$ , where  $-1$  corresponds to the generation of very small and contracted movements, while  $1$  corresponds to very wide and large movements. SPC is equal to zero when the behavior of the agent is generated without any expressivity control; in this case, the agent performs the gesture with the amplitude that was defined by the system designer. It is accomplished by contracting or expanding the goal wrist end-effector positions according to the sectors illustrated in McNeill's diagram [54], as well as adjusting the elbow's swivel angle.
- 2) *Temporal Extent (TMP)*: This refers to the global duration of gestures (e.g., quick versus sustained). This parameter modifies the speed of execution of the whole gestures, that is, of all its

phases at the same time: preparation, hold, stroke, and retraction. Its values range in the interval  $[-1, 1]$ : if TMP is equal to zero, then the behavior of the agent is generated without any expressivity control; gestures are slower as the value of the parameter decreases in value and faster as the parameter's value increases.

- 3) *Fluidity (FLD)*: This refers to the smoothness and continuity of movement (e.g., smooth/graceful versus sudden/jerky). This parameter's values range in the interval  $[-1, 1]$ . FLD is equal to zero when the behavior of the agent is generated without any expressivity control. Higher values describe smooth and continuous movement execution, while lower values create discontinuity in the movements. This is accomplished by varying the continuity parameter of Kockanek–Bartels splines used for interpolation. It also alters timing thresholds for retracting the arms to a neutral position beside the waist and may connect a sequence of gestures together to provide a more fluid motion.
- 4) *Power (PWR)*: This determines gesture movement properties that make the gesture look stronger/weaker or more tense/relaxed. Its values range in the interval  $[-1, 1]$ . PWR is equal to zero when the behavior of the agent is generated without any expressivity control. This parameter acts only on the duration of the gesture's stroke phase and on the interpolation curves' shape that are computed to generate the agent's limb movement during this phase. Higher values reduce the duration of the stroke phase (while the overall gesture duration remains the same; see the TMP parameter): this results in higher acceleration of the head and limbs during the stroke phase, making the stroke movement look more powerful and impulsive. The interpolation curves that are computed to generate the agent's limb movement are varied according to the power parameter: we alter the tension and bias parameters of Tension, Continuity, Bias (TCB) interpolation splines, for example, by increasing/decreasing tension as power is increased/decreased.

The remaining two expressivity parameters (Overall Activation and Repetitivity) were not considered for mapping with the expressive motion cues extracted with EyesWeb XMI, as they do not describe the quality of movement but are related to the number of performed gestures.

### C. Movement Expressivity Mapping

The generation of expressive copying behavior by the embodied agent is based on a movement expressivity mapping which is performed starting from the expressive motion cues automatically extracted during the video analysis. The motion cues are mapped onto the agent's expressivity parameters. Fig. 4 shows the correspondence between the automatically extracted motion cues and the embodied agent's expressivity parameters: the CI is mapped onto the SPC, since they both provide a measure of the amplitude of movements; the VEL onto the TMP, as they both refer to the speed of movements; the ACC onto the PWR, as both are indicators of the acceleration of the movements; and the FL onto the FLD, as they both refer to the degree of the smoothness of movements.

Each gesture performed by the actor is described by continuous values of the expressive motion cues. On the other hand, for the embodied agent, expressivity control is applied to the gesture's stroke, which is the culminant phase of the gesture, and expressivity must be described for the agent by a single value for each gesture. For this reason, the design required appropriate indicators to be defined to drive the synthesis process, so as not to lose information about the expressivity of the actor's gestures: the maximum of VEL and ACC, the minimum of the CI (i.e., maximum expansion), and the average

