Computational Models of Expressive Behaviors for a Virtual Agent

Radosław Niewiadomski,
TSI, Telecom ParisTech

Sylwia Julia Hyniewska
Université de Genève, FPSE;
TSI, Telecom ParisTech

Catherine Pelachaud
CNRS – LTCI, Telecom ParisTech

1. Introduction

Embodied Conversation Agents, ECAs, are virtual entities with human-like communicative capabilities. They are used as a new form of human-machine interface. They often have a human-like appearance (see Figure 1) and are able to sustain a conversation with users or virtual agents (Granström et al., 2002; Cassell et al., 2000; Kopp et al., 2003; Marsella and Gratch, 2009; Gratch et al., 2007). ECAs communicate through verbal and nonverbal means such as facial expressions, hand and arm movements, body posture, and prosody. They can indicate a point in space (Martin et al., 2009), describe with their hands the shape of an object (Bergmann and Kopp, 2009), punctuate their speech with beats and eyebrow movements (Cassell 2001; Granström et al., 2002), and display emotional expressions (Niewiadomski et al., 2009a, 2009b) or social signals (Bickmore and Picard, 2005). They are able to manage speaking turn exchange (Thórisson, 2002) and even show some ecologically valid listening behavior (Bevacqua et al., 2010).

Figure 1: Examples of ECAs. (a) Sam and Alex (source J. Cassell); (b) Social cognition (source J. Gratch and S. Marsella) ; (c) Max agent (source S. Kopp).

ECAs have basically dual functions. On one hand they can serve various functions in a human-machine interface, and on the other they are tools to study human communication. Examples of the first application are web agents that serve provided information, or pedagogical agents that provide support to students in their learning process (Johnson et al., 2000; Moreno and Flowerday, 2006; Graesser et al., 2007). Lately research has been undertaken to create ECAs as companions for young kids or for the elderly (Bickmore et al., 2008; Tartaro and Cassell, 2008). ECAs are fully parameterized entities. By controlling each of their parameters, one by one or in combination, one can study their role and impact. Visual prosody has been investigated by looking at the tight temporal synchrony of voice intonation and eyebrow movement (Krahmer and Swerts, 2004; Cavé et al., 1996). The role of body posture in turn taking (Cassell et al., 1999), eye movements in deictic function (Raidt et al., 2007), importance of multimodal behavior in communicating emotional state (Martin et al., 2006) have also been examined.

Expressivity of ECAs is crucial. It gives liveliness to the virtual humanoid and allows it to communicate with others. Expressivity embeds two factors: the set of multimodal signals that is conveyed as well as their coherency in the communicative and emotional context. When building an ECA one has to create a large repertoire of multimodal signals but also to elaborate control mechanisms to determine when an expression should be displayed and which one it should be. Studies have shown that humans tend to interact with virtual humanoids in a similar way as they do with humans (Schilbach et al., 2006; Brave et al., 2005) and they have similar expectations in terms of politeness (Reeves and Nass, 1996; Rehm and André, 2005) and cultural rules (Endrass et al., 2009). Agents portraying emotional behaviors and following different emotion strategies such as empathy are perceived as more trustworthy and friendly; users enjoy interacting with them more (Brave et al., 2005; Partala and Surakka, 2004; Prendinger et al., 2005, Ochs et al., 2008). When communicating, the full body is at work. Gaze, hand gesture, facial expression, intonation, body posture are used in a coordinated manner to convey semantic, syntactic, attitudinal and emotional information (Kendon, 2004; McNeill, 1992; Ekman, 1979; Argyle and Cook, 1976; Heylen, 2006).

Regarding decoding and encoding of emotions, faces are a great vehicle of communication (Ekman, 2003b; Kaiser and Wehrle, 2001). They are able to display very subtle expressions where a slight difference in timing or muscular activity can convey different meanings. In this chapter we will pay particular attention to the modeling of expressions of emotions. We will mainly present computational models of expressions of emotions, the aim of which is to capture this palette of possibilities.

In the next section we present how the different psychology theories explain emotional expressive behaviors. Section 3 describes various computational models of static and dynamic facial expressions as well as multimodal displays of emotions. The representation of behavior expressivity is presented in the following section. Section 5 deals with the regulation of emotional expressions. We end this chapter with some concluding remarks.

2. Background – expressions of emotion

Emotional expressions are one of the most important forms of human nonverbal communication. Human expressions, even if they are usually spontaneous displays of emotional states, can also be posed, for example by actors (Gosselin et al., 1995). In this section we present different theoretical approaches to emotional expressions. In more details, in the next subsection we present main theories of facial expressions of emotions while in Section 2.2 we focus on multimodal expressions of emotions. Section 2.3 is dedicated to the models of full body expressive emotional behaviors. Finally, in Section 2.4, we discuss the socially adequate emotional displays in ECAs. We put special attention to facial behavior regulation.

2.1 Theoretical background for emotional facial behavior

In this section we present the theoretical basis of several models of expressive behaviors. In order to preserve the credibility of virtual agents' behaviors, most of the existing computational models is based on one or more theories of emotions. Even if these theories often offer only partial explanation of the humans' emotional expressive behaviors they may assure the validity of the mathematical models used in virtual agents. In this section we concentrate on theories that have been used so far in computational models of emotions.

2.1.1 Discrete representations

Discrete emotion theorists claim there is a limited number of fundamental emotions which often are called "primary" or "basic" emotions (e.g. Izard, 1971; Plutchik, 1980). Each of these prototypical emotions is characterized by a specific adaptive function, expression (e.g. specific facial behavior), physiological pattern, distinctive conscious experience (a feeling), and instrumental action (Keltner and Buswell, 1997; Manstead et al., 2005). Emotions are common to different cultures and their manifestations are universally understood (Ekman, 1999) while an expression is a characteristic and prototypical signal of one of these emotions. Therefore, establishing the list of distinct facial expressions is equivalent to establishing a list of candidates for basic emotions. The most often used list of basic emotions was proposed by Ekman and it contains six emotional states: anger, disgust, fear, joy, sadness, and surprise (Ekman and Friesen, 1975).

The small number of basic emotions is not sufficient to explain all human facial displays. In this tradition, a variety of emotional expressions commonly observed is often explained by introducing some "derivative" facial expressions. According to some theorists (e.g. Ekman and Friesen, 1975; Izard, 1992; Tomkins and McCarter, 1964) expressions can be mixed or blended in order to obtain expressions of compound emotional states like: "sadness and happiness at the same time". The different types of blends can appear in the form of superposition of emotions or masking of one emotion by another. Devillers et al. (2005) observe that these expressions occur often in everyday life situations. The resulting expression has some elements of facial expressions of both emotions. According to Ekman these facial expressions are obtained by the composition of expressions over different face areas. For instance in the case of a superposition of two emotions, the display is composed of one emotional expression for the upper face area and a different one for the lower face (Ekman and Friesen, 1975). Moreover, according to some discrete emotion theorists (e.g.: (Ekman and Friesen, 1975)), emotional expressions are affected by combinations of biological and learned factors. In particular, an expression serves to communicate an internal state, but it can be regulated according to the sociocultural factors. These factors can affect to some extent the pattern and timing of any expression. Thus, any facial expression can be for example concealed or simulated. The displays of "derivative" expressions mentioned above differ from the spontaneous expressions of basic emotions.

2.1.2 Dimensional representations

According to the dimensional emotion theorists, emotions are defined as variations on a continuum of some dimensions. Several dimensional models have been proposed. Russell's circumplex model of emotions is one of the most well known, with the dimensions of valence (pleasure-displeasure) and arousal (Russell, 1980). Regarding emotional expressions it predicts the link between each facial muscle action and its position on the activation and valence level. The information from the face is not considered sufficient however to enable the attribution of a concrete emotional label: the emotional expression would be more easily classified in terms of valence and arousal. The emotional labels are only attributed when the context is known.

Another bidimensional theory has been proposed by Schlossberg (1952), who organized facial expressions in the pleasure-displeasure and the attention-rejection dimensions. Then he ordered six expressions in this circle: happiness, surprise, fear, anger disgust, contempt, according to the observed confusions between them.

