CHAPTER 11

HUMAN AND VIRTUAL AGENT EXPRESSIVE GESTURE QUALITY ANALYSIS AND SYNTHESIS

Radoslaw Niewiadomski[1], *Maurizio Mancini*[2], *Stefano Piana*[2]
[1] Telecom ParisTech - 37, rue Dareau, 75014 Paris, France
niewiado@telecom-paristech.fr
[2] InfoMus Lab, University of Genoa, viale Causa 13, 16145 Genoa, Italy
maurizio.mancini@unige.it, steto84@infomus.org

1. Introduction

Nonverbal communication is an essential element of human-human communication, equally important as the verbal message. It consists of bodily *nonverbal signals*, including *facial expressions*, *hand/arm gestures*, *posture shifts*, and so on (Argyle 1998). In particular, the communicative meaning of gestures usually depends on two components: *shape* and *expressive quality*. While the role of the former, for instance the configurations of the hand in time, is well known (McNeill 1996), studies about the latter, i.e., "how a particular mental intention is communicated through gesture's expressive quality", are quite recent. Nevertheless it has been experimentally shown that gesture's expressive qualities may communicate social relations and communicative intentions, such as: emotional states (Castellano et al. 2007), affiliation (Lakens and Stel 2011), cultural background (Rehm 2010), dominance (Varni et al. 2009, Jayagopi et al. 2009), agreement (Bousmalis et al. 2009) or group co Cohesion (Hung and Gatica-Perez 2010). Some authors also suggest that gesture's *expressivity* could be important in the communication of social states such as empathy (Varni et al. 2009) or even sexual interest (Grammer et al. 2000).

In this Chapter, the expression "*expressive gesture quality*" refers to those features of nonverbal behaviors that describe how a specific gesture is performed, for example its temporal dynamics, fluidity or energy. In the domain of *Human-Computer Interaction (HCI)* expressive gesture quality is important at least for two reasons. First of all, researchers try to detect the expressive qualities in human nonverbal behavior and to infer their communicative meaning. In this case the final goal is the recognition of, for example, the user's emotional state or mood. Second, expressive gesture synthesis is studied in the design and implementation of virtual agents, i.e., anthropomorphic autonomous characters displayed, for instance, on the computer screen that use various verbal and nonverbal forms of

communication (Cassell et al. 2000). Virtual agents may modulate their expressive gesture quality to better transmit their communicative intentions, e.g., their emotional state or mood, to the user.
This Chapter presents an overview of studies that enhance the communicative capabilities of human-computer interfaces by taking into account the expressive qualities of nonverbal behavior. It also presents a detailed description of two systems for expressive gesture quality analysis and synthesis in HCI. In more details; the next section is split into three parts: in the first and second part, we review some of the methods described in the literature for expressive gesture quality analysis and synthesis by illustrating various algorithms; in the third part we present systems in which a continuous expressive gesture quality analysis and synthesis *loop* is performed. In Section 3 we present a case study – the analysis and synthesis of some expressive gesture features in the EyesWeb XMI platform for the creation of multimodal applications (Camurri et al. 2007) and the virtual agent called Greta (Niewiadomski et al. 2011). Finally, in Section 4, we provide a conclusive overview by comparing existing algorithms for gesture quality analysis and synthesis in HCI.

2. State of the art

2.1 Expressive gesture quality analysis

In HCI, a central role is played by automated gesture analysis techniques, aiming to extract and describe physical features of human behavior and use them to infer information related to, for example, their emotional state, personality, or social role. The ability for systems to understand users' behavior and to respond to them with appropriate feedback is an important requirement for generating socially tuned machines (Schröder et al. 2011, Urbain et al. 2010). Indeed, the expressive gesture quality of movement is a key element both in understanding and responding to users' behavior.
Many researchers (Johansson 1973, Wallbott and Scherer 1986, Gallaher 1992, Ball and Breese 2000, Pollick 2004) investigated human motion features and encoded them into categories. Some authors refer to body motion using dual qualifiers such as slow/fast, small/large, weak/energetic, unpleasant/pleasant. Behavior expressivity has been correlated to energy in communication, to the relation between temporal/spatial features of gestures, and/or to personality/emotion. Harald G. Wallbott (Wallbott 1998) deems that behavior expressivity is related to the notion of quality of the mental, emotional, and/or physical state, and the intensity of this state. Behaviors do not only encode content information, that is, "*What is communicated*" through a gesture shape, but also expressive information, that is, "*How it is communicated*" through the manner of execution of the gesture.

There exist at least two important aspects of expressive gesture quality analysis, that is, the low level feature detection and its high level interpretation in the terms of its eventual communicative meaning. Both of them have received important contributions in the last years. In next two subsections we present both these aspects.

2.1.1 Expressive gesture features detection

Several low level features were proposed to describe the expressivity of the movement. Theories from arts and humanities, such as for example Laban's Effort theory (Laban and Lawrence 1947) are some of the sources analysis techniques are grounded on. Several algorithms have been proposed to measure the features that can be extracted from a movement. Interestingly, the same features can be computed in many different ways.

The *Spatial Extent* and the *Fluidity* of movement are two such features that are often analyzed. Among others Cardakis and colleagues (Cardakis et al. 2007) analyze the user's gesture extent by measuring the distance between two hands whereas hands' fluidity is computed as the sum of the variance of the norms of the hands' motion vectors. Similarly, (Camurri et al. 2004a) compute the *Contraction Index*, which is the ratio between the area of the minimum rectangle surrounding the actor body and the body silhouette. In (Bernhardt and Robinson 2007)'s work the maximum distance of hand and elbow from body is taken. (Camurri et al. 2004a) approximate movement fluidity with the *Directness Index*, revealing whether a movement follows a straight line or a sinuous trajectory. (Mazzarino et al. 2007) propose two different methods to measure fluidity. First, they estimate hand's fluency by finding gesture start and ending time, then they determine the amount of movement phases in a given time window: the lower the number of phases the higher the fluency. Second, they analyze discrepancies between different body parts' movements: fluency is then evaluated by comparing the quantity of motion of upper and lower body. More recently, Mazzarino and Mancini (Mazzarino and Mancini 2009) propose to estimate human movement *Smoothness* by computing the correlation between trajectory curvature and velocity in a given time window. Smoothness is computed for each frame of a video and for a particular point on the body. Thus it is possible to establish several points (e.g., the two hands) and compute smoothness separately for each of them.