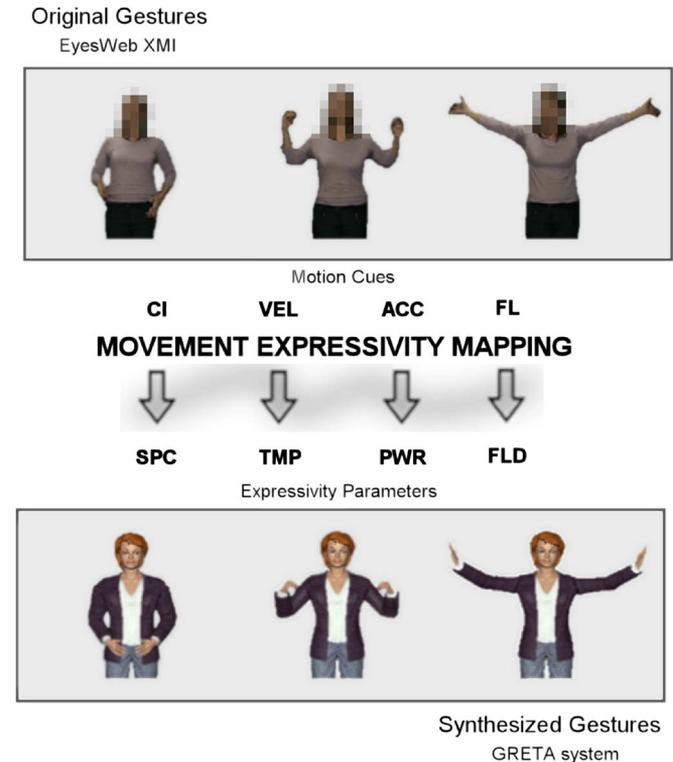


Fig. 4. Example of movement expressivity mapping and copying behavior by the agent. The original gesture displayed in this example was extracted from the GEMEP corpus [46], [47].

value of the FL are considered for each gesture performed by the actor. 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.

The embodied agent generates expressive copying behavior: the agent generates gestures that reproduce the same expressivity of the gestures performed by the actor (see Fig. 4). The copying is performed only at the expressive level, whereas information about the shape (configuration of hand/arm during the gesture execution) of the gesture performed by the actor is not considered. Perceptual experiments reported in Section IV investigated the impact of different types of gestures on the perception of emotions communicated by the agent.

## IV. EXPERIMENT

The overall purpose of the experiment was to verify the extent to which participants could recognize the affective content expressed by the agent using a single channel of information (i.e., movement expressivity). Expressive copying behavior was generated for the agent according to the process described in Section III and evaluated through a perceptual experiment, which is detailed next.

### A. Overview

A set of gestures from an extract of videos of the GEMEP corpus [46], [47], a corpus of acted emotional expressions, was analyzed with EyesWeb XMI. Three videos of the corpus were considered, with three different emotions (*anger*, *joy*, and *sadness*) expressed by an actor observed by a frontal camera. Performances were not scripted: actors were free to use gestures of their own choice. The expressive motion cues described in Section III-A were extracted from the videos and reproduced in the embodied agent following the process reported in Section III (Fig. 4).

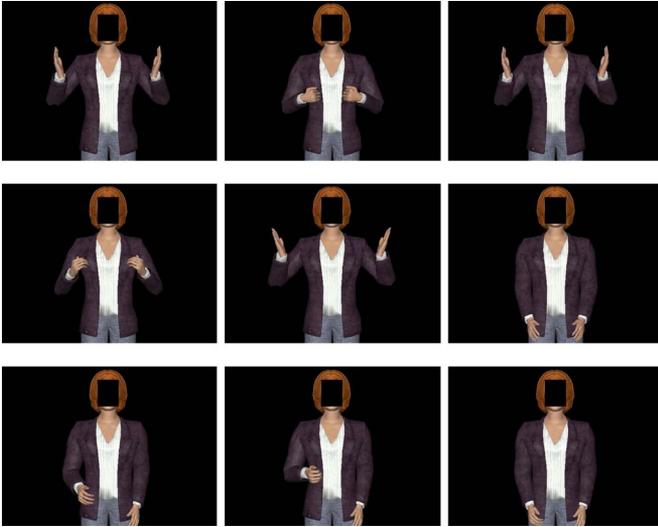


Fig. 5. Three *beat gestures* performed by the embodied agent. The face is hidden so as to ensure that it does not influence the perception of emotions from the participants.

*Beat gestures* were synthesized for use in the experiment. They were chosen as they are conversational gestures whose shape does not convey any obvious emotional expression or meaning [54]. Typical beats are short movements of the hand or fingers up and down. Three different beat gestures were synthesized in the embodied agent: 1) raising the hands in front of the body and performing twice a stroke by opening the hands in an outer movement and back (*beat gesture n.1*); 2) raising the hands in front of the body and performing a stroke by opening the hands in an outer movement (*beat gesture n.2*); and 3) raising and lowering the fist by performing a stroke (*beat gesture n.3*) (Fig. 5). The beat gestures were synthesized and recorded in videos according to the following specifications. Each gesture was synthesized with no expressivity control (neutral baseline) and with the 4 expressivity parameters modulated from their original neutral values based on the expressivity extracted from the emotional gestures performed by the actor, producing 12 videos in total, with 4 videos for each beat gesture: 1 neutral, 1 expressing anger, 1 expressing joy, and 1 expressing sadness. Table I summarizes the gestures considered in the experiment.