Three independent dimensions are used in the PAD model (Pleasure-Activation-Dominance) (Mehrabian, 1980, 1995), while the model proposed by Fontaine et al. (2007) uses four dimensions for the representation of emotions: evaluation-pleasantness, potency-control, activation-arousal, and unpredictability. The latter model enables to predict emotions in terms of dimensions and their behavior characteristics, e.g. a state with a high unpredictability would be characterized by raised eyebrows.

2.1.3 Componential process model representations

According to appraisal theorists, an emotional state is created by the significance given to different elements of an event (Scherer, 1999; Arnold, 1960; Lazarus, 1991). This evaluation is subjective, with respect to the well being of the individual. Thus the mental state is function of the subjective appraisal and is not driven by a preprogrammed reaction. Furthermore, appraisal theorists see emotions as dynamic episodes that lead to changes on different levels, from cognitive and motivational processes to motor and physiological responses, or action preparation (e.g. Frijda, 1986; Roseman and Smith, 2001; Scherer, 1999).

Among these theories, the Componential Process Model (CPM, see Scherer 1999) is often used because of its concrete predictions on the level of emotional behaviors that are linked theoretically to eliciting mechanisms. In the CPM each step of the subjective appraisal is linked to a facial response. Those facial movements are defined by Scherer and Ellgring (2007), researchers from the componential approach, in terms of action units (AU) which represent the position of particular muscles during an expression. The CPM states that it is the accumulation of the AU resulting from the step by step evaluation that creates the final emotional expression. Some neurophysiological studies have been realized already to confirm part of the predicted sequence of cognitive evaluations (e.g. Grandjean and Scherer, 2008).

On the level of the facial responses, the number of emotional expressions is very large, as the various elements of the facial expressions (AU) can co-occur in different patterns.

There is a great number of very differentiated emotional states and they are captured by labels only through a process of grouping of different states, through some kind of averaging and central tendencies. Scherer names these “averaged” states “modal emotions” (Scherer, 2001). The modal emotions result from cognitive evaluation outcomes that are predominant in our daily lives. What is more, these modal states are labeled in a large number of languages through short verbal expressions, mostly single words.

2.1.4 *Behavioral ecology* representation

The *behavioral ecology view* (Fridlund, 1994) dispenses with the central role of emotions in creating facial expressions. Instead, there exist only facial displays that depend on intentions of the displayer and that are specific to the context. According to Alan Fridlund, facial displays are rather means of social communication than manifestations of internal emotional states. People may show a huge variety of facial displays which are not restricted to the expressions of their emotions. On the other hand, facial expressions like a *smile* or an *angry face* can be associated with a variety of internal states. The behavioral ecology supporters give also an alternative explanation of the concept of facial regulation of emotional displays. An expression described elsewhere as a fake one (e.g., a voluntary smile) is not an imperfect simulation of a spontaneous expression, but it is a distinct social communicative act.

2.2 Multimodal expression of emotions

In the previous section we presented main theories and predictions on emotional facial displays. Recently the research on emotional displays starts being concerned with modalities different from the face. Indeed some studies show that emotions can be also related to body movements (Wallbott, 1998; Pollick, 2004). Among others, the component process model proposed by Scherer (1999) claims that an emotion is a dynamic episode that produces a sequence of response patterns on the level of gestures, voice and face (Wallbott and Scherer, 1986; Scherer and Ellgring, 2007). Moreover, some observational studies have explored the complexity of emotional expressions in terms of their dynamics and/or multimodality. Keltner (1995) studied the sequence of facial and gestural movement in embarrassment. He relied on the analysis of their appearance frequencies in audio-visual data. Shiota and colleagues studied three positive emotions: awe, amusement, pride (Shiota, et al., 2003). They showed that the three have expressions that are more than prototypical static facial expressions as described in Ekman's and Friesen's work (1975). They would rather be expressed by a set of possible signals, sometimes with asynchronous onsets, offsets and apices. Consequently, the emotional expression is not to be seen as categorical and not all elements have to be present in an expression at the same time, for such to be recognized as a display of a particular emotional state.

2.3 Emotional expression regulation

Although a direct link between an expression and an underlying emotion is often looked for, there is much evidence that the expressed emotion is not always the felt one (Fernandez-Dols and Ruiz-Belda 1995; Ekman and Friesen, 1969). Indeed, the expressions of felt emotions are often suppressed, amplified, or feigned according to the social context (La France and Hecht 2005) or personal goals (Saarni and Weber 2005). Among others, obeying to rules and social conventions often inclines people not to show their emotions in certain situations. For example, even if someone is angry with his boss, he will probably avoid expressing his emotions in front of his superior. Obeying these social rules gives a person a set of advantages of being socially acceptable.

In the Componential Process Model (Scherer, 1999), an emotion expression is driven by what Scherer calls "push and pull effects". The first one is the drive of internal changes on the exterior, motor changes, such as facial expression. The second one is the drive by social expectations and communicative codes (Scherer and Ellgring, 2007). Contrary, in behavioral ecology view any expression (either felt or fake one) is an important signal communicating different meaning. The management of facial behavior has also been discussed in the *discrete emotions theory* (e.g. Ekman and Friesen, 1969). According to the discrete emotion theorists, humans have a set of distinct expressions related to basic emotions at their disposal; however these can be altered through a secondary process that acts on the normal, spontaneous, facial displays. The management of facial behavior has been studied by Ekman and Friesen, two authors who introduced the concept of *display rules*. Display rules are a set of procedures for the management of emotional displays. They reflect knowledge about how to act appropriately. There are three necessary conditions for applying display rules: knowledge, motivation, and behavior. People must *know* which facial expression is appropriate in a specific context. They must *want* to control the spontaneous facial reactions. Finally, they must *be able* to show an adequate facial display. Thus, the application of display rules implies more or less conscious control of facial behavior.

Application of a specific display rule was studied by many researchers. They individuated a set of variables which influence the facial behavior management. Some of them are used to distinguish between cultures with different types of facial behavior (Americans vs. Japanese) (Kupperbusch et al., 2005; Matsumoto, 1990). A second group of variables explains the facial behavior of individuals in one culture. The use of some rules of facial behavior management may depend on the interlocutor (e.g. superior, child, Ekman and Friesen, 1975). We behave in a different way with our superior at work and differently with a friend. Others – so called situational display rules depend more on the situation than particular people involved. Thus, a certain type of behavior is expected during a funeral or wedding, irrespectively of who is dead or married (Saarni and Weber, 2005). Finally, personal display rules contain particular habits and effects of past experiences as well as personality traits (Ekman and Friesen, 1975).

2.4 Expressivity

Some researchers contribute to the study of emotional expressions by focusing on their *expressivity* i.e. on the external, visible qualities of a movement, like its speed, amplitude, fluidity and so on. Expressivity is an integral part of the communication process as it can provide information on the emotional state, mood and personality of a person (Wallbott, 1998).

Several works on behavior expressivity of virtual agents are based on Laban Movement Analysis (LMA, Laban and Lawrence,1974) that is an annotation scheme for describing human movement. In his theory, Laban proposed four main categories to describe a movement: body, effort, shape, and space. The first one describes structural and physical characteristics of the human body while moving. The *effort* component refers to the dynamics of the movement while the *shape* category to the way the body interacts with itself and its environment. The last category concerns movement in connection with the environment (e.g. spatial intention).

Many researchers like Wallbott and Scherer (1986), Gallaher (1992), and Pollick (2004) have investigated human motion characteristics 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 with the energy in the communication, with the relation between temporal/spatial characteristics of gestures, and/or between personality/emotion.

For Wallbott (1998) the expressivity is related to the notion of quality of the mental, emotional and/or physical state and to its quantity, somehow linked to the intensity factor of the mental/emotional/physical state. Behaviors encode not only the content information, that is, “What is communicated” through a gesture shape, but also the expressive information, that is, “How it is communicated” through the manner of execution of the gesture. Gallaher (1992) conducted a set of experiments to find the characteristics of the movement that capture the expressive style of the displayer. She gathers them into four dimensions: expressiveness, animation, expansiveness, and coordination. Moreover, Wallbott and Scherer (1986) and Pollick (2004) showed that the expressive qualities of nonverbal behaviors can be used to recognize emotional state in discrete and/or dimensional approaches.