The other group of expressive characteristics of a gesture focuses on different temporal aspects of its realization. For instance temporal aspects of gesture in (Bernhardt and Robinson 2007) are measured by the average hand (and elbow) speed. A slightly different method to estimate the temporal quality of the gesture is used in (Mancini and Castellano 2007). For this purpose the authors compute the velocity of the barycenter of the hand.

Movement *Power* and *Impulsivity* were also addressed by different

computational methods. In (Cardakis et al. 2007) power is the first derivative of the motion vectors, whereas (Mancini and Castellano 2007) operationalize power by the acceleration of the hand. More complex algorithms are used to compute impulsivity of the movement. In (Mazzarino and Mancini 2009) it is characterized as a local peak in the time series of quantity of motion. For this purpose the authors detect any significant rise of quantity of motion in a given time window.
Finally, the *Overall Body Activation* is analyzed by Camurri and colleagues (Camurri et al. 2004a) through the computation of the *Quantity of Motion* (*QoM*). It is measured as the difference of the person's body silhouettes area computed on consecutive video frames. In (Cardakis et al. 2007) the user's movement is estimated as the sum of the motion vectors of color-tracked user's hands.

2.1.2 Communicative meaning of expressive gesture qualities

The high-level meaning of expressive gesture features has been recently investigated, to extract the communicative high-level message of gestures and body movements. (Camurri et al. 2003, 2006, Castellano 2006) classified expressive gesture in human full-body movement (music and dance performances) and in motor responses of participants exposed to music stimuli: they identified parameters deemed important for emotion recognition and showed how these parameters could be tracked by automated recognition techniques.
Other studies show that expressive gesture analysis and classification can be obtained by means of automatic image processing (Drosopoulos et al. 2003, Balomenos et al. 2005) and that the integration of multiple modalities (facial expressions and body movements) is successful for multimodal emotion recognition (Gunes and Piccardi 2005).
Several systems have been proposed in which visual feedback/response is provided by analyzing some features of the users' behavior. In such systems the input data can be obtained from dedicated hardware (joysticks, hand gloves, etc), audio, and video sources. SenToy (Paiva et al. 2003) is a doll with sensors in its arms, legs and body. Several body positions of the doll are associated with emotional states. According to how the users manipulate the doll, they can influence the emotions of characters in a virtual game: depending on the expressed emotions, the synthetic characters perform different actions. (Taylor et al. 2005) developed a system in which the reaction of a virtual character is driven by the way in which the user plays a music instrument. (Kopp et al. 2003) designed a virtual agent able to imitate natural gestures performed by humans using motion-tracked data. When mimicking, the agent extracts and reproduces the essential form features of the gesture stroke, which is the most important gesture phase. (Reidsma et al. 2006) 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.

Castellano and colleagues (Castellano et al. 2007) investigate how emotional states can be communicated through speech, face, and gesture both in a separate and in a joint way. In particular, gestures are analyzed by tracking user's hands and computing some meta-movement features. At first, the user's body silhouette is extracted from the input video by performing background subtraction. Then, hands are localized using skin color tracking and their geometrical barycenter is determined. The authors extract two types of indicators: movement cues and features. Movement cues correspond to data computed directly from points (hands' barycenter) moving on a 2D plane (the video frame), like speed, acceleration, and fluidity. Movement features are meta-indicators, that is, descriptors of the movement cues variation over time, like initial and final slope, maximum value, number of peaks, and so on. These cues are provided as input to a Bayesian classifier determining which emotional state could be associated with the performed gesture. Those Bayesian classifiers are used before and after performing data fusion between modalities. In the first case, the classifiers are applied separately to speech, face, and gesture features. Separate results are then combined via a voting algorithm. In the second case, there is a single classifier that receives all the features coming from different modalities as input.

In (Sanghvi et al. 2011), children's emotional reaction while playing chess with a robot is evaluated by analyzing their upper body movements and posture. For this, computer vision algorithms are applied (e.g., CAMShift) to extract the children's body silhouette from the input video. Body (frontal/backward) lean angle and curvature of the back, Quantity of Motion, and a Contraction Index are determined. Then, the authors compute the 1st, 2nd, and 3rd derivative of each feature's time-series and the histograms of each derivative. The result is a stream of features that is classified using sixty-three classifiers. The best results are obtained by the ADTree and OneR classifiers and they show that Quantity of Motion and the 2nd derivative of the movement features are the most significant indicators in discriminating the user's emotional state.

(Kleinsmith and Berthouze 2011) revise psychological and neuroscientific works demonstrating the importance of both gesture movement and form in the process of human affect recognition. By showing participants a set of, for example, emotional upright and upside-down reversed videos, researchers prove that emotion recognition is still possible but with lower rate, revealing the contribution of form in the process. The authors present a system for recognition of affective postures based on sequences of static postures. Non-acted expressions of affect are collected by motion capturing the body joints rotations of human video game players at the time when the game is won or lost. The system is composed by two separate modules: the

first one for classification of static postures and the second one for classification of a sequence of postures. Each posture is described as a vector of body joints rotations. The posture classification module provides as output the probability distribution for three states: defeated, triumphant, and neutral. Then a decision rule is applied to each of the three states in the complete sequence of postures, providing the cumulative probability that each state is present in the sequence.

Previous work from the same authors (Kleinsmith et al. 2011) focuses on the recognition of non-acted affective states grounded only on body postures. In particular, they present models based on low-level description of body configuration. Each body posture is represented by a vector of features: each feature represents the normalized rotation of one of the body joints around one of the three axes (e.g., rotation of the left/right shoulder/elbow around the x/y/z axis). Then models for automatic recognition of four emotional states (frustrated, triumphant, concentrating, and defeated) are defined by providing 103 postures represented by their vectors of low-level features and the corresponding label.

In (Berthouze 2012), the body movements of video game players are analyzed both by observers and from motion capture data to understand their role, e.g., movements that are functional to the game vs. movements that express affect. The motion capture data consists of the players' body joints rotation and the amount of body movements is computed by a normalized sum of all of the joints over a game session.

2.2 Expressive gesture quality synthesis

Two approaches for expressive gesture generation are widely used: animation based on motion capture data and procedural animation. First of all, expressive movement can be re-synthesized from motion capture data. An example of such an approach is proposed by (Tsuruta, Choi, Hachimura 2010) to generate emotional dance motions. In this work the authors parameterize "standard" captured motions by modifying the original speed of motion or altering its joint angles. Emotional dance motions are parameterized by a small number of parameters obtained empirically. Five emotional attitudes are considered: neutral, passionate, cheerful, calm, and dark. The parameters influence directly the joint values of a very simple body model consisting of 6 degrees of freedom (DOF), namely knees, waists, and elbows.