Participants were required to indicate which emotion was expressed by the agent in each of the videos, according to its gestures. All the generated videos were modified so as to obscure the face of the agent, to ensure that it did not influence the participants' judgments.

The objective of the experiment was to verify whether the proposed expressive motion cues, reproduced at the expressive level by the agent, are effective at communicating the same emotion conveyed by the original movement.

### B. Pre-experiment User Study

Twenty-six participants (20 males and 6 females, average age: 29 years old) were asked to rate the emotions conveyed by the three gestures performed by the actor in the GEMEP corpus. These were the gestures selected for synthesis with the embodied agent. Participants were asked to rate, for each of the three videos, anger, joy, and sadness on a scale from 1 to 100. The ratings of each emotion were averaged over all the participants. Results showed that participants easily recognized anger (87.7%), while joy and sadness were recognized less successfully (18.9% and 52.1%, respectively). Note that these results

TABLE I  
BEAT GESTURES SYNTHESIZED BY THE EMBODIED AGENT  
CONSIDERED IN THE EXPERIMENT

Condition	Type of Gesture Performed
C1.1 BEAT-1-NEUTRAL	Beat gesture n.1 with no expressivity control (1 video)
C1.2 BEAT-1-ANGER	Beat gesture n.1 with anger expressivity (1 video)
C1.3 BEAT-1-JOY	Beat gesture n.1 with joy expressivity (1 video)
C1.4 BEAT-1-SADNESS	Beat gesture n.1 with sadness expressivity (1 video)
C2.1 BEAT-2-NEUTRAL	Beat gesture n.2 with no expressivity control (1 video)
C2.2 BEAT-2-ANGER	Beat gesture n.2 with anger expressivity (1 video)
C2.3 BEAT-2-JOY	Beat gesture n.2 with joy expressivity (1 video)
C2.4 BEAT-2-SADNESS	Beat gesture n.2 with sadness expressivity (1 video)
C3.1 BEAT-3-NEUTRAL	Beat gesture n.3 with no expressivity control (1 video)
C3.2 BEAT-3-ANGER	Beat gesture n.3 with anger expressivity (1 video)
C3.3 BEAT-3-JOY	Beat gesture n.3 with joy expressivity (1 video)
C3.4 BEAT-3-SADNESS	Beat gesture n.3 with sadness expressivity (1 video)

refer to the average ratings, reported as percentages, of each emotion when that specific emotion was intended to be expressed by the actor.

While these results do not represent a direct baseline against which to compare the results of the main experiment, they provide an indication of how participants perceived the emotions conveyed by the original gestures, and they will be used in the next sections to interpret the main findings.

### C. Experiment

Thirty-seven participants (29 males and 8 females, average age: 24 years old) took part in the experiment. Each participant was asked to observe 12 videos over 12 conditions (see Table I).

The videos were shown to the participants in a random order while ensuring a different sequence for each participant. Participants were informed that the aim of the study was to investigate the relationship between emotions and movement expressivity in an expressive virtual agent. Each participant was presented with the following instructions: “During the experiment, you will see, one at a time, 12 videos. In each video, a virtual agent, whose face has been hidden, performs one gesture. For each video, we ask you to observe the gesture, to pay attention to its expressivity (that is, the way the gesture is performed, rather than the type of gesture performed), and to assess which emotion/emotions has/have been expressed.” Participants were asked to observe the gestures in the video recordings and to provide an assessment of the emotions (*anger*, *joy*, and *sadness*) conveyed by each gesture, on a scale from 0 to 10. Participants were allowed to watch each video as many times as they wanted.

## V. RESULTS AND DISCUSSION

In order to investigate the effect of the type of movement generated by the agent on the participants' ratings of emotions, statistical analysis was performed. An automatic clustering procedure for identification of extreme values found three outliers, leaving a total of 34 samples.