3. Computational models of emotional expressions

Computational models of emotional expressions have received a growing interest. The models of expressive behaviors are crucial for virtual characters believability (Aylett, 2004). The existing models of emotional behavior are based on one or more theories presented in the previous section. In this section we present a short survey of existing computational models of emotional expressions for virtual agents. First we present models that rely on theory of discrete emotions. They combine elements of different facial expressions to obtain new ones. Then we will show various models that are based on the dimensional theories. In Section 3.3 we show first implementations of Scherer's componential theory. Finally, we show recent trials to generate multimodal expressions of emotions.

3.1. Models based on discrete facial expressions

The discrete emotion approaches provide concrete predictions on several emotional expressions. The idea of universality of the most common expressions of emotions was particularly sought to enable the generation of “well recognizable” facial displays. What is more, the unitary nature of the expressions was particularly attractive for its simplicity for early computational models: the different elements (e.g. action units, AU) of each expression are predicted to have a common development, with only one starting (onset) and one ending (offset) point and a common apex. However easy to categorize in terms of evoked emotions, the expressions based on discrete theory are still oversimplified. One method to enrich the emotional behavior of a virtual character, while relying on discrete facial expressions, is to introduce blends (see Section 2.1.1). In works of Bui (2004), Niewiadomski and Pelachaud (2007a, 2007b) and Mao et al. (2008) these expressions are modeled using fuzzy methods.

Bui (Bui, 2004) uses a set of fuzzy rules to determine the blending expressions of six basic emotions based on works of Ekman and Friesen (1975). A subset of rules is attributed to each pair of emotions. The fuzzy inference determines the degrees of muscle contractions of the final expression as a function of the input emotions intensities. Expressions are obtained by combining expressions from the lower and upper face of both emotions. The rules are derived from Ekman and Friesen’s work (1975) where expressions of blend of emotions are detailed. For example in the blend expression of surprise and fear, the upper face will display fear and the lower face surprise.

Niewiadomski and Pelachaud (2007a) have proposed another computational model of facial expressions based on fuzzy methods and discrete emotions theory. They use a face partition approach for the generation of various types of expressions like superposed, masked, fake and inhibited expressions. In this approach each facial expression is defined over 8 areas (brows, upper eyelids, eyes direction, lower eyelids, cheeks, nose, lips, and lips tension). Each part of the face displays an emotion. In complex expressions different emotions can be expressed on different areas of the face. The algorithm uses fuzzy rules to assign which signals of which simple expression go on the 8 different facial areas. For example, in the expression of "sadness masked by happiness", sadness is shown on the eyebrows and upper eyelids area while happiness is displayed on the lips areas. The method is extended to any simple facial expression of emotion by the application of fuzzy similarity (Niewiadomski and Pelachaud, 2007b).

In Mao et al.’s (2008) model the generation of the facial expression is realized in three layers, the first one being the physiological, the second emotional and the third social layer. The authors consider: 14 physiological variables on the physiological level (e.g. adrenaline, blood pressure or sneezing), 36 predefined emotional expressions (e.g. fear, reproach or satisfaction) and six social expressions on the social level (e.g. disagreement or winking) and the resulting facial behavior is composed of the output of each layer processed separately, while taking in count the priorities given to each layer.

In particular, the emotional layer outputs displays of complex emotional states. In more details, the complex expression results from the processing of fuzzy relation matrix between 36 predefined facial expressions and a number of emotions (i.e. the emotional state of the agent). This matrix contains many-to-many mappings between the fuzzy emotion vector and the fuzzy facial expression vector.

Each value *(e, f)* in this matrix is a degree of membership expressing the probability that an emotion *e* is mapped to the expression *f*. Given an emotional states vector as input, the output is the fuzzy facial vector that is defuzzified to obtain a concrete (complex) expression. In the algorithm, the first and last layers work in parallel. The first layer may influence the way facial behavior is realized while the last layer may facilitate or inhibit emotional expressions and/or use some social signals instead of the direct expression of an internal state. These layers define a hierarchical system. The output of each layer may be modified by the output of the layer which has a higher priority.

3.2 Dimensional based models

Models based on the dimensional approach are most often used because they allow the creation of a variety of expressions with subtle differences for related emotional states. In the models presented in this section, while the dimensional approach is crucial, they often propose a fusion with the discrete approach. For the purpose of this chapter we divided the dimensional-based models into two sets: one group uses the linear interpolation between several discrete emotions, the other uses the experimental empirical data and tries to associate some parts of facial expressions with certain areas of multidimensional space.

3.2.1 Linear interpolation

Several models combine the dimensional and the discrete approaches. They use the expressions of simple emotional states and they apply on them some arithmetical operations according to the position they occupy in some multi dimensional spaces. The model called Emotion Disc (Ruttkay et al., 2003) uses a bi-linear interpolation between two basic expressions and the neutral one. In the Emotion Disc six expressions are spread evenly around the disc according to the Schlossberg's model (1952, see Section 2.1.2), while the neutral expression is represented by the centre of it. The distance from the centre of the circle represents the intensity of expression. The spatial relations are used to establish the expression corresponding to any point of the Emotion Disc.

Two models by Tsapatsoulis and colleagues (2002) and by Albrecht and colleagues (2005) use a similar approach to compute new emotional displays. Both models use the expressions of two “neighboring” emotions to compute the facial expressions for non-basic emotions. For this purpose they use different multidimensional spaces, in which emotional labels are placed. In both approaches new expressions are constructed starting from the six Ekmanian expressions: anger, disgust, fear, happiness, sadness, and surprise. In more detail, in Tsapatsoulis et al.’s (2002) model two different approaches are used. First of all, a new expression can be derived from a basic one by “scaling” it. In the second approach a new expression is generated by looking for the spatially closest two basic emotions as defined within the dimensional spaces proposed by Whissell (1989) and Plutchik (1980). Then the parameters of these expressions are weighted with their coordinates. Albrecht et al. (2005) proposed an extended approach. The authors use a three dimensional space of emotional states defined by activation, evaluation, and power and anatomical model of the face based on FACS (Ekman et al., 2002).

The PAD model (see Section 2.2) was used in a study where participants navigated in a 3D space with corresponding facial animations using a 3D control device (Courgeon et al., 2008). In this work eight expressions (fear, admiration, anger, joy, reproach, relief, distress, satisfaction) were attributed to the extreme points of the three dimensions (valence, activation and dominance) while an interpolation of facial parameters defining an expression allowed for the generation of intermediate expressions (Courgeon et al., 2008).

3.2.2 Empirically grounded dimensional models of facial behavior

Several models of emotional behavior link separate facial actions with some emotional dimensions like valence. Interestingly most of them use the PAD model which is a three dimensional model defining emotions in terms of pleasure (P), arousal (A) and dominance (D) (Mehrabian, 1980, see Section 2.2). Among others, Zhang and colleagues proposed an approach for the synthesis of facial expressions from PAD values (Zhang et al., 2007). First, the authors proposed a new parameterization of facial expressions: Partial Expression Parameters (PEPs). Each PEP defines a facial movement in a specific area of the face. A perceptive study evaluated how their set of PEPs is linked to participants' attributions of P, A, and D values. The validity of the expressions generated from PAD values was confirmed in an evaluation study, where participants had to attribute the PAD and emotional labels to the animations (Zhang et al., 2007).

Another facial expression model based on the Russell and Mehrabian's three dimensional model was proposed by Boukricha et al., (2009). An empirical study enabled the authors to map a correspondence between randomly generated facial expressions composed of several action units as defined with FACS (Ekman et al., 2002) and ratings in term of PAD values. These PAD ratings resulted from naive participants' evaluation of bipolar adjectives using a Likert scale (Semantic Differential Measures of Emotional State or Characteristic (Trait) Emotions, as proposed by Mehrabian and Russell (1974). The evaluated expressions were placed in the dimensional space, where Dominance takes one of two discrete values (high or low dominance) while Pleasure and Activation values are mapped into a continuous space. A facial expressions control space is constructed with multivariate regressions, which enables the authors to associate a facial expression to each point in the space.

A similar method was applied previously by Grammer and Oberzaucher (2006), whose work relies only on the two dimensions of pleasure and arousal. Their model can be used for the creation of facial expressions relying on the action units defined in the FACS (Ekman et al., 2002) and situated in the two dimensional space. The authors also run a perceptive study to place randomly generated facial expressions in the dimensional space and performed a multiple multivariate regression, enabling a mapping between AUs and the two dimensions. This model was validated by checking the position of the six basic emotions in their 2D space.