Several models were proposed for procedural animation. In Allbeck and Badler (2003), the choice of nonverbal behavior and the movement quality depends on the agent’s personality and emotional state. The way in which the agent performs its movements is influenced by a set of high-level parameters derived from Laban Movement Analysis (Laban and Lawrence, 1947), and implemented in the Expressive Motion Engine, EMOTE (Chi et

al. 2000). The authors use two of the four categories of the Laban's annotation scheme: Effort and Shape. Effort corresponds to the dynamics of the movement and it is defined by 4 parameters: space (relation to the surrounding space: direct / indirect), weight (impact of movement: strong / light), time (urgency of movement: sudden / sustained), and flow (control of movement: bound / free). The Shape component describes body movement in relation to the environment. In Allbeck and Badler's model this component is described using three dimensions: horizontal (spreading / enclosing), vertical (rising / sinking) and sagittal dimension (advancing / retreating). EMOTE acts on the animation as a filter. The model adds expressivity to the final animation. It can also be used to express some properties of the virtual agent or its emotional state. For this purpose, the EMOTE parameters were mapped to the emotional states (OCC model, (Ortony et al. 1988)) and personality traits (OCEAN model, (Goldberg 1993)).

Neff and Fiume (Neff and Fiume 2002, 2003) propose a pose control model that takes into account several features of nonverbal behavior such as the timing of movement, the fluent transition between different poses and its expressive qualities. For each body posture different properties can be defined like its tension, amplitude, or extent. The model allows a human animator to vary, for example, how much space a character occupies during a movement or to define whether the posture should be relaxed or tensed. In more detail (Neff and Fiume 2002) propose a model of tensed vs. relaxed nonverbal behaviors in a physically based animation framework. This model is based on the observed relation between expressive features of movement and forces of gravity and momentum. The system allows an animator to control explicitly the tension for each DOF in the character animation by taking into account the gravity and external forces. Consequently, tensed movements are short, more accelerate and without overshoot, while relaxed ones start slowly and with a delay (necessary to overcome inertia) and finish with a visible overshoot.

In a more recent work (Neff and Fiume 2003), the same authors focus on three different aspects of movement i.e., timing, amplitude, and spatial extent of gesture. First, they propose a sequential realization of the movements where different joints are no longer animated at the same time but they move in a sequence (i.e., some of the joints follow the others) to make movement more fluid (natural). The human animator can specify the offset and the type (forward or reverse successions of joints) manually; then the system automatically propagates the time shifts to all joints, which for this purpose are organized in a hierarchical structure. They also model the amplitude of the movements i.e., they adjust the ranges over which a motion occurs. The amplitude of the movements is modeled by multiplying the distance of the joints for each pose (keyframe) of the animations. For each inter-pose two deltas are calculated: one measuring their distance from

the average to the end of the pose and the second measuring the distance from the average to the end state of the previous pose, then they are multiplied by the amplitude factor. The average values are applied to poses from the second to n-1 in a sequence. A similar mechanism is used to calculate the spatial extent, i.e., the space where the action is realized.

(Hartmann et al. 2005) define and implement a set of parameters that allow one to alter the way in which an agent expresses its actual communicative intention. This model is based on perceptual studies conducted by Wallbott and Scherer (Wallbott and Scherer 1986, Wallbott 1998) and Gallaher (Gallaher 1992). These works define a large number of dimensions that characterize gesture expressivity. Hartmann and colleagues (Hartmann et al. 2005) implement six of these dimensions (see Section 3 for more details). Three of them, namely spatial extent, temporal extent and power, act on the parameters defining the gestures and the facial expressions. They modify respectively the amplitude of a signal (that corresponds to the physical displacement of a facial feature or the wrist position), the movement duration (linked to the execution velocity of the movement), and the dynamic properties of movement (namely acceleration). Another dimension, fluidity, modifies the movements trajectory as well as works over several behaviors of a given modality. In the latter case it specifies the degree of fluidity between consecutive behaviors. The last two dimensions: overall activity and repetitivity refer to the quantity of signals and to their repetition.

Szczuko and colleagues (Szczuko et al. 2009) propose a fuzzy controller that can be used to modify a key-frame based animation by applying two expressive features: fluidity and level of exaggeration. In this approach an animation is modified by adding some additional gesture phases, namely preparation, overshoot, and phase movement hold. The preparation is a slight movement in a direction opposite to the main direction while the overshoot is an additional movement of the last bone in the chain that overpasses the target position and goes back. According to the authors adding these phases to an animation modifies its expressive quality without changing the communicative meaning of the displayed behavior. In the paper, the authors propose an approach that allows a human animator to specify explicitly the fluidity of animation as well as its expressiveness through a set of discrete values: {fluid, middle, abrupt} and {natural, middle, exaggerated}. This approach is based on fuzzy rules defined in a perceptual study. According to the given values of expressiveness and fluidity the fuzzy rules control the duration and the amplitude of additional phases.

A similar approach, proposed in (Kostek and Szczuko 2006), can be used to generate emotionally characterized body animations. In this approach the authors, first manually create a set of animations of certain gestures. Second, the emotional content and intensity of these animated gestures are

evaluated in a perceptive study. Third, in order to find the relation between perceived emotion and the low-level features of evaluated gestures animations the authors apply rough set exploration algorithms on the low-level features of animated gestures (i.e., the amplitude, the length and the speed of every joint's movement). The authors show that the most significant features for communicating emotional content are related to the duration of the gesture phases (stroke position, duration of hold, preparation and stroke phase). They argue that the duration of the gesture phases is the most significant feature in communicating emotional states; it is more important than the amplitude of the movement. Finally they also propose values of low-level gesture features that can be applied to any gesture animation in order to modify its emotional content.
(Hsieh and Luciani 2005, 2006) use a physically based particle modeling approach to model several modern dance figures. Their models allow the user to control the expressive qualities of movement such as light vs. strong, free vs. bound, or sudden vs. sustained. In more details, for each dance figure the authors define one physically based particle model. Each figure needs its own description using different set of parameters (e.g., masses, forces) that influences the dynamics of the movement. Thus, every dance figure is defined using different set of mathematical equations that describe its forces, velocity, potential, and kinetic energy. The models describe the energy flow for each figure rather than its spatial criteria. So, instead of defining explicitly its trajectory, a movement is generated implicitly by the involved forces. The single models can be combined in more complex behaviors.