Three repeated measures *t*-tests ( $N = 34$ ) were performed for each beat gesture, for a total of nine *t*-tests. In each *t*-test, the type of gesture performed was considered as the independent variable (two levels), while the emotion rating was considered as the dependent

TABLE II  
DEPENDENT AND INDEPENDENT VARIABLES FOR THE  $t$ -TESTS

	Variable name	Variable type	Levels
Beat n.1 t-test n.1	Type of gesture Emotion rating	Independent Dependent	2: C1.1, C1.2 1: anger
Beat n.1 t-test n.2	Type of gesture Emotion rating	Independent Dependent	2: C1.1, C1.3 1: joy
Beat n.1 t-test n.3	Type of gesture Emotion rating	Independent Dependent	2: C1.1, C1.4 1: sadness
Beat n.2 t-test n.1	Type of gesture Emotion rating	Independent Dependent	2: C2.1, C2.2 1: anger
Beat n.2 t-test n.2	Type of gesture Emotion rating	Independent Dependent	2: C2.1, C2.3 1: joy
Beat n.2 t-test n.3	Type of gesture Emotion rating	Independent Dependent	2: C2.1, C2.4 1: sadness
Beat n.3 t-test n.1	Type of gesture Emotion rating	Independent Dependent	2: C3.1, C3.2 1: anger
Beat n.3 t-test n.2	Type of gesture Emotion rating	Independent Dependent	2: C3.1, C3.3 1: joy
Beat n.3 t-test n.3	Type of gesture Emotion rating	Independent Dependent	2: C3.1, C3.4 1: sadness

variable (one level). Each  $t$ -test aimed to explore whether there is a significant difference in the rating of a specific emotion between a gesture expressing that specific emotion by modulating the expressivity parameters and a gesture performed with no expressivity control. Beat gestures with no expressivity control were used as a neutral baseline against which to compare the emotion ratings of the expressivity-modulated beat gestures. Table II provides an overview of the dependent and independent variables considered in the  $t$ -tests for the three beat gestures. Means and standard deviations for the  $t$ -tests are reported in Table III: significant differences between mean values are highlighted with a star.

**Beat gesture n. 1:** Results showed a significant effect of the type of gesture on the emotion rating. The first  $t$ -test analyzed the difference in ratings of anger between the conditions C1.2 and C1.1. When anger is expressed (C1.2), the ratings of anger are significantly higher than when the gesture is performed without expressivity control (C1.1, neutral baseline) [ $t(33) = 3.59$ ;  $p < 0.001$ ]. The second  $t$ -test analyzed the difference in ratings of joy between the conditions C1.3 and C1.1. When joy is expressed (C1.3), the ratings of joy are significantly higher than when the gesture is performed without expressivity control (C1.1, neutral baseline) [ $t(33) = 2.45$ ;  $p = 0.01$ ]. The third  $t$ -test analyzed the difference in ratings of sadness between the conditions C1.4 and C1.1. When sadness is expressed (C1.4), the ratings of sadness are higher than when the gesture is performed without expressivity control (C1.1, neutral baseline), and the difference approaches significance [ $t(33) = 1.67$ ;  $p = 0.05$ ].

*Conclusion: When the embodied agent performs the beat gesture n. 1 with the expressivity of a particular emotion, the ratings of that emotion by the participants are significantly higher than when the embodied agent performs the same gesture without expressivity control.*

**Beat gesture n. 2:** When the embodied agent performs the beat gesture n.2, there is a significant effect of the type of gesture on the emotion rating of anger and joy, but not sadness. The first  $t$ -test analyzed the difference in ratings of anger between the conditions C2.2 and C2.1. When anger is expressed (C2.2), the ratings of anger are significantly higher than when the gesture is performed without expressivity control (C2.1, neutral baseline) [ $t(33) = 1.77$ ;  $p < 0.05$ ]. The second  $t$ -test analyzed the difference in ratings of joy between the conditions C2.3 and C2.1. When joy is expressed (C2.3), the ratings of joy are significantly higher than when the gesture is performed without expressivity control (C2.1, neutral baseline) [ $t(33) = 4.28$ ;  $p < 0.001$ ]. The third  $t$ -test analyzed the difference in ratings of sad-

ness between the conditions C2.4 and C2.1. When sadness is expressed (C2.4), the ratings of sadness are higher, but not significantly, than when the gesture is performed without expressivity control (C2.1, neutral baseline) [ $t(33) = 0.21$ ;  $p > 0.05$ ].