Arya and colleagues (2009) propose a perceptually valid model for emotion blends. In a perceptive study human participants had to create facial expressions associated to mixed emotions on a 3D face model. For this purpose they were asked to illustrate short stories with blending expressions. A PCA analysis was performed to find the association between facial actions and emotions. Fuzzy rules are generated from the statistical analysis of the facial expressions created in the experiment by the participants. These rules associate specific facial actions with the three-dimensional space of valence, arousal and agency. Contrary to Bui whose fuzzy rules were activated depending on the intensity of emotions, in Arya et al.'s (2009) model the fuzzy values in 3D space are used to activate the agent's face.

Recently, Stoiber et al. (2009) proposed an interface for the generation of facial expressions of a virtual character. The interface allows one to generate facial expressions of the character using 2D custom control space. The underlying graphics model is based on the analysis of the deformation of a real human face. For this purpose an image database has been elaborated. It contains sequences of unlabelled images of an actor performing various facial expressions of different intensities. Facial expressions are created through the deformation of both the geometry and the texture of a facial model. The approach is based on the principal component analysis of the facial expressions database. The geometrical and textural variation of the human face is detected and automatically described in a low-dimensional space. Instead of using any existing emotional theory the authors directly analyze the physical deformations of a face. Consequently they do not rely on a predefined emotional dimensional space (see models presented in Section 3.2.1), but they build a custom 2D control space. With such an approach, it is assured that each point of the space corresponds to a realistic facial expression. Using this interface the user can create a large variety of emotional expressions of different intensities, including various mixed expressions, as well as the fluent sequences of expressions.

3.3 Beyond static expressions of emotions

As we have seen in earlier section, most research on models of emotional displays focus on facial expressions. Recent results in psychology (e.g. Keltner, 1995; Shiota et al., 2003) show however that several emotions are expressed rather by a set or a sequence of different nonverbal behaviors which are arranged in a certain interval of time rather than by a static facial expression. The expressions of emotional states are dynamic and they can be displayed over different modalities like the face, gaze and head movement (Keltner, 1995), gestures (Keltner, 1995), or even posture (Pollick, 2004; Wallbott, 1998). The multimodal expression of emotions in ECA gained recently interest. Among others, Clavel et al. (2009) studied the role of the face and posture in the recognition of ECAs' emotional expressions through two studies. One study showed that the integration of the facial and postural changes into the ECAs' emotional behavior affects users' overall perception of basic emotions, and have an impact on the attribution of the valence and activation values to the animations. A second study shows an improvement of the emotion recognition when facial and postural changes are congruent. The authors observed that the judgments were mainly based on the information sent by the face, although adding congruent postures improves the interpretation of the facial expression (Clavel et al., 2009).

Below we present some recent models of emotional displays that introduce the multimodality and sequentiality into nonverbal emotional ECAs behavior. First we focus on the works which model the dynamics of emotional display. Then we present models that enable the generation of dynamically changing multimodal expressions of emotions. Often they combine several theories, like the discrete with the componential (Niewiadomski et al., 2009a, 2009b) or the discrete with the dimensional one (Lance and Marsella, 2007).

3.3.1 Dynamic expressions of emotions based on observational studies

The dynamics of emotional expressions is modeled by Xueni Pan et al. (2007). First of all, certain sequences of signals (facial expressions and head movements) were extracted from a video-corpus. From this real data Pan et al. built a directed graph (called a motion graph) in which the arcs are the observed sequences of signals and nodes are possible transitions between them. The different paths in the graph correspond to different expressions of emotions. Thus, new animations can be generated by reordering the observed displays. Mana and Pianesi (2006) use Hidden Markov Models to model the dynamics of emotional expressions during speech acts. The system was trained on the acted samples that contain non-sense utterances and acted emotional facial expressions with different intensities. Then it was used to animate the Xface toolkit (Balci et al., 2007).

3.3.2 Models based on the componential appraisal theory

Some researchers were inspired by the Componential Process Model (Scherer, 2001), which states that different cognitive evaluations of the environment lead to specific facial behaviors. Paleari and Lisetti (2006) and Malatesta et al. (2009) focus on the temporal relations between different facial actions predicted by the Sequential Evaluation Checks (SECs) of the CPM model. Some of these (novelty, intrinsic pleasantness, conduciveness and coping potential; see Scherer, 2001) are linked to expectations in terms of particular sub-expressions that are defined in terms of facial action units (AU; Ekman, Friesen and Hager, 2002). In Paleari et al.'s (2007) work the different facial parameters are activated at different moments. The final animation that is generated on the virtual agent's face is a sequence of several sub-expressions linked to the SECs cognitive evaluations.
Also in Malatesta et al.'s (2009) work the emotional expressions are created manually from sequences predicted in Scherer's theory (2001). Differently from Paleari and Lisetti's work, each expression is derived from the addition of a new AU to the former ones. What is more, Malatesta et al. (2009) compared an additive approach (where each predicted AU contributes to a facial expression by being activated along the former AUs) with a sequential one (where each predicted AU is activated after the offset of former AUs and not along). Results show an above chance level recognition in the case of

the additive approach, whereas the sequential approach gives recognition results marginally above random choice (Malatesta et al., 2009).

3.3.3 Multimodal sequential expressions of emotions based on observational studies

In this section we present recent models that generate both sequential and multimodal expressions of emotions. Lance and Marsella (2007, 2008) propose a model of emotional gaze shifts towards an arbitrary target. In their approach they combine two animation methods: motion capture based and procedural animation. First of all, Lance and Marsella (2007) focus on the head and body movements in gaze shifts. They map these movements into the PAD dimensional model. For this purpose, a set of parameters called Gaze Warping Transformation (GWT) describing how the emotional multimodal displays differ from the neutral ones has been extracted from the recordings of acted emotional displays. The head and body movement data was captured through three motion sensors. The Gaze Warping Transformations are a combination of temporal scaling and spatial transformation parameters that are obtained from the difference between two gaze shifts directed from the same point to the same target, one being emotionally expressive and the other emotionally neutral. Finally, any emotionally expressive gaze shifts towards an arbitrarily placed target can be produced by applying GWTs to any captured neutral gaze shift.

In a second stage, a procedural model of eye movement has been added to the gaze shifts model (Lance and Marsella, 2008). This model uses several classes of gaze movements which are composed of saccade and vestibulo-ocular reflex movements. Finally, the procedural eye movement is automatically added to head and torso movement generated with GWT. Consequently an emotional state of the agent is expressed using the GWT, while the procedural model of eye movement ensures realistic motion. The model is able to generate sequences of gaze shifts expressing the same or different emotional states.

The model presented by Niewiadomski et al. (2009a, 2009b) generates emotional *expressions* that may be composed of nonverbal behaviors displayed over different modalities, of a sequence of behaviors or of expressions within one modality that change dynamically. The model is based on both video annotations and on the data reported from the literature (see Section 2.2). Videos are annotated with FACS (Ekman et al., 2002) for the face and with free labels for the other parts of the body (e.g., head nod, raising of an arm). From the analysis of the multimodal annotation emerges that these signals are not arranged freely and that some relational patterns can be established. Signals description is gathered into two sets: behavior set and constraint set. Each emotional state has its own behavior set, which contains signals that might be used by the virtual agent to display that emotion. Then, relations between the signals of one behavior set are described in a constraint set with an XML-based language. Two types of constraints are considered. *Temporal* constraints specified by arithmetic operators define the relations between the starting and ending time of a signal while *appearance* constraints describe more general relations between the signals e.g. if two signals cannot appear together, or that the appearance of one signal always coincides with the appearance of another one.

Figure 2: An example of a multimodal expression, based on the annotation of panic fear.

The algorithm takes as input a given emotion and generates a sequence of signals i.e. the animation of a given duration *t* composed of a sequence of signals on different modalities. It does so by choosing an appropriate subset of signals from the behavior set, their durations, and order of display. Each generated sequence has to satisfy all temporal and appearance constraints. The model was integrated within the Greta agent (Niewiadomski et al., 2009c). Figure 2 illustrates a sequence of signals related to the panic fear emotion. This algorithm can generate different animations, each satisfying the temporal and appearance constraints, for a given emotion. Figure 3 shows two sequences for a same emotion, embarrassment.

Figure 3: Two examples of a multimodal expression of embarrassment.