2.3 Expressive gesture quality loop between humans and machines

Applications offering real-time expressive gesture quality analysis and synthesis are becoming widely developed in multimodal interactive scenarios. Their final aim is the creation of credible and natural multimodal interaction between human and machine: that is, they aim to create an *expressive gesture quality loop*, i.e., bidirectional interaction that exploits the communicative role of gesture expressivity, as illustrated in Figure 1.

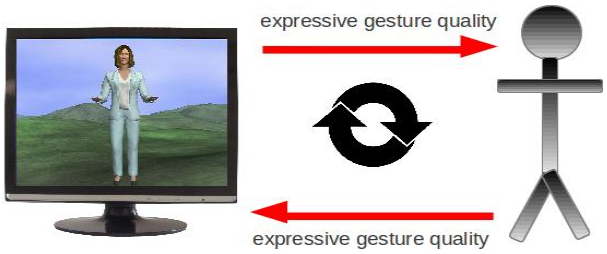

Figure 1: The user's expressive gesture quality is analyzed to influence the agent; in a

symmetrical way, the agent's expressive gesture quality influences the user's behavior.

The idea of an interactive HCI system that intentionally uses the expressive qualities of the behavior was proposed by Caridakis and his colleagues in (Caridakis et al. 2007). Their system allows a virtual agent to mimic the qualitative features of the user's behavior. For this purpose recognized expressive features of the human behavior are mapped to the agent behaviors. Consequently, the agent does not repeat the human movements but its individual behavior is modified to fit the user's expressive behavior profile. The system was only partially implemented and it was not working in a real-time. Thus, natural human-machine interaction was not possible. Instead, the gesture expressivity parameters detected with the gesture analysis module were used to control manually a virtual agent that implements Hartmann's model of expressive behavior (Hartmann et al. 2005; see also Section 2.2 and 3). Caridakis and colleagues (Caridakis et al. 2007) proposed a mapping between the expressive features of human behavior and the agent's expressivity parameters: the sum of the variance of the norms of the motion vectors is associated to the agent's *Fluidity*; the first derivative of the motion vector to *Power*; the distance between hands to *Spatial Extent*; the sum of the motion vectors to *Overall Activity.*

A similar solution was proposed more recently by (Mancini and Castellano 2007). Differently from Cardakis et al.'s work Mancini and Castellano build a truly interactive system that takes as input the video data, extracts high-level behavior features using the EyesWeb XMI platform (Camurri, et al 2007), and finally synthesizes them with a virtual agent. Similarly to (Caridakis et al. 2007), the agent copies only the expressive qualities of the movement of the human but realizes different gestures. The video input is treated with EyesWeb XMI to perform quantitative analysis of the human movement in real-time. The virtual agent uses the expressive model proposed by (Hartmann et al. 2005) (see previous Section). The following mapping between the features detected by EyesWeb XMI and the agent's expressivity quality of movement is then performed: *Contraction Index* is mapped to *Spatial Extent*, *Velocity* of movement to *Temporal Extent*, *Acceleration* to *Power*. Finally *Directness Index*, a measure of the movement straightness, is associated with *Fluidity*.

(Pugliese and Lehtonen 2011) propose the creation of an *enactive loop*: a user and a virtual agent simulate the situation in which two humans are in the same room but they are separated by a glass, so they can communicate only by body movements. But, as it happens in the above applications, the interaction is only at the expressive gesture quality level. The proposed system allows defining a mapping between the user's detected *Quantity of Motion* and *Distance* (the distance of the person from the glass) onto the same agent's movement features. However, such mapping is free, that is,

the agent could simply imitate the user or it could, for example, respond with opposite behaviors (e.g., if the users moves quickly the agent moves slowly and so on).

3. A case study: the EyesWeb XMI platform and the Greta ECA

In this section now introduce a concrete example of two existing systems for expressive gesture quality analysis and synthesis, implementing some of the algorithms described in the previous Sections. The EyesWeb XMI platform is a modular system that allows both expert (e.g., researchers in computer engineering) and non-expert users (e.g., artists) to create multimodal installations in a visual way (Camurri et al. 2007). The platform provides modules, called blocks, that can be assembled intuitively (i.e., by dragging, dropping, and connecting them with the mouse) to create programs, called patches, that exploit system's resources such as multimodal files, webcams, sound cards, multiple displays and so on.
The Greta (Niewiadomski et al. 2011) is a virtual agent able to communicate verbally as well as nonverbally various communicative intentions. Concerning nonverbal communication it is able to display facial expressions, gestures, torso and head movements. Greta is controlled using two XML-like languages BML, and FML-APML. It is a part of several interactive multimodal sytems working in rel-time, e.g., SEMAINE (Schröder et al. 2011) or AVLaughterCycle (Urbain et al. 2010).

3.1 Expressive gesture quality analysis framework

In this Section we describe a framework for multi-user nonverbal expressive gesture quality analysis. Its aim is to facilitate the construction of computational models to analyze the nonverbal behavior explaining the emotions expressed by the users.

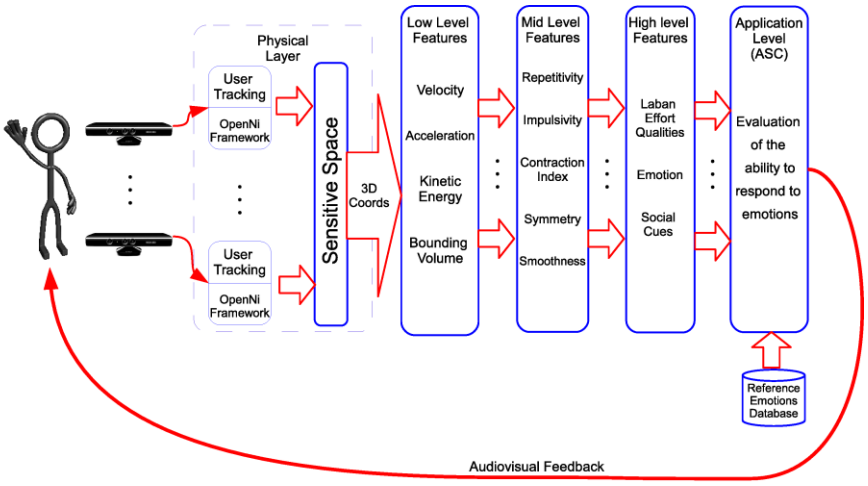

Figure 2: Framework for multi-user nonverbal expressive gesture analysis.

