Affect expression in ECAs: application to politeness displays

Radosław Niewiadomski[1] and Catherine Pelachaud[2]

Telecom ParisTech, Paris, France[1]

LTCI-CNRS, Paris, France[2]

Abstract

In this paper we present our Embodied Conversational Agent (ECA) capable of displaying a vast set of facial expressions to communicate its emotional states as well as its social relations. Our agent is able to superpose and mask its emotional states as well as fake or inhibit them. We defined *complex facial expressions* as expressions arising from these displays. In the following, we describe a model based on fuzzy methods that enables to generate complex facial expressions of emotions. It uses fuzzy similarity to compute the degree of resemblance between facial expressions of the ECA. We also present an algorithm that adapts the facial behaviour of the agent depending on its social relationship with the interactants. This last algorithm is based on the theory of politeness by Brown and Levinson (1987). It outputs complex facial expressions that are socially adequate.

Key words: embodied conversational agent, facial expressions, nonverbal politeness, fuzzy similarity, social context

1 Introduction

Embodied conversational agents (ECA) are 3D virtual entities with a human-like appearance able to communicate with the human users or with other ECAs. Similarly to humans, ECAs use various forms of verbal and nonverbal signals like speech, gestures, gaze, and in particular facial expressions. Early evaluation studies have shown the advantages of agents displaying facial expressions. They are more engaging (Walker et al. (1994)), and the comprehension of message is improved (Elliott (1997)). Despite these results ECAs usu-

Email addresses: `radoslaw.niewiadomski@telecom-paristech.fr`[1], `catherine.pelachaud@telecom-paristech.fr`[2].

ally use only a small subset of human facial repertoire mainly the neutral expression and the six facial expressions of the so-called "basic" emotions: anger, disgust, fear, happiness, sadness, and surprise (Ekman and Friesen (1975)).

Human facial expressions may serve several functions. They may signal emotional states, but also communicate intentions and attitudes (Ekman and Friesen (1975); Poggi (2005)). They are influenced by many social and cultural factors. A series of experiments has shown that people modify their spontaneous facial expressions depending on interpersonal relations (e.g. Buck et al. (1992); Ekman and Friesen (1969); France and Hecht (2005); Manstead et al. (2005); Wagner and Lee (2005); Wagner and Smith (1991)). Different types of display strategies like showing fake, inhibited, or masked expressions are used. The ability to control and to know when to control emotional expressions (i.e. suppress, substitute, or simulate expressions of emotions) is one of the skills often referred to as *emotional intelligence* (Goleman (1995)). By analogy to human beings, we expect that embodied conversational agents can also benefit from emotional intelligence. Emotionally effective and socially competent agents are more likely to build a successful relationship with a human user. According to Reeves and Nass (1996) people have some implicit expectations about the social and emotional behaviour of the new media. Computers need to respect social rules and, in particular, rules of interpersonal relations. The violation of social norms by the computer is viewed as social incompetence and is perceived as offensive (Reeves and Nass (1996); Pentland (2004); Vinciarelli et al. (2009)).

Before presenting our algorithms we introduce some terms related to this work. In this paper, we distinguish between *simple* and *complex* facial expressions. By *simple* facial expressions we intend spontaneous facial displays of emotional states (which can be described by one-word label) e.g. display of anger or contempt. We use the term *complex facial expressions* to describe expressions that arise from the combination of several simple facial displays (e.g. superposition of two expressions) or which are voluntarily modified by the displayer (e.g. inhibited or masked expressions). Different types of facial expressions like masked (e.g. anger masked by happiness), superposed (e.g. anger and fear displayed contemporarily), fake (e.g. simulated surprise), or inhibited (e.g. suppressed disappointment) expressions are complex facial expressions resulting from the combination or modulation of simple expressions of emotion like "the expression of anger" or "the expression of disappointment".