4. Behavior expressivity

The qualitative execution of a gesture and of an expression is very representative for an emotional state (Wallbott and Scherer, 1986). Behavior expressivity has been studied through perceptual tests (Wallbott, 1998), analysis models (Caridakis et al., 2006) (see Figure 4) and also in the context of dance (Laban and Lawrence, 1974).

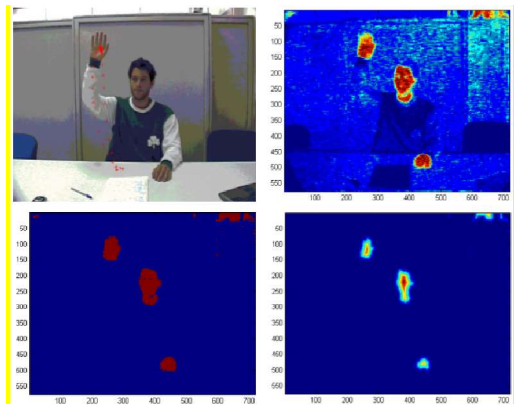

Figure 4: Analysis of behavior expressivity (source A. Raouzaiou, (Caridakkis et al., 2006)).

In ECAs' domain, Ruttkay and colleagues have proposed a representation language that encompasses behavior styles (Ruttkay and Noot, 2005). An agent is described along a large number of dimensions ranging from its cultural, professional, emotional and physical states. All these dimensions act on the way the agent behaves.

Figure 5: Variation of the spatial extent parameter (Hartmann et al., 2006).

Several agent systems are able to modulate the execution of nonverbal behaviors. In Allbeck and Badler's (2003), the choice of nonverbal behavior and the movement quality depends on the agent's personality and its emotional state. The way in which the agent performs its movements is influenced by a set of high level parameters derived from the Laban Movement Analysis (Laban and Lawrence, 1974, see also Section 2.4), 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 schema: *Effort* and S*hape.* The Effort corresponds to the dynamics of the movement and it is defined in this model by 4 parameters: space: (relation with 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 the 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 (advancing / retreating). Once it has been computed, 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 character or its emotional state. For this purpose the EMOTE parameters were mapped to the emotional states (OCC model, Ortony et al. (1990)) and personality traits (OCEAN model, Goldberg (1993)).

Neff and Fiume (2002, 2003) proposed a pose control model that takes into account several features of the nonverbal behavior such as the timing of movement, the fluent transition between different poses and its expressive qualities. In particular, for each defined body posture different properties can be defined like its tension, amplitude or extent. The model allows an 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.

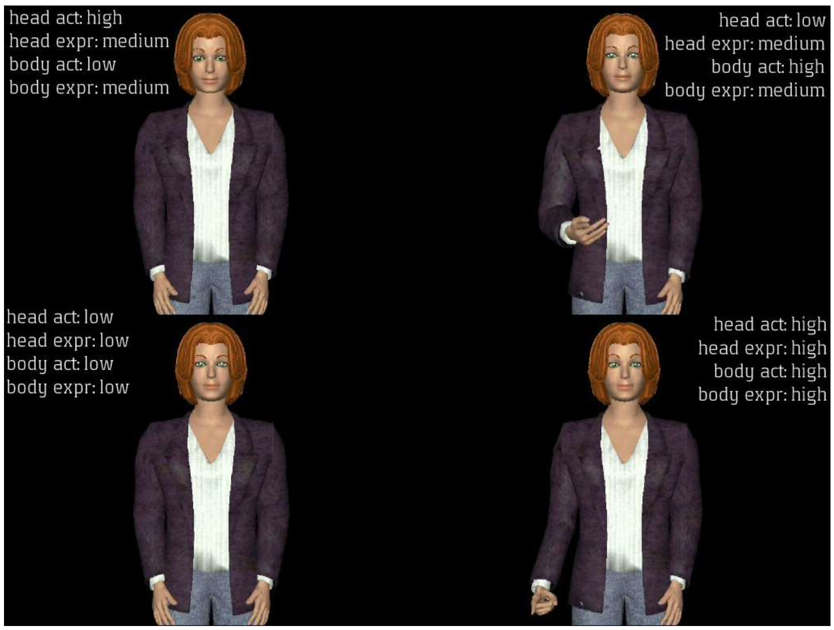

Figure 6: Examples of distinctive agents (Mancini and Pelachaud, 2008). The four agents displayed different communicative behaviors depending on their behavior expressivity.

Hartmann et al. (2006) proposed a nonverbal behavior expressivity model based on perceptual studies conducted by Wallbott and Scherer (1986), Gallaher (1992), and Wallbott (1998). These works define a large number of dimensions that characterize expressivity. Hartmann et al. have implemented six of them (see Figure 5). Three of them, 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, movement duration (linked to the execution velocity of the movement), and the dynamic properties of movement (namely the acceleration). Another dimension, fluidity, works over several behaviors of a given modality. It specifies the degree of fluidity between consecutive behaviors. The last two dimensions, global activity and repetitivity, refer to the quantity of signals and to their repetition. This model has been used to create distinctive agents, i.e. agents whose communicative behaviors can be distinguished one from the others (Mancini and Pelachaud, 2008) (see Figure 6). In their model Mancini and Pelachaud use these six parameters to define both overall and dynamic qualities of the agent nonverbal behaviors. They introduce two structures: the *Baseline* is a set of fixed parameters that represent the overall agent behavior through time, while the *Dynamicline* corresponds to the expressive qualities of the nonverbal behaviors influenced by the current communicative goals and emotional state of the agent. This approach allows one to generate animations that differ in the quality of movement for different agents as well as to modify the expressive qualities of nonverbal behavior for a particular emotional state. Thus, for example, if an agent has a general tendency (Baseline) to perform few movements with a slow speed and low amplitude then in a sad state it could simply stop moving, as sadness tends to reduce the activation parameter; on the other hand, an agent with a general tendency of moving a lot with fast and large movements, will continue to make movements even when being sad, although with a lower frequency, speed and amplitude.

5. Emotional expression regulation models

When we interact we take into account several factors to decide in a more or less conscious way which intentions and emotional states to communicate. These factors encompass the social and affective relations we have with our interlocutor, our respective roles, the power one has on the interlocutor, and so on. Growing up and living in a given society we learn not to be impulsive and to control our behavior. We are aware of the expressions that can be displayed in different circumstances and to whom. Ekman and Friesen (1969) use the term *display rules* to describe the rules that are prescribed in our social, cultural, professional and affective environments. ECAs are viewed as social entities and as such it is important to consider social context when determining which emotional display the ECA should show.

Several research teams studied the role of appropriate emotional displays on the perception of virtual agents. An inattentive application of facial expressions to virtual agents may influence negatively the user's evaluation of the agent. In Walker et al.'s study (Walker et al., 1994), people liked less the facial interface that displayed a negative expression comparing to the one which showed a neutral expression. It does not mean that the negative expressions are not desirable at all. In a card game scenario the negative expressions expressed by an agent (if situated in the context) were evaluated positively. The agent that expressed only positive expressions, irrespectively of the situation, was evaluated worse that the other one which expressed also negative emotions (Becker et al., 2005). Thus even in ECAs the emotional displays should be adequate to the context. Niewiadomski et al. (2010) evaluated the believability, warmth and competence of virtual agents showing various nonverbal emotional displays. In their evaluation study, the authors used three different versions of the agent showing (non) appropriate and/or (non) plausible emotional displays. Interestingly human subjects evaluated more warm the agent showing non plausible expressions than the one showing non appropriate (i.e. not adapted to the context) but plausible emotional displays.

There exist several models that adapt the verbal context to the situation (e.g. Walker et al., 1996; André et al., 2004). Surprisingly only few models of facial behavior regulation of an ECA have been proposed so far. Prendinger and Ishizuka modeled "social role awareness" in animated agents (Prendinger and Ishizuka, 2001). They introduced a set of procedures called "social filter programs". These procedures are a kind of rules for facial expression management. Defining social filter programs, the authors considered both social conventions (politeness) and personalities of interlocutors. The social filter program defines the intensity of an expression as the function of a social threat (power and distance), user personality (agreeableness, extroversion), and the intensity of emotion. As a result, it can either increase or decrease the intensity of facial expression, or even totally inhibit it. The evaluation study shows the influence of 'social role awareness" on the perception of "naturalness" of the agent.