That is, the computed features are used to analyze users' nonverbal behavior, the emotions they express, and the level of social interaction. For this purpose we use the Eyesweb XMI software platform (Camurri et al. 2004b, Camurri et al. 2007), video cameras or Kinect sensors, and we extract several low-level movement features (e.g., movement energy) that are used to compute mid and high-level features (e.g., impulsivity or smoothness) in a multilayered approach, possibly up to labeling emotional states or social attitudes. The choice of these features is motivated by previous works on the analysis of emotion using a minimal set of features (e.g., Camurri et al. 2003, Glowinski et al. 2011).

An overview of the software architecture is sketched in Figure 2. In the following Sections we describe a subset of the proposed framework's components. The first one is the *Physical Layer*, performing measurements on the user's physical position in space as well as user's body joints configuration. The *Low Level Features* include those features directly describing the physical features of movements, such as its speed, amplitude and so on. The *Mid Level Features* can be described by models and algorithms based on the low level features, for example the movement smoothness can be computed given its velocity and curvature.

3.1.1 Physical Layer

For the purpose of the expressive gesture quality analysis described in this case study we use one or more Kinect sensors. Kinect is a motion sensing input device by Microsoft, originally conceived for the Xbox. We implement support for multiple such devices in the EyesWeb XMI platform, enabling to track movement in a larger sensitive space, e.g., tracking a user that moves in separate rooms. This also allows a higher number of users to be tracked

simultaneously: each Kinect device focuses on a different space area, next the data captured by each device are merged to obtain a single sensitive area. The motion capture measurements provided by each Kinect device are then processed to share the same absolute reference system.
Multiple users' detection and tracking is performed thanks to different EyesWeb XMI software modules (blocks). To communicate with the Kinect sensor, EyesWeb XMI supports both the OpenNi framework (version 1.5.4.0) and the Microsoft Kinect SDK (version 1.6). The two APIs support the streaming of both color images and depthmaps captured by the Kinect's optical sensors. A depthmap is a grayscale image where the color intensity represents the distance from the sensor measured in mm. Both OpenNi and Microsoft Kinect SDK support user segmentation and tracking by providing 2D and 3D measurements of multiple users joints: the two API are similar in term of speed and real-time performances (the Microsoft API is slightly less influenced by occlusions), the sets of joints tracked by the two APIs are similar but the Microsoft one can track a bigger number of joints (see Figure 3).
Using the OpenNi Framework, EyesWeb supports the automatic calibration of the user tracking system, and provides functionalities to save configuration files. This feature allows to avoid the tuning phase of Kinect, which consists of the automatic calibration phase requiring from ten to fifteen seconds (during the tuning phase the tracking measurements are less precise).

Figure 3: 2D coordinates of the tracked joints produced by the Miscrosoft SDK (on the left) and by the OpeNi API (on the right).

Two different EyesWeb blocks, called "Kinect Extractor OpenNi" and "Kinect Extractor SDK", were developed to interface with Kinect using the OpenNi or the Microsoft APIs; both blocks provide the data from the Kinect's sensors. Multiple instances of these blocks may be used in a single application in order to use several Kinect devices at the same time; for each device, the outputs provided by the blocks are: a set of tracked users, the image from the color camera or, alternatively, the image from the infrared camera (available only using OpenNi), an image representing the reconstructed depthmap, where the distance from the sensor is mapped to a gray-level in the image. The block developed to support the Microsoft SDK can also output information about face tracking and an audio stream.

3.1.2 Low-level features

Wallbott identified movement expansiveness as a relevant indicator for distinguishing between high and low arousal emotional states (Wallbott 1998). He also observed that the degree of movement energy is an important factor in discriminating emotions. In his study, highest ratings for the energy features corresponded to hot anger and joy while lowest values corresponded to sadness and boredom. De Meijer highlighted that emotional warmth and empathy are usually expressed by open arms (De Meijer 1989). Camurri and colleagues (Camurri et al. 2003) showed that movement activity is a relevant feature in recognizing emotion from the full-body movement of dancers. Results showed that the energy in the anger and joy performances were significantly higher than in the grief ones. From the above studies we defined and implemented two low-level user's full-body movement features: Bounding Volume and Kinetic Energy. These features can be considered a 3D extension of the two previously developed 2D low-level features Contraction Index and Motion Index (or Quantity of Motion) (see Section 2.1 for details):

- *Bounding Volume* (BV) - It is the normalized volume of the smallest parallelepiped enclosing the user's body. Figure 4 shows an example of BV computation. The BV can be considered as an approximation of the user's degree of body "openness": for example, if the user stretches her arms outside or upside, and so the BV, increases.
- *Kinetic Energy* (KE) - It is computed from the speed of the user's body segments, tracked by Kinect, and their percentage mass as referred by (Winter 1990). In particular the full-body kinetic energy KE is equal to:

$$E_{FB} = \frac{1}{2} \sum_{i=0}^{n} m_i v_i^2$$

where m_i is the mass of the *i-th* user's body segment (e.g., head, right/left shoulder, right/left elbow and so on) and v_i is the velocity of the *i-th* segment, computed as the difference of the position of the segment at the current Kinect frame and the position at the previous frame.

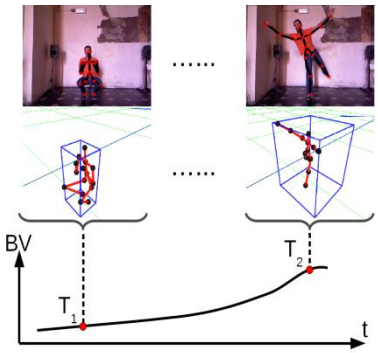

Figure 3: Bounding Volume real-time computation: at time T1 the user has a contracted posture, thus the BV value is very low; at time T2 the user reaches an expanded posture, exhibiting a higher value for the BV.