In this paper our focus is on the creation of emotional facial expressions and how their displays are modulated during an interaction. It is just one aspect of emotional intelligence. Our aim is to endow ECAs with the ability to display different types of complex facial expressions with particular attention paid to their various meanings and communicative roles. However in this paper we do not discuss the issue whether ECAs are able to have felt emotions or are solely able to display them or even to refer to an emotional state sometimes called emotional emblems (Ekman and Friesen (1975)). We focus on perceivable facial features of emotions and how they can be displayed by ECAs. Then, in the last section of this paper, we present an application of our complex facial expression algorithm. We introduce a virtual agent that uses a variety of facial expressions to manifest its relationship with a user (or another agent). We focused on situations in which different rules of facial behaviour management are applied; we studied how these situations influence the spontaneous displays of emotions. Consequently, our agent uses facial expressions not only to express its emotional states but also to manifest its social relations.

The remaining part of this paper is structured as follows. In the next section we present the psychological foundations of our work, while section 3 is dedicated to an overview of expression synthesis algorithms and of agents able to adapt themselves to social context. Section 4 presents the main concepts of our system. In sections 5 and 6, the model of complex facial expressions is described, while in section 7 we present a model of facial behaviour management. Finally we discuss the limitations of this work in section 8 and we conclude the paper in section 9.

2 Background

Darwin (1872) postulated that facial expressions are primarily the answer of evolution to potential danger. Even if we cannot neglect our biological origins, nowadays, the human facial expressions are mostly an instrument of communication. Like many other forms of interaction, they are regulated by some informal rules (France and Hecht (2005)) and personal goals (Saarni and Weber (2005)). In this section we discuss the role of facial displays in context-dependent situations.

2.1 The meaning of facial expressions

The explanation and description of the phenomenon of emotional displays in social context have not been agreed upon until now. In the *emotional expression view* (e.g. Ekman and Friesen (1975); Matsumoto (1990)) facial expressions are affected by combinations of biological and learned factors. An expression serves to communicate internal emotional state, but it can be regulated according to some socio-cultural factors. These factors can affect to some extent the pattern and timing of expression. Any facial expression can be concealed or simulated and this voluntary modification can be perceived by an observer. On the other hand the behavioural ecology view (e.g. Fridlund (1994)) dispenses with the central role of emotions in creating facial expressions. Facial displays depend on the intentions of the displayer and are specific to the (social) context. They are rather means of social communication. Therefore, people smile when they want to express readiness to play and not when they are happy (see Manstead et al. (2005) for detailed discussion). In this theory a fake expression (e.g., voluntary smile) is not imperfect or flawed simulation of spontaneous expression but it is a distinct social communicative act. Another explanation of facial expressions can be found in componential appraisal theories (e.g. Scherer (2001); Wehrle et al. (2000)) which share the assumption that single components of facial expressions like raising the eyebrows or opening the mouth have an independent meaning. These facial signals indicate the specific appraisal results. Thus facial expressions are the evidence of appraisal processes that can be influenced by social situation (Wehrle et al. (2000)).

We mainly based our works on the emotional expression view according to which management of facial behaviour is a secondary process that alters the displays of inner states. In our research we avoid considering the cognitive level (i.e. the cognition process that leads to an emotional state) but we generate the perceivable facial expressions to be displayed by our agent. Facial expressions in our approach are not generated from appraisals but are the results of composition of expressions. Interestingly, the concept of composing facial expressions from smaller elements occurs in Ekman's explanation of complex expressions (see section 2.2 for details) as well as in Scherer's componential theory. It is also supported by results of perceptive experiments about perception of partial facial expressions (Bassili (1979); Constantini et al. (2005)). According to these experiments positive emotions are mainly perceived from the expression of the lower face (e.g., smile), while negative emotions are perceived from the upper face (e.g., frown).

2.2 Complex facial expressions

According to supporters of *emotional expression view* facial expressions are a good indicator of emotional states (Ekman and Friesen (1975); Izard (1977); Wagner et al. (1986)). They do not always correspond to felt emotions, they can also be fake (showing an expression of an unfelt emotion), masked (masking a felt emotion by an unfelt emotion), superposed (showing mixed felt emotions), inhibited (masking the expression of emotion with the neutral expression), suppressed (de-intensifying the expression of an emotion), or exaggerated (intensifying the expression of an emotion) (see Niewiadomski (2007) for detailed discussion).