The agent called Reflexive Agent (de Carolis et al., 2001) is also able to adapt its expressions of emotions according to the situational context. This agent analyses various factors in order to decide about either displaying or not its emotional state: emotional nature factors (i.e. valence, social acceptance, emotion of the addressee) and scenario factors (i.e. personality, goals, type of relationship, type of interaction). In particular the Reflexive Agent uses regulation rules that define for which values of these factors a concrete emotion can (or cannot) be displayed (de Carolis et al., 2001).

In Niewiadomski and Pelachaud (2007a) a virtual agent can inhibit an expression, mask it by another one, or show a fake expression depending on the social context. The facial expression management model is based on 2 out of the 3 variables defined in Brown and Levinson's (1987) theory of politeness, namely social and power distances. The output of the management module is a complex expression (see Section 3.1) defined as the modulation of the agent's default emotional displays depending on the valence of the agent's displayed emotion and the social distance and power between interlocutors.

Niewiadomski et al. (2008) studied the appropriate emotional displays of a virtual agent in empathic situations. They distinguish between four types of expressions: the egocentric expression of a felt emotional state, the expression of empathic emotion, and two different blends of both of them (superposition and masking). According to the evaluation results people find more adequate facial expressions that contain elements of emotion of empathy. Thus the agents should rather hide their real emotional states and show the empathic emotions.

6. Conclusion

In this chapter we have presented various computational models of expressive behaviors for virtual agents. These models are based on different theories of emotions. Some models, in particular the early models, consider mainly the expressions of the six basic emotions as predicted by the discrete emotion theory, in particular by Ekman (2003b). Models based on the dimensional approach compute facial expressions of an emotion specified by its coordinates in 2D or 3D space using a combinatorial approach. These approaches ensure the creation of a large variety of facial expressions.

Both of these types of models view the expressions of emotions as conveyed mainly by static facial expressions. Other works, based on the appraisal theory, have developed dynamic models of behaviors; that is the emotional behavior of an agent is obtained by a rapid succession of sub-expressions. The sub-expressions can be defined either from theory, and more particularly on Scherer's work (2001), or from video corpus annotation.

Few models encompass multimodality. Emotions are not conveyed solely through the face but also with the whole body. The modalities are not independent from one another. They are coordinated to convey a coherent message. Last but not least, behavior expressivity is an important factor of emotion communication and has been implemented within virtual agents.

Latest works are looking into subtle and complex research areas such as multimodality or social context. Some interesting results have been obtained. However more research needs to be undertaken, especially in modeling a large variety of emotional expressions and how social and cultural context modulate their communication. When conversing with a human user, in a particular context, for a given role and application, agents need to follow the corresponding socio-cultural rules and display emotional expressions accordingly, Agents should be endowed with the capacity to mask their emotional states, to exaggerate one, etc. Another direction of research regards the modeling of individual differences in expressions. Agents have generic behaviors. They can hardly be distinguished one from the others. It is a necessary step to develop agents differing not only by their geometry but also by their expressive behaviors.

Acknowledgments

The authors of this Chapter have been partially supported by the EU funded Social Signals Processing Network of Excellence (SSPNET) and by the French projects ANR-IMMEMO and ANR-CECIL.

7. References

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

Albrecht, I., Schröder, M., Haber, J., Seidel, H.P. (2005). Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. Special issue of Journal of Virtual Reality on "Language, Speech and Gesture", 8(4) 201-212.

André, E., Rehm, M., Minker, W., Buhler, D. (2004). Endowing spoken language dialogue systems with emotional intelligence. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P., (eds.), Affective Dialogue Systems, Springer Verlag, 178-187.

Argyle, M., Cook, M. (1976). Gaze and Mutual Gaze. Cambridge University Press.

Arnold, M. B. (1960). Emotion and personality. New York: Columbia University Press.

Arya, A., Di Paola, S., Parush, A. (2009). Perceptually Valid Facial Expressions for Character-based Applications. To appear in: International Journal of Computer Games Technology. Also available at: http://www.csit.carleton.ca/~arya/pubs.html

Aylett, R.S. (2004). Agents and affect: why embodied agents need affective systems. Invited paper, 3rd Hellenic Conference on AI, Samos, May 2004, Springer Verlag LNAI 3025, 496-504.

Becker, C., Wachsmuth, I., Prendinger, H., Ishizuka, M. (2005). Evaluating affective feed-back of the 3D agent Max in a competitive cards game. In: Tao, J., Tan, T., Picard, R.W., (eds.), Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII), Beijing, China, Springer, vol. 3784, 466-473.

Bergmann, K., Kopp, S. (2009). Increasing expressiveness for virtual agents–Autonomous generation of speech and gesture for spatial description tasks. In: Decker, K., Sichman, J., Sierra, C., Castelfranchi, C. (eds.), Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), 361-368.

Bevacqua, E., Prepin, K., Niewiadomski, R., de Sevin, E., Pelachaud, C. (2010). Greta: Towards an Interactive Conversational Virtual Companion. In: Wilks Y. (ed.), Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues, John Benjamins Publishing Company, 143-156.

Bickmore, T., Picard, R.W. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transactions on Computer Human Interaction (ToCHI), 59, 1, 21-30.

Bickmore, T., Pfeifer, L., Yin, L. (2008). The Role of Gesture in Document Explanation by Embodied Conversational Agents, International Journal of Semantic Computing, 2, 1, 47–70.

Boukricha, H., Wachsmuth, I., Hofstaetter, A., Grammer, K. (2009). Pleasure-Arousal-Dominance Driven Facial Expression Simulation. In: Proceedings of Third International Conference on Affective Computing and Intelligent Interaction (ACII 2009), Amsterdam, The Netherlands, 119-125.

Brave, S., Nass, C., Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies, 62, 161–178.

Brown, P., Levinson, S.C. (1987). Politeness: some universals on language usage, Cambridge University Press.

Bui, T. D. (2004). Creating Emotions and Facial Expressions for Embodied Agents. PhD Thesis, Center for Telematics and Information Technology, University of Twente.

Caridakis G, Raouzaiou A, Karpouzis K, Kollias, S. (2006). Synthesizing gesture expressivity based on real sequences. Workshop ‘‘Multimodal Corpora. From Multimodal Behaviour Theories to Usable Models’’. In: 5th international conference on language resources and evaluation

(LREC’2006), Genova, Italy, 27th May 2006, 19–23.

De Carolis, B., Pelachaud, C., Poggi, I., De Rosis, F. (2001). Behavior Planning for a Reflexive Agent. Proceedings of the 17th international joint conference on Artificial Intelligence (IJCAI 2001), Seattle, Washington, USA, 1059-1064.

Cassell, J., Bickmore, T.W., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H.H., Yan, H. (1999). Embodiment in conversational interfaces: Rea. In Proceedings of Proceedings of the SIGCHI conference on Human factors in computing systems (CHI’99), 520-527.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., Yan, H. (2000). Human Conversation as a System Framework: Designing Embodied Conversational Agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press, 29-63.

Cassell, J. (2001). Embodied Conversational Agents: Representation and Intelligence in User Interface. AI Magazine, 22(3): 67-83.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. In Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-1996), 2175-2178.

Chi, D., Costa, M., Zhao, L., Badler, N. (2000). The EMOTE model for effort and shape. In: Proceedings of SIGGRAPH'2000, 173-182.

Clavel, C., Plessier, J., Martin, J.-C., Ach, L., Morel, B. (2009). Combining Facial and Postural Expressions of Emotions in a Virtual Character. In: Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA 2009), Amsterdam, The Netherlands, 287-300.

Courgeon, M., Martin, J.C., Jacquemin, C. (2008). User's gestural exploration of different virtual agents' expressive profiles. In: Proceedings of the 7th international joint conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), Estoril, Portugal, 1237-1240.

Devillers, L., Abrilian, S., Martin, J.-C. (2005). Representing real life emotions in audiovisual data with non basic emotional patterns and context features, First International Conference on Affective Computing and Intelligent Interaction (ACII'2005), Beijing, China, 519-526.

Ekman, P. (1979). About brows: Emotional and conversational signals. In: von Cranach, M., Foppa, K., Lepenies, W., D. Ploog, D. (eds.), Human Ethology. Cambridge: Cambridge University Press, 169- 248.

Ekman, P. (1999). Facial Expressions. In: Dalgleish, T., Power, T. (eds.) The Handbook of Cognition and Emotion. Sussex, U.K.: John Wiley and Sons, Ltd., 301-320.