3.1.3 Mid Level features: Impulsivity Index

Impulsive body movements are those performed quickly, with a high energy and by suddenly moving spine/limbs in a straight direction. We adopt this definition after reviewing some literature about human movement analysis and synthesis. Wilson and colleagues (Wilson et al. 1996) found that stroke gestures have a lower number of execution phases compared to other conversational gestures, suggesting that their execution is shorter in time. In Laban's theory impulsive gestures have quick Time and free Flow, that is, they are executed quickly with energy flowing through body in a consciously uncontrolled way. Finally, Bishko (Bishko 1991) defines impulsive gestures as "an accent leading to decreasing intensity".
The measure we propose for the Impulsivity Index mid-level feature is a combination of the two low-level features Kinetic Energy (KE) and Bounding Volume (BV). KE is firstly used to identify the gesture duration *dt*: for example, using an adaptive threshold, when KE becomes greater than the threshold we identify the gesture beginning time; when KE goes below the threshold we identify the ending time. Then, if KE is higher than a fixed energy threshold and the gesture length *dt* is lower than a fixed time threshold then Impulsivity Index is equal to the ratio between the variation of BV and the gesture length *dt*:

```
let time threshold = 0.45 s;
let energy threshold = 0.02;
if (KE ≥ energy threshold) then evaluate GestureTimeDuration dt;
if (dt ≤ time threshold)) then ImpulsivityIndex = ΔBV/dt;
```

3.2 Expressive gesture quality synthesis

In the virtual agent called Greta (Niewiadomski et al., 2011) expressive gesture synthesis is implemented according to the model proposed by Hartmann and colleagues (Hartmann et al. 2005). The procedural animation generated with Greta can be modified by the use of six high-level parameters: *Overall Activity* (OAC), *Repetitivity* (REP), *Spatial Extent* (SPC), *Temporal Extent* (TMP), *Fluidity* (FLD) and *Power* (PWD).

Two of them, namely the OAC and REP influence the general agents activity. The first one corresponds to the general amount of activity. As this parameter increases (or decreases), the number of nonverbal behaviors increases (or decreases). Each nonverbal behavior has a value associated with it; the OAC parameter is associated to a threshold: only gestures that overpass this threshold are displayed. The second general parameter, REP, specifies the possibility that the gesture stroke will be repeated. In this manner the generation of rhythmic repetitions of the same behavior is allowed.

The four remaining parameters SPC, TMP, FLD and PWR can be specified for each gesture separately. SPC controls the amplitude of movement. It implements McNeill diagram (McNeill, 1996), where wrist positions are defined in terms of 17 sectors. SPC can be used to create expanded vs. contracted gestures. TMP controls temporal aspect of movement. It modifies the speed of execution of gesture phases. Gestures are slow if the value of the parameter has a negative value, and fast when the value is positive. It is calculated according to the Fitts's law (Fitts, 1954) that predicts that the time required to moving to a target area in function of the target distance. Consequently the value of temporal extent changes the duration of each phase, and the execution time of keyframes. The third parameter, FLD, controls the smoothness and continuity of movement (e.g. smooth, graceful versus sudden or jerky). Higher values allow smooth and continuous execution of movements while lower values create discontinuity in the movements. Fluidity is modeled in two different manners. First of all its value influences the interpolation of the gesture, i.e., the trajectory of the movement. For this purpose it is mapped to the continuity parameter of *Tension-Continuity-Bias* (TCB) splines. Secondly, it also influences the transition between two gestures. The higher is the fluidity the more likely two gestures will be fluently co-articulated without passing through a rest intermediate position. Finally the last parameter, PWR, controls the dynamic properties of the movement. Powerful movements are expected to have higher acceleration and deceleration magnitudes. They also should exhibit less overshoot. This effect is obtained by mapping the power value to tension and bias parameters of TCB spline curves. Thus, similarly to the fluidity parameter, power modifies the trajectory of the movement. By controlling the PWR value one may specify gestures that are weak and relaxed or, oppositely, strong and tensed.

4. Conclusion

In this Chapter we focused on the expressive quality of gesture in the human-computer interaction perspective. We discussed two aspects which are complementary: expressive gesture quality analysis and synthesis. We argued that often the same features detected in human behavior are synthesized on virtual agents. At the same time, different computational methods, both for analysis and synthesis, may focus on the same expressive feature. For this reason we conclude the Chapter comparing these features and computational methods, both for analysis and synthesis. The result of this comparison is presented in Table 1.

Table 1 shows that, in particular, Fluidity and Power do not have a single universal interpretation. These features are also difficult to analyze and synthesize. Indeed, in the evaluation of the model by Hartmann and colleagues (Hartman et al. 2005) these two expressive features received the lowest recognition rates when synthesized with the Greta agent. It is easy to notice that sometimes different algorithms model similar features using different names; as it is in the case of Power and Tension.

To summarize this chapter, one can observe that there is a growing interest in the analysis and synthesis of the expressive quality of gesture. It also spreads over other research domains such as robotics (e.g. Le et al. 2011). However more research is needed to specify and model expressive features of nonverbal behavior in a more realistic way. The recent developments of less invasive motion capture based methods seem to be a promising methodology to study expressive gesture quality and to build new expressive models. The creation of efficient broad-consumer tools, such as Kinect, opens new challenges in the domain of the expressive gesture quality analysis.

Measure	Analysis	Synthesis
Spatial aspects of movement	maximum distance of hand and elbow (Bernhardt and Robinson 2007) distance between two hands (Cardakis et al. 2007) minimum rectangle surrounding the body (contraction index, Mancini and Castellano, 2007)	joint distance from the body (Hartmann et al. 2005), (Neff and Fiume 2003)
Temporal aspects of movement	average hand and elbow speed (Bernhardt and Robinson 2007) hand barycenter velocity (Mancini and Castellano 2007)	gesture phase duration computed according to the Fitt's law (Hartmann et al. 2005)

Fluidity, Smoothness	sum of variance of the norms of the motion vectors (Caridakis et al. 2007) approximated by the Directess: ratio between the length of the shortest line and the actual trajectory (Camurri et al. 2004a) discrepancy between movements of different body parts (Mazzarino et al. 2007) relation of the sequence of motion and non-motion phases (Mazzarino et al. 2007)	sequential joint activation (Neff and Fiume 2003) duration of and amplitude of the preparation and overshoot phase (Szczuko et al. 2009) trajectory interpolation procedure (Hartmann et al. 2005) concatenation of two gestures (Hartmann et al. 2005)
Energy, Power, Impulsivity	hand and elbow acceleration (Bernhardt and Robinson 2007) first derivate of motion vectors (Caridakis et al. 2007) hand acceleration (Mancini and Castellano 2007) peak in the time series of quantity of motion (Mazzarino et al. 2009) kinetic energy of user's joints (Piana et al. 2012)	tension and bias of TCB spline lines (Harmann et al. 2005)
Tension, Stiffness		timing and amplitude of gesture phases (Ness and Fiume, 2002)
Overall activity	sum of the motion vectors (Cardakis et al. 2007) Quantity of motion, velocity (Mancini and Castellano 2007)	threshold based gesture selection algorithm (Hartmann et al 2005)

Table 1: A comparison of different approaches for expressive gesture quality analysis and synthesis.