*Conclusion: When the embodied agent performs the beat gesture n. 2 with the expressivity of a particular emotion, the ratings of that emotion by the participants are significantly higher than when the embodied agent performs the same gesture without expressivity control in the case of anger and joy, but not sadness.*

**Beat gesture n. 3:** Results showed a significant effect of the type of gesture on the emotion rating of anger, but not joy and sadness. The first  $t$ -test analyzed the difference in ratings of anger between the conditions C3.2 and C3.1. When anger is expressed (C3.2), the ratings of anger are higher than when the gesture is performed without expressivity control (C3.1, neutral baseline) [ $t(33) = 5.20$ ;  $p < 0.001$ ]. The second  $t$ -test analyzed the difference in ratings of joy between the conditions C3.3 and C3.1. When joy is expressed (C3.3), the ratings of joy are higher, but not significantly, than when the gesture is performed without expressivity control (C3.1, neutral baseline) [ $t(33) = 0.20$ ;  $p > 0.05$ ]. The third  $t$ -test analyzed the difference in ratings of sadness between the conditions C3.4 and C3.1. When sadness is expressed (C3.4), the ratings of sadness are higher, but not significantly, than when the gesture is performed without expressivity control (C3.1, neutral baseline) [ $t(33) = 0.66$ ;  $p > 0.05$ ].

*Conclusion: When the embodied agent performs the beat gesture n. 3 with the expressivity of a particular emotion, the ratings of that emotion by the participants are significantly higher than when the embodied agent performs the same gesture without expressivity control only in the case of anger.*

Overall, by utilizing the gestures performed without expressivity control as the neutral baseline against which success was evaluated, anger is recognized in all the beat gestures performed by the embodied agent, joy in two beat gestures out of three, and sadness in one. It is to be noted that anger was also the emotion that was best recognized in the gestures performed by the actor in the user study conducted prior to the experiment, while the recognition rates of joy and sadness were lower. Therefore, it is not surprising that this difference emerges also in the emotion rating of the gestures performed by the embodied agent.

Furthermore, the results show that, despite the fact that expressivity does not change, perception of the same emotion may vary across gestures. These results point to the conclusion that gesture type is likely to play an important role in emotion communication, in addition to movement expressivity.

## VI. CONCLUSION

This paper has investigated the role of expressive copying behavior in the perception of emotion in humans and virtual agents. It considers whether synthesized motions may convey the same affective content as original human motions, even when they are not exact copies of the originals. Lightweight motion cues were automatically extracted from video recordings of gestures performed by a human actor and mapped onto expressivity parameters relating to characteristics of the quality of motion of a virtual agent in order to create video recordings of synthesized motions. The results suggest that participants tended to associate the synthesized gestures with the same affective content as the originals conducted by the human actor.

An issue of importance in the reported experiment is the use of a corpus based on acted emotional expressions [46], [47]. It should be noted that video acted portrayals are not ideal, as they are open to varying skill-level and artistic interpretation of the actor and may contain exaggerated expressions. Importantly, however, the situation in which they are recorded can be precisely controlled, and thus, it is

TABLE III  
MEAN VALUES AND STANDARD DEVIATIONS OF RATINGS OF ANGER, JOY, AND SADNESS FOR DIFFERENT CONDITIONS OF TYPE OF GESTURE

	Ratings anger	Ratings joy	Ratings sadness
Beat n.1	* C1.2 — Mean 3.06 S.D. 3.14 * C1.1 — Mean 1.24 S.D. 1.78	* C1.3 — Mean 4.76 S.D. 2.92 * C1.1 — Mean 3.12 S.D. 3.22	* C1.4 — Mean 1.74 S.D. 1.96 * C1.1 — Mean 0.97 S.D. 1.93
Beat n.2	* C2.2 — Mean 2.00 S.D. 2.71 * C2.1 — Mean 1.26 S.D. 1.85	* C2.3 — Mean 5.68 S.D. 3.25 * C2.1 — Mean 3.09 S.D. 3.08	C2.4 — Mean 1.06 S.D. 1.56 C2.1 — Mean 1.00 S.D. 1.60
Beat n.3	* C3.2 — Mean 6.12 S.D. 1.95 * C3.1 — Mean 3.88 S.D. 2.90	C3.3 — Mean 0.68 S.D. 1.77 C3.1 — Mean 0.59 S.D. 1.67	C3.4 — Mean 2.65 S.D. 2.70 C3.1 — Mean 2.26 S.D. 2.45

possible to capture a clear high-quality corpus. Nonetheless, it is desirable in future work to seek ways of eliciting emotional expressions in natural settings. This is a highly challenging research direction, as it investigates copying behavior of more subtle and naturalistic expressions. Basic emotions were considered in this paper due to the explorative nature of this research, but it is expected that movement expressivity plays a key role also in the expression of more subtle emotions.