Ekman, P. (2003a). Darwin, masking, and facial expression. Ann. N.Y. Acad. Sci. 1000, 205-221.

Ekman, P. (2003b). The Face Revealed. Weidenfeld and Nicolson, London.

Ekman, P., Friesen, W.V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1, 49-98.

Ekman, P., Friesen, W. (1975). Unmasking the Face. A guide to recognizing emotions from facial clues. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Ekman, P., Friesen, W.V., Hager, J.C. (2002). Facial Action Coding System (FACS).

Endrass, B., Rehm, M, André, E. (2009). Culture-specific communication management for virtual agents. In Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 09), 281-288.

Fernandez-Dols, J.M., Ruiz-Belda, J.A. (1995). Are Smiles a Sign of Happiness? Gold Medal Winners at the Olympic Games. Journal of Personality and Social Psychology, 69 (6), 1113-1119.

Fontaine, J.R., Scherer, K.R., Roesch, E.B., Ellsworth, P. (2007). The world of emotion is not two-dimensional. Psychological Science, vol. 13, 1050-1057.

Fridlund, A. (1994). Human facial expression: An evolutionary view. Academic Press, San Diego, CA.

Frijda, N.H. (1986). The emotions Cambridge; New York: Cambridge University Press.

Gallaher, P. (1992). Individual differences in nonverbal behavior: Dimensions of style, Journal of

Personality and Social Psychology, 63, 133–145.

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

Gosselin, P., Kirouac, G., Dore, F.Y. (1995). Components and Recognition of Facial Expression in the Communication of Emotion by Actors. In: Ekman, P., Rosenberg, E.L. (eds.), What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), Oxford University Press, 243-267.

Graesser, A. Chipman, P., King, B., McDaniel, B., D'Mello, S. (2007). Emotions and Learning with AutoTutor. In: Proceeding of the 2007 conference on Artificial Intelligence in Education, 569-571.

Grammer, K., Oberzaucher, E. (2006). The Reconstruction of Facial Expressions in Embodied Systems: New Approaches to an Old Problem, ZIF Mitteilungen, 2, 14-31.

Grandjean, D., Scherer, K.R. (2008). Unpacking the cognitive architecture of emotion processes. Emotion, 8(3), 341-351.

Granström, B., House, D., Beskow, J. (2002). Speech and gestures for talking faces in conversational dialogue systems. In: Granström, B., House, D., Karlsson, I. (eds.) Multimodality in language and speech systems. Dordrecht: Kluwer, 209-241.

Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R. (2007). Creating Rapport with Virtual Agents. 7th International Conference on Intelligent Virtual Agents, Paris, France, 125-138.

Hartmann, B., Mancini, M., Pelachaud, C. (2006). Implementing Expressive Gesture Synthesis for Embodied Conversational Agents, In Proceedings of International Gesture Workshop, Lecture Notes in Computer Science vol. 3881, Springer, 188-199.

Heylen, D.K.J. (2006). Head gestures, gaze and the principles of conversational structure. International Journal of Humanoid Robotics, 3(3), 241-267.

Izard, C. E. (1977). Human Emotions. New York: Plenum.

Izard, C.E. (1992). Basic emotions, relations among emotions, and emotion-cognitive relations. Psychological Review, 99, 561-565.

Jonhson, W.L., Rickel, J.W., Lester, J. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in interactive Learning Environment. International Journal of Artificial Intelligence in Education, Vol. 11, 47-78.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. Journal of Personality and Social Psychology, 68, 441- 454.

Kaiser, S., Wehrle, T. (2001). Facial Expressions as Indicators of Appraisal Processes. In: Scherer, K.R., Schorr, A., Johnstone, T., (eds.), Appraisal Processes in Emotions: Theory, Methods, Research, Oxford University Press, New York, 285-300.

Keltner, D., Buswell, B.N. (1997). Embarrassment: Its distinct form and appeasement functions. Psychological Bulletin, 122, 250-270.

Kendon, A. (2004), Gesture: Visible Action as Utterance, Cambridge University Press.

Kopp, S., Jung, B., Leßmann, N., Wachsmuth, I. (2003). Max - A Multimodal Assistant in Virtual Reality Construction, KI-Künstliche Intelligenz, 4, 17-23.

Krahmer, E., Swerts, M. (2004). Signaling and detecting uncertainty in audiovisual speech by children and adults, In INTERSPEECH-2004, 1141-1144.

Kupperbusch, C., Matsumoto, D., Kooken, K., Loewinger, S., Uchida, H., Wilson-Cohn, C., Yrizarry, N. (2005). Cultural Influences on Nonverbal Expressions of Emotion. In: Philippot P., Feldman, R.S., Coats, E.J. (eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press, 17-44.

Laban, R., Lawrence, F. (1974). Effort: Economy in body movement. Boston, Plays Inc.

Lance, B.J., Marsella, S.C. (2007). Emotionally expressive head and body movements during gaze shifts. In: Proceedings of the 7th International Conference on Intelligent Virtual Agents (IVA), Springer, 72-85.

Lance, B.J., Marsella, S.C. (2008). A model of gaze for the purpose of emotional expression in virtual embodied agents. In Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal, 199-206.

La France, M., Hecht M.A. (2005). Option or Obligation to Smile: The Effects of Power and Gender and Facial Expression. In: Philippot, P., Feldman, R.S., Coats, E.J. (eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction), Cambridge University Press, 45-70.

Lazarus, R. (1991). Emotion and adaptation. New York: Oxford University Press

Malatesta, L., Raouzaiou, A., Karpouzis, K., Kollias, S.D. (2009). Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. Applied Intelligence. 30(1), 58-64.

Balci, K., Not, E., Zancanaro, M., Pianesi., F. (2007). Xface open source project and SMIL-agent scripting language for creating and animating embodied conversational agents. ACM Multimedia, 1013-1016.

Mancini, M., Pelachaud, C. (2008). Distinctiveness in multimodal behaviors. In: Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'08, Estoril Portugal, 159-166.

Manstead, A.S.R., Fischer, A.H., Jakobs, E.B. (2005). The Social and Emotional Functions of Facial Displays. In: Philippot, P., Feldman, R.S., Coats, E.J. (eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction), Cambridge University Press, 287-316.

Mao, X., Xue, Y., Li, Z.,Bao, H. (2008). Layered Fuzzy Facial Expression Generation: Social, Emotional and Physiological. In: Or, J. (ed.), Affective Computing, Focus on Emotion Expression, Synthesis and Recognition, I-Tech Education and Publishing, Vienna, Austria, 185-218.

Marsella, S., Gratch, J. (2009). EMA: A process model of appraisal dynamics. Cognitive Systems Research, Vol. 10, No. 1, 70-90.

Martin, J.-C., Buisine, S., Pitel, G., Bernsen, N.O. (2006). Fusion of Children's Speech and 2D Gestures when Conversing with 3D Characters. Dutoit, T., Nigay, L., Schnaider, M. (eds.), Journal Signal Processing. Special Issue on Multimodal Human-computer Interfaces, vol. 86 (12), Elsevier, 3596-3624.

Martin, J.-C., Sansonnet, J.-P., Braffort, A., Verrecchia, C. (2009). Informing the Design of Deictic Behaviors of a Web Agent with Spoken and Sign Language Video Data. 8th International Gesture Workshop (GW 2009), Feb 25-27, 2009. Bielefeld University, Germany.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. Motivation and Emotion, 14 (3), 195-214.

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. Genetic, Social, and General Psychology Monographs, vol. 121, 3, 339-361.

McNeill, D. (1992). Hand and mind - what gestures reveal about thoughts. University of Chicago Press, IL.

Mehrabian, A., Russell, J.A. (1974). An approach to environmental psychology. MIT Press, Cambridge.

Mehrabian, A. (1980). Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies. Cambridge: Oelgeschlager, Gunn & Hain.

Moreno, R., Flowerday, T. (2006). Students’ choice of animated pedagogical agents in science learning: A test of the similarity attraction hypothesis on gender and ethnicity. Contemporary Educational Psychology, 31, 186-207.

Mana, N., Pianesi, F. (2006). HMM-based synthesis of emotional facial expressions during speech in synthetic talking heads. In: Proceedings of the 8th international conference on Multimodal interfaces (ICMI 2006), 380-387.