Acknowledgement

We would like to thank Dr. E. Bevacqua from National Engineering School of Brest (France), Dr. G. Volpe from University of Genoa (Italy), and Jennifer Hofmann from University of Zurich (Switzerland) for their valuable comments and suggestions.

This research has been partially funded by the European Community Seventh Framework Programme (FP7/2007-2013) ICT, under grant agreement No. 289021 (ASCInclusion) and grant agreement No. 270780

(ILHAIRE - Incorporating Laughter into Human Avatar Interactions: Research and Experiments, http://www.ilhaire.eu).

BIBLIOGRAPHY

Allbeck, J., & Badler, N. (2003). Representing and Parameterizing Agent Behaviors. In H. Prendinger and M. Ishizuka, (eds.) Life-like Characters: Tools, Affective Functions and Applications, Springer, Germany, 2003, pp. 19-38.

Argyle, M. (1998). Bodily Communication. Methuen & Co., London, 2nd edition.

Ball, G., & Breese, J. (2000). Emotion and personality in a conversational agent. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, (eds.), Embodied Conversational Characters. MIT Press, Cambridge.

Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., & Kollias, S. (2005). Emotion analysis in man-machine interaction systems. In S. Bengio and H. Bourlard, (eds.), Machine Learning for Multimodal Interaction, volume 3361 of Lecture Notes in Computer Science, pp. 318–328, Springer Verlag.

Bernhardt, D. & Robinson, P. (2007). Detecting Affect from Non-stylised Body Motions. In A.C. Paiva, R. Prada, and R.W. Picard (eds.), Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction (ACII '07), Springer-Verlag, Berlin, Heidelberg, 59-70.

Bianchi-Berthouze, N. (2012). Understanding the role of body movement in player engagement. Paper accepted to Human-Computer Interaction.

Bishko, L. (1991). The use of Laban-based analysis for the discussion of computer animation. In The 3rd Annual Conference of the Society for Animation Studies.

Bousmalis, K., Mehu, M., & Pantic, M. (2009). Spotting Agreement and Disagreement: A Survey of Nonverbal Audiovisual Cues and Tools. In Proceedings of IEEE International Conference Affective Computing and Intelligent Interfaces, 1-9.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques. International Journal of Human-Computer Studies, Elsevier Science, 59:213–225.

Camurri, A., Mazzarino, B., & Volpe, G. (2004). Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library. In A. Camurri, G. Volpe, (eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag, 460-467.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri, G. Volpe, (eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag, 20–39.

Camurri, A., Castellano, G., Ricchetti, M., & Volpe, G. (2006). Subject interfaces: measuring bodily activation during an emotional experience of music. In S. Gibet, N. Courty, and J.F. Kamp (eds.), Gesture in Human-Computer Interaction and Simulation, volume 3881, Springer Verlag, 268–279.

Camurri, A., Coletta, P., Varni, G., & Ghisio, S. (2007). Developing multimodal interactive systems with EyesWeb XMI. In Proceedings of the 2007 Conference on New Interfaces for Musical Expres-sion (NIME07), ACM, 302–305.

Caridakis, G., Raouzaiou, A., Bevacqua, E., Mancini, M., Karpouzis, K., Malatesta, L., & Pelachaud, C. (2007). Virtual agent multimodal mimicry of humans, In: J.-C. Martin, P. Paggio, P. Kühnlein, R. Stiefelhagen and F. Pianesi (eds.), Language Resources and Evaluation, special issue of Multimodal Corpora For Modelling Human Multimodal Behavior, Springer Netherlands, pp. 367-388.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E.F. (2000). Embodied Conversational Agents, The MIT Press.

Castellano, G. (2006). Human full-body movement and gesture analysis for emotion recognition: a dynamic approach. Paper presented at HUMAINE Crosscurrents meeting, Athens.

Castellano, G., Kessous, L., & Caridakis, G. (2007). Multimodal emotion recognition from expressive faces, body gestures and speech. In F. de Rosis, (ed.), Proceedings of the Doctoral Consortium of 2nd International Conference on Affective Computing and Intelligent Interaction.

Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). The EMOTE model for effort and shape. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '2000). ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 173-182.

Drosopoulos, A., Balomenos, T., Ioannou, S., Karpouzis K., & Kollias, S. (2003). Emotionally-rich man- machine interaction based on gesture analysis. In Human-Computer Interaction International, volume 4, pages 1372–1376.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, volume 47, number 6, pp. 381–391.

Gallaher, P.E. (1992). Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology, 63(1):133–145.

Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro M., & Scherer, K. (2011). Toward a minimal representation of affective gestures, Affective Computing, IEEE Transactions on. 2(2), 106 –118.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. American Psychologist 48 (1): 26–34.

Grammer, K., Kruck, K., Juette, A., & Fink, B. (2000). Non-verbal behavior as courtship signals: the role of control and choice in selecting partners. Evolution and Human Behavior, 21(6), Elsevier, 371-390.

Gunes, H., & Piccardi, M. (2005). Fusing face and body display for bi-modal emotion recognition: Single frame analysis and multi-frame post integration. In Proceedings of the First international conference on Affective Computing and Intelligent Interaction, pp. 102-111.

Hartmann, B., Mancini, M., Buisine, S., & Pelachaud, C. (2005). Design and evaluation of expressive gesture synthesis for embodied conversational agents. In: Third International Joint Conference on Autonomous Agents & Multi-Agent Systems, Utrecht, Holland, 1095-1096.

Hsieh, C.-M., & Luciani, A. (2006). Minimal dynamic modeling for dance verbs. Journal of Visualization and Computer Animation, pp. 359-369.

Hsieh, C.-M., & Luciani, A. (2005). Generating dance verbs and assisting computer choreography. ACM Multimedia'2005, 774-782.

Hung, H., & Gatica-Perez, D. (2010). Estimating Cohesion in Small Groups using Audio-Visual Nonverbal Behavior, IEEE Transactions on Multimedia, Vol. 12, No. 6.

Jayagopi, D., Hung, H., Yeo, C., & Gatica-Perez, D. (2009). Modeling Dominance in Group Conversations using Non-verbal Activity Cues, IEEE Trans. on Audio, Speech, and Language Processing, Special Issue on Multimodal Processing for Speech-based Interactions, Vol. 17, No. 3, pp. 501-513.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14:201–211.