It is also to be remarked that the experimental results provided evidence that the choice of a particular type of gesture may affect the perception of emotion. Even in cases where the expressivity is the same, a specific gesture type may be more suitable than another to communicate a specific emotion. This result should be taken into consideration in the design of agents engaging in copying behavior, as it suggests that a combination of the type of movement performed and the way the movement is performed is important for a successful communication of emotion. This raises questions regarding the choice of the type of gestures to be used for these purposes, as this should take into account the context of the final application, from the specific interaction scenario, to the characteristics of the target users (e.g., gender, preferences, culture, etc.).

Finally, the overall aim of this paper has been to elucidate some basic aspects toward the overall challenge of creating an affective loop between users and agents. It should be noted that the perceptual study has been conducted in an “offline” mode: participants are not actively engaged in a dynamic real-time interaction with the agent themselves. Further work must better address this issue, to consider the ratings of participants actively involved in the ongoing interaction. This constitutes an intriguing and challenging line of future research, as it would necessarily involve participants’ perceptions of the agent’s copying behavior in the context of their perceptions of their own behavior, involving more complex processes than those considered here.

#### ACKNOWLEDGMENT

The authors would like to thank T. Bänziger and K. Scherer for the videos from the GENEVA Multimodal Emotion Portrayals corpus. The authors would also like to thank a number of people, namely, C. Pelachaud, A. Camurri, M. Demurtas, A. Vasalou, and E. Roesch, for their advice, support, and help.

#### REFERENCES

- [1] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, Sep. 1997.
- [2] C. Breazeal, “Role of expressive behaviour for robots that learn from people,” *Philos. Trans. Roy. Soc. B*, vol. 364, no. 1535, pp. 3527–3538, Dec. 2009.
- [3] EU I-SEARCH Project. [Online]. Available: [www.isearch-project.eu](http://www.isearch-project.eu)
- [4] C. Conati, S. Marsella, and A. Paiva, “Affective interactions: The computer in the affective loop,” in *Proc. 10th Int. Conf. IUI*, New York, 2005, p. 7-7.
- [5] K. Höök, “Affective loop experiences: Designing for interactional embodiment,” *Philos. Trans. Roy. Soc. B*, vol. 364, no. 1535, pp. 3585–3595, Dec. 2009.
- [6] D. Robinson and L. Smith-Lovin, “Emotion display as a strategy for identity negotiation,” *Motiv. Emotion*, vol. 23, no. 2, pp. 73–104, 1999.
- [7] M. Álvarez, R. Galn, F. Matía, D. Rodríguez-Losada, and A. Jiménez, “An emotional model for a guide robot,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 5, pp. 982–992, Sep. 2010.
- [8] R. Van Baaren, R. Holland, K. Kawakami, and A. van Knippenberg, “Mimicry and prosocial behavior,” *Psychol. Sci.*, vol. 15, no. 1, pp. 71–74, Jan. 2004.
- [9] S. Baron-Cohen, “The empathizing system: A revision of the 1994 model of the mindreading system,” in *Origins of the Social Mind*, B. Ellis and D. Bjorklund, Eds. London, U.K.: Guilford Publ. Inc., 2005.
- [10] C. Peters, “A perceptually-based theory of mind model for agent interaction initiation,” *International Journal of Humanoid Robotics (IJHR), Special Issue: Achieving Human-Like Qualities in Interactive Virtual and Physical Humanoids*, vol. 3, no. 3, pp. 321–340, 2006.
- [11] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.
- [12] R. el Kaliouby and P. Robinson, “Real-time inference of complex mental states from facial expressions and head gestures,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.—Workshop Real Time Comput. Vis. Human Comput. Interact.*, 2004, p. 154.
- [13] A. Chakraborty, A. Konar, U. Chakraborty, and A. Chatterjee, “Emotion recognition from facial expressions and its control using fuzzy logic,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 4, pp. 726–743, Jul. 2009.
- [14] G. Castellano, G. Caridakis, A. Camurri, K. Karpouzis, G. Volpe, and S. Kollias, “Body gesture and facial expression analysis for automatic affect recognition,” in *Blueprint for Affective Computing: A Sourcebook*, K. R. Scherer, T. Bänziger, and E. B. Roesch, Eds. Oxford, U.K.: Oxford Univ. Press, 2010.
- [15] B. Schuller, R. J. Villar, G. Rigoll, and M. Lang, “Meta-classifiers in acoustic and linguistic feature fusion-based affect recognition,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Philadelphia, PA, 2005, pp. 325–328.
- [16] H. Gunes and M. Piccardi, “Automatic temporal segment detection and affect recognition from face and body display,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 64–84, Feb. 2009.
- [17] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan, “Detecting user engagement with a robot companion using task and social interaction-based features,” in *Proc. ICMI-MLMI*, Cambridge, MA, 2009, pp. 119–126.
- [18] C. Breazeal, “Regulation and entrainment in human–robot interaction,” *Int. J. Robot. Res.*, vol. 21, no. 10/11, pp. 883–902, Oct. 2002.
- [19] K. Dautenhahn, “Trying to imitate—A step towards releasing robots from social isolation,” in *Proc. Perception Actions Conf.*, 1994, pp. 290–301.
- [20] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency, “Virtual rapport,” in *Proc. 6th Int. Conf. IVA*, Marina del Rey, CA, 2006, pp. 14–27.
- [21] R. Byrne and A. Russon, “Learning by imitation: A hierarchical approach,” *Behav. Brain Sci.*, vol. 21, no. 5, pp. 667–684, Oct. 1998.
- [22] K. R. Scherer, “On the nature and function of emotion: A component process approach,” in *Approaches to Emotion*, K. R. Scherer and P. Ekman, Eds. Hillsdale, NJ: Erlbaum, 1984, pp. 293–317.
- [23] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe, “Multimodal analysis of expressive gesture in music and dance performances,” in *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer Verlag, 2004, ser. LNAI, pp. 20–39.
- [24] R. T. Boone and J. G. Cunningham, “Children’s decoding of emotion in expressive body movement: The development of cue attunement,” *Dev. Psychol.*, vol. 34, no. 5, pp. 1007–1016, Sep. 1998.
- [25] H. G. Wallbott, “Bodily expression of emotion,” *Eur. J. Soc. Psychol.*, vol. 28, no. 6, pp. 879–896, Dec. 1998.