Neff, M., Fiume, E. (2002). Modeling Tension and Relaxation for Computer Animation, ACM SIGGRAPH Symposium on Computer Animation, 81-88.

Neff, M., Fiume, E. (2003). Aesthetic Edits for Character Animation. ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 239-244.

Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C. (2009c). Greta: an interactive expressive ECA system. In: Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems, Budapest, 1399-1400.

Niewiadomski, R., Demeure, V., Pelachaud, C. (2010). Warmth, competence, believability and virtual agents. To appear in: Proceedings of the 10th International Conference on Intelligent Virtual Agents, Philadelphia, USA.

Niewiadomski, R., Hyniewska, S., Pelachaud, C. (2009a). Evaluation of Multimodal Sequential Expressions of Emotions in ECA. In: Proceedings of Third International Conference on Affective Computing and Intelligent Interaction (ACII 2009), Amsterdam, Holland, 635-641.

Niewiadomski, R., Hyniewska, S., Pelachaud, C. (2009b). Modeling emotional expressions as sequences of behaviors. In: Proceedings of the 9th International Conference on Intelligent Virtual Agents, Amsterdam, Holland, 316-322.

Niewiadomski, R., Ochs, S., Pelachaud, C. (2008). Expressions of empathy in ECAs. In: Proceedings of the 8th International Conference on Intelligent Virtual Agents, Tokyo, Japan, 37-44.

Niewiadomski, R., Pelachaud, C. (2007a). Model of Facial Expressions Management for an Embodied Conversational Agent. In: Proceedings of Second International Conference on Affective Computing and Intelligent Interaction (ACII'2007), Lisbon, Portugal, 12-23.

Niewiadomski, R., Pelachaud, C. (2007b). Fuzzy Similarity of Facial Expressions of Embodied Agents. In: Proceedings of the 7th International Conference on Intelligent Virtual Agents, Paris, France, 86-98.

Ochs, M., C. Pelachaud, D. Sadek (2008). An Empathic Virtual Dialog Agent to Improve Human-Machine Interaction, Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'08, Estoril Portugal, 89-96.

Ortony, A., Clore, G.L., Collins, A. (1998).The Cognitive Structure of Emotions. Cambridge University Press, Cambridge, UK.

Paleari, M., Lisetti, C. (2006). Psychologically grounded avatars expressions. In: First Workshop on Emotion and Computing at KI 2006, 29th Annual Conference on Artificial Intelligence, Bremen, Germany.

Paleari, M., Grizard, A., Lisetti, C.L. (2007). Adapting psychologically grounded facial emotional expressions to different anthropomorphic embodiment platforms. In: 20th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2007), May 7-9, 2007, Key West, USA.

Pan, X., Gillies, M., Sezgin, T.M., Loscos, C. (2007). Expressing complex mental states through facial expressions. In: Second International Conference on Affective Computing and Intelligent Interaction (ACII), Springer, 745-746.

Partala, T. Surakka, V. (2004). The effects of affective interventions in human-computer interaction. Interacting with Computers, 16, 295–309.

Plutchik, R. (1980). Emotions: A psychoevolutionary Synthesis, Harper and Row, New York.

Pollick, F.E. (2004). The features people use to recognize human movement style. In: Camurri, A., Volpe, G. (eds.) Gesture-Based Communication in Human-Computer Interaction - GW 2003, ser. Lecture Notes in Artificial Intelligence, NewYork: Springer-Verlag, no. 2915, 10–19.

Prendinger, H., Yingzi, J., Kazutaka, K., Ma, C. (2005). Evaluating the interaction with synthetic agents using attention and affect tracking. In Proceedings of the fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), 1099-1100.

Prendinger, H., Ishizuka, M. (2001). Social role awareness in animated agents. Proceedings of the fifth international conference on Autonomous agents, 270-277.

Raidt, S., Bailly, G., Elisei, F. (2007). Analyzing and modeling gaze during face-to-face interaction. In: Proceedings of 7th International Conference on Intelligent Virtual Agents, IVA'2007, 403-404.

Reeves, B., Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places, Cambridge University Press.

Rehm, M., André, E. (2005). Catch me if you can - Exploring Lying Agents in Social Settings, Dignum, F., Dignum, V., Koenig, S., Kraus, S., Singh, M.P., Woolridge, M. (eds.), Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Utrecht, The Netherlands, ACM, 937-944.

Roseman I.J., Smith, C.A. (2001). Appraisal Theory: Overview, Assumptions, Varieties, Controversies, Scherer, K.R., Schorr, A., Johnstone, T. (eds.), Appraisal Processes in Emotion: Theory, Methods, Research, Oxford University Press, 3-34.

Ruttkay, Z., Noot, H., ten Hagen, P. (2003). Emotion Disc and Emotion Squares: tools to explore the facial expression face, Computer Graphics Forum, 22 (1), 49-53.

Ruttkay, Z., Noot, H. (2005). Variations in Gesturing and Speech by GESTYLE, International Journal of Human Computer Studies, Special Issue on 'Subtle Expressivity for Characters and Robots', 62 (2), 211-229.

Saarni, C., Weber, H. (2005). Emotional Displays and Dissemblance in Childhood: Implications for Self-Presentation. In: Philippot, P., Feldman, R.S., Coats, E.J. (eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction), Cambridge University Press, 71-108.

Schilbach, L., Wohlschläger, A., Krämer, N., Newen, A., Shah, N., Fink, G., Vogeley, K. (2006). Being with virtual others: neural correlates of social interaction. Neuropsycholgia, vol. 44, 718-730.

Scherer, K.R. (1999). Appraisal theories. In Dalgleish, T., Power, M. (eds.) Handbook of Cognition and Emotion, Chichester: Wiley, 637–663.

Scherer, K.R. (2001). Appraisal considered as a process of multilevel sequential checking. In: Scherer, K., Schorr, A., Johnstone, T. (eds). Appraisal Processes in Emotion: Theory, Methods, Research. Oxford University Press, 92-119.

Scherer, K.R., Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? Emotion, 7, 158-171.

Schlosberg, H. (1952). The description of facial expression in terms of two dimensions, Journal of Experimental Psychology, 44 (4), 229-237.

Shiota, M. N., Campos, B., Keltner, D. (2003). The faces of positive emotion: Prototype displays of awe, amusement, and pride. Annals of the New York Academy of Sciences, vol. 1000, 296-299.

Stoiber, N., Séguier, R., Breton, G. (2009). Automatic design of a control interface for a synthetic face. Proceedings of the 13th international conference on Intelligent user interfaces, 207-216.

Tartaro, A., Cassell, J. (2008). Playing with Virtual Peers: Bootstrapping Contingent Discourse in Children with Autism. Proceedings of International Conference of the Learning Sciences (ICLS), June 24-28, Utrecht, Netherlands, 382-389.

Thórisson, K.R. (2002). Machine Perception of Multimodal Natural Dialogue. In: McKevitt, P.S., Nulláin, Ó., Mulvihill, C. (eds.), Language, Vision and Music,- Amsterdam: John Benjamins, 97-115.

Tomkins, S.S. McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. Perceptual Motor Skill, 18, 119-158.

Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Crowie, R. and Douglas-Cowie, E. (2002). Emotion Recognition and Synthesis Based on MPEG-4 FAPs. In: Pandzic I., Forchheimer R. (eds.), MPEG-4 Facial Animation. The standard, implementations, applications, John Wiley and Sons, UK, 141-167.

Walker, J., Sproull, L., Subramani, R. (1994). Using a human face in an interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, Boston, Massachusetts, 85-91.

Walker, M., Cahn, J., Whittaker, S. (1996). Linguistic style improvisation for lifelike computer characters, Proceedings of the AAAI Workshop on AI, Alife and Entertainment.

Wallbott, H. (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, 690-699.

Whissell, C. (1989). The Dictionary of Affect in Language, In: Plutchik, R., Kellerman, H., (eds.), Emotion: Theory, Research, and Experience, volume 4: The Measurement of Emotions, Academic Press, Inc., SanDiego, 113–131.

Zhang, S., Wu, Z., Meng, H. M., Cai, L. (2007). Facial Expression Synthesis Using PAD Emotional Parameters for a Chinese Expressive Avatar. In: Proceedings of Second Affective Computing and Intelligent Interaction Conference (ACII 2007), Lecture Notes in Computer Science, Springer, 24-35.