Kleinsmith, A., & Bianchi-Berthouze, N. (2011). Form as a Cue in the Automatic Recognition of Non-acted Affective Body Expressions, Proceedings of Affective Computing and Intelligent Interaction 2011, Lecture Notes in Computer Science, v. 6974, pp. 155-164.

Kleinsmith, A., Bianchi-Berthouze, N., & Steed, A. (2011). Automatic recognition of non-acted affective postures. IEEE Transactions on Systems, Man, and Cybernetics 41(4):1027–1038.

Kopp, S., Sowa, T., & Wachsmuth, I. (2003). Imitation games with an artificial agent: From mimicking to understanding shape-related

iconic gestures. In A. Camurri, G. Volpe, (eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag, pp. 436–447.

Kostek, B., & Szczuko, P. (2006). Rough Set-Based Application to Recognition of Emotionally-Charged Animated Character's Gestures, Transactions on Rough Sets V, Book Series Title: Lecture Notes, 146-166.

Laban, R., & Lawrence, F.C. (1947). "Effort", Macdonald & Evans, USA.

Lakens, D., & Stel, M. (2011). If They Move in Sync, They Must Feel in Sync: Movement Synchrony Leads to Attributions of Rapport and Entitativity. Social Cognition, 29(1), 1–14. Guilford Publications.

Le, Q.A., Hanoune, S., & Pelachaud, C. (2011). Design and implementation of an expressive gesture model for a humanoid robot. In: 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011), Bled, Slovenia, pp. 134 - 140.

Mancini, M, & Castellano, G. (2007). Real-time analysis and synthesis of emotional gesture expressivity, In Proceedings of the Doctoral Consortium of 2nd International Conference on Affective Computing and Intelligent Interaction.

Mazzarino, B., Peinado, M., Boulic, R., Volpe, G., & Wanderley, M.M. (2007). Improving the Believability of Virtual Characters Using Qualitative Gesture Analysis. In S. Gibet, N. Courty, and J.F. Kamp (eds.), Gesture in Human-Computer Interaction and Simulation, volume 3881, Springer Verlag, pp. 48-56.

Mazzarino, B., & Mancini, M. (2009). The Need for Impulsivity & Smoothness - Improving HCI by Qualitatively Measuring New High-Level Human Motion Features, Proceedings of the International Conference on Signal Processing and Multimedia Applications SIGMAP, Milan, Italy, pp. 62-67.

McNeill, D. (1996). Hand and Mind: What Gestures Reveal about Thought. Chicago, Illinois, USA: University Of Chicago Press.

Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions, Journal of Nonverbal Behavior. 13(4), 247–268.

M. Neff and E. Fiume. Modeling tension and relaxation for computer animation. In Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA '02), 2002

Neff, M., & Fiume, E. (2003). Aesthetic edits for character animation. In Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA '03).

Niewiadomski, R., Bevacqua, E., Le, Q.A., Obaid, M., Looser, J., & Pelachaud, C. (2011). Cross-media agent platform, In Proceedings of 2011 Web3D ACM Conference, Paris, France, pp. 11-19.

Ortony, A., Clore, G.L., & Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge: Cambridge University Press.

Paiva, A., Chaves, R., Piedade, M., Bullock, A., Andersson, G., & Höök, K. (2003). Sentoy: a tangible interface to control the emotions of a synthetic character. In AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM Press, pp. 1088–1089.

Piana, S., Mancini, M., & Camurri, A. (2012). Automated analysis of non-verbal expressive gesture, Sixth International Workshop on Human Aspects in Ambient Intelligence, International Joint Conference on Ambient Intelligence, Pisa, Italy.

Pollick, F.E. (2004). The features people use to recognize human movement style. In A. Camurri, G. Volpe, (eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag, pp. 10–19.

Pugliese, R., Lehtonen, K. (2011). A Framework for Motion Based Bodily Enaction with Virtual Characters, In: H. Vilhjálmsson, S. Kopp, S. Marsella, K. Thórisson (eds.), Proceedings of Intelligent Virtual Agents 2011, Lecture Notes in Computer Science, v. 6895, Springer Berlin / Heidelberg, pp. 162-168.

Rehm, M. (2010). Non-symbolic gestural interaction for AmI. In: H. Aghajan, R. L.-C. Delgado, J.C. Augusto (eds): Human-Centric Interfaces for Ambient Intelligence, ACM Press, 327-345.

Reidsma, D., Nijholt, A., Poppe, R., Rienks, R., & Hondorp, H. (2006). Virtual rap dancer: invitation to dance. In Conference on Human Factors in Computing Systems, Montréal, Québec, Canada. ACM Press, pp. 263–266.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., & Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In Proceedings of the 6th international conference on Human-Robot Interaction, New York, NY, USA. ACM, pp. 305–312.

Schröder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M., & Wollmer, M. (2011). Building Autonomous Sensitive Artificial Listeners. IEEE Transactions on Affective Computing volume 3, number 2, pp. 165 - 183.

Szczuko, P., Kostek, B., & Czyżewski, A. (2009). New Method for Personalization of Avatar Animation. In K. Cyran, S. Kozielski, J. Peters, U. Stanczyk, A. Wakulicz-Deja (eds.), Man-Machine Interactions, Advances in Intelligent and Soft Computing, 435- 443.

Taylor, R., Torres, & D., Boulanger, P. (2005). Using music to interact with a virtual character. In The 2005 International Conference on New Interfaces for Musical Expression, pp. 220–223.

Tsuruta, S., Choi, W., & Hachimura, K. (2010). Generation of Emotional Dance Motion for Virtual Dance Collaboration system, Digital Humanities 2010, UK, pp. 368-371.

Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., Picart, B., Tilmanne, J., & Wagner, J. (2010). AVLaughterCycle. Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation, Journal of Multimodal User Interfaces, vol. 4, n. 1, pp. 47-58.

Varni, G., Camurri, A., Coletta, P., & Volpe, G. (2009). Toward a Real-Time Automated Measure of Empathy and Dominance. CSE (4), pp. 843-848.

Wallbott, H.G. (1998). Bodily expression of emotion. European Journal of Social Psychology, 28:879– 896.

Wallbott, H.G., & Scherer, K.R. (1986). Cues and channels in emotion recognition. Journal of Personality and Social Psychology, 51(4):690–699.

Wilson, A., Bobick, A., & Cassell, J. (1996). Recovering the temporal structure of natural gesture. In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition.

Winter, D. (1990). Biomechanics and motor control of human movement. John Wiley & Sons, Inc., Toronto.