- [26] P. Ekman and W. Friesen, "Detecting deception from the body or face," *J. Person. Social Psychol.*, vol. 29, no. 3, pp. 288–298, 1974.
- [27] M. De Meijer, "The contribution of general features of body movement to the attribution of emotions," *J. Nonverbal Behav.*, vol. 13, no. 4, pp. 247–268, Dec. 1989.
- [28] B. de Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Philos. Trans. Roy. Soc. B*, vol. 364, no. 1535, pp. 3575–3584, Dec. 2009.
- [29] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Proc. 2nd Int. Conf. ACHI*, vol. 4738, ser. LNCS, A. Paiva, R. Prada, and R. W. Picard, Eds., Lisbon, Portugal, Sep. 12–14, 2007, pp. 59–70, Springer-Verlag.
- [30] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Proc. 2nd Int. Conf. ACHI*, vol. 4738, ser. LNCS, A. Paiva, R. Prada, and R. W. Picard, Eds., Lisbon, Portugal, Sep. 12–14, 2007, pp. 71–82, Springer-Verlag.
- [31] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated analysis of body movement in emotionally expressive piano performances," *Music Perception*, vol. 26, no. 2, pp. 103–119, Dec. 2008.
- [32] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact.*, Lausanne, Switzerland, 2011, pp. 305–312.
- [33] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 1027–1038, Aug. 2011.
- [34] J. Cassell, "More than just another pretty face: Embodied conversational interface agents," *Commun. ACM*, vol. 43, no. 4, pp. 70–78, 2000.
- [35] C. Peters, G. Castellano, M. Rehm, E. Andre, G. Volpe, A. Camurri, A. Raouzaoui, K. Rapantzikos, K. Karpouzis, and A. Vasalou, "Fundamentals of agent perception and attention modelling," in *Emotion-Oriented Systems: The Humaine Handbook*, P. Petta, C. Pelachaud, and R. Cowie, Eds. New York: Springer-Verlag, 2011.
- [36] R. M. Maatman, J. Gratch, and S. Marsella, "Natural behavior of a listening agent," in *Proc. 5th Int. Conf. IVA*, Kos, Greece, 2005, pp. 25–36.
- [37] S. Kopp, T. Sowa, and I. Wachsmuth, "Imitation games with an artificial agent: From mimicking to understanding shape-related iconic gestures," in *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer Verlag, 2004, ser. LNAI, pp. 436–447.
- [38] D. Reidsma, A. Nijholt, R. Poppe, R. Rienks, and H. Hondorp, "Virtual rap dancer: Invitation to dance," in *Proc. CHI Extended Abstracts Human Factors Comput. Syst.*, 2006, pp. 263–266.
- [39] L. Riek and P. Robinson, "Real-time empathy: Facial mimicry on a robot," in *Proc. Workshop AFFINE ICMI*, Chania, Crete, Greece, 2009.
- [40] J. C. Lester, S. G. Towns, and P. J. FitzGerald, "Achieving affective impact: Visual emotive communication in lifelike pedagogical agents," *Proc. Int. J. Artif. Intell. Educ.*, vol. 10, pp. 278–291, 1999.
- [41] T. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput.-Human Interact.*, vol. 12, no. 2, pp. 293–327, Jun. 2004.
- [42] A. Camurri, P. Coletta, G. Varni, and S. Ghisio, "Developing multimodal interactive systems with EyesWeb XMI," in *Proc. Conf. New Interfaces Musical Exp.*, 2007, pp. 305–308.
- [43] C. Pelachaud, "Multimodal expressive embodied conversational agents," in *Proc. 13th Annu. ACM Int. Conf. Multimedia*, 2005, pp. 683–689.
- [44] G. Castellano and M. Mancini, "Analysis of emotional gestures for the generation of expressive copying behaviour in an embodied agent," in *Proc. Adv. Gesture-Based Human-Comput. Interact. Simul.: 7th Int. GW, Lisbon, Portugal, May 2007, Revised Selected Papers*, vol. 5085, ser. LNAI, M. S. Dias, S. Gibet, M. Wanderley, and R. Bastos, Eds., Berlin, Germany, 2009, pp. 193–198, Springer-Verlag.
- [45] K. R. Thórisson, T. List, C. Pennock, and J. DiPirro, "Whiteboards: Scheduling blackboards for interactive robots," in *Proc. AAAI Workshop Modular Construct. Human-Like Intell.*, Pittsburgh, PA, Jul. 9–13, 2005.
- [46] T. Bänziger and K. Scherer, "Using actor portrayals to systematically study multimodal emotion expression: The GEMEP Corpus," in *Affective Computing and Intelligent Interaction 2007*, vol. 4738, R. W. Picard, Ed. Berlin, Germany: Springer-Verlag, 2007, ser. LNCS, pp. 476–487.
- [47] T. Bänziger and K. Scherer, "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus," in *Blueprint for Affective Computing: A Sourcebook*, K. R. Scherer, T. Bänziger, and E. B. Roesch, Eds. Oxford, U.K.: Oxford University Press, 2010, pp. 271–294.
- [48] A. Camurri, B. Mazzarino, and G. Volpe, "Analysis of expressive gesture: The Eyesweb Expressive Gesture Processing Library," in *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915, A. Camurri and G. Volpe, Eds. Berlin, Germany: Springer Verlag, 2004, ser. LNAI, pp. 460–467.
- [49] H. G. Wallbott and K. R. Scherer, "Cues and channels in emotion recognition," *J. Person. Soc. Psychol.*, vol. 51, no. 4, pp. 690–699, 1986.
- [50] G. Volpe, "Computational models of expressive gesture in multimedia systems," Ph.D. Dissertation, Faculty of Eng., Univ. Genova, Genoa, Italy, Apr. 2003.
- [51] B. Hartmann, M. Mancini, S. Buisine, and C. Pelachaud, "Design and evaluation of expressive gesture synthesis for embodied conversational agents," in *Proc. 3rd Int. Joint Conf. AAMAS*, Utrecht, Jul. 2005, pp. 1095–1096.
- [52] D. Tolani, "Inverse kinematics methods for human modeling and simulation," Ph.D. dissertation, Univ. Pennsylvania, Philadelphia, PA, 1998.
- [53] B. Hartmann, M. Mancini, and C. Pelachaud, "Implementing expressive gesture synthesis for embodied conversational agents," in *Proc. Gesture Workshop*, 2005, pp. 188–199.
- [54] D. McNeill, *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: Univ. Chicago Press, 1992.