Facial areas composition. According to Paul Ekman complex facial expressions are obtained by the composition of expressions over different facial areas. For instance in the case of superposition of two emotions the final expression is composed of the upper facial area of one expression and the lower facial area of another one (Ekman and Friesen (1975)). The boundary between the upper and the lower face is not precisely defined: for certain pairs of emotions (e.g., anger and sadness) the eyes are included in the upper facial area, while for other pairs they are not (Ekman and Friesen (1975)). Ekman and Friesen described eighteen different expressions of superposition for pairs involving six emotions (Ekman and Friesen (1975); Ekman (2003b)). Not every possible combination of the upper and the lower faces is plausible e.g. sadness in the superposition with happiness is expressed by the upper face region, and happiness by the lower face. The opposite case does not occur.

Reliable features of emotional expressions. Humans also distinguish the expression of felt emotion from the expression of fake emotion (Ekman and Friesen (1969); Frank et al. (1995); Gosselin et al. (1995)). A list of *deception clues* i.e. the features of expressions that are useful in distinguishing between fake and felt expressions have been proposed (Ekman and Friesen (1975); Ekman (1985, 2003a)). Humans are not able to voluntarily control all their facial muscles. Expressions of felt emotions may be associated with specific facial features like raised brows in the case of sadness (Ekman and Friesen (1975)) or some *orbicularis oculi* activity in the case of happiness (Ekman (2003b)). Such *reliable features* lack in fake expressions as they are difficult to do voluntarily. On the other hand, people are not able to fully inhibit felt emotions. According to the *inhibition hypothesis* (Ekman (2003b)), the same elements of facial expressions, which are difficult to show voluntarily in the case of unfelt emotions, are also difficult to inhibit in the case of felt emotions. Ekman and Friesen (1975) enumerated reliable features that leak over for unfelt emotions. Felt and fake expressions can also be distinguished by their variation of symmetry, synchronisation, and timing (Ekman and Friesen (1975); Ekman (1985)). Fake expressions are more often asymmetric (Ekman (2003a)), more abrupt (Frank et al. (1995); Ekman and Friesen (1982)) and longer than felt ones (Ekman (2003a)).

Complex expressions in natural setting. In a study that analysed and annotated spontaneous human behaviour Devillers et al. (2005) observed different types of expressions like *blended*, *masked* and *sequential expressions*. These complex expressions were found to be displayed nearby as much as expressions of simple emotions (33% vs. 46% of all occurrences) (Devillers et al. (2005)).

2.3 Context-dependent facial displays

The management of facial behaviour has been studied intensively. Ekman and Friesen (1975) have introduced the concept of *display rules* - they noticed that the representatives of different cultures show different facial expressions in the same context independently of their emotional states. Display rules reflect knowledge about how to act appropriately. There are three factors underlying the use of display rules: knowledge, motivation, and behaviour. People must know which facial expression is appropriate in a specific context. They must want to control their spontaneous facial reactions. Finally, they must be able to show an adequate facial display. Thus, the application of display rules implies the more or less conscious control of facial behaviour. The rules of facial displays management have at least two different origins. First, display rules refer to the conventions about followed by members of a particular group, class, or subculture. The second type called personal display rules contains particular habits and effects of past experiences as well as personality traits. The rules of facial behaviour management can depend on situations (e.g. weddings, or funerals), cultures (e.g. Japan vs. American), or certain features of the interlocutor (e.g. superior, child) (Ekman and Friesen (1975)). In this work we focus on the rules of facial behaviour management that are used to adapt the expressions according to the interpersonal relations.

3 State of art

Although different types of complex facial expressions occur often in real-life (see section 2), they are rarely considered in virtual agents. In section 3.1 we present existing models of facial behaviour while in section 3.2, we describe how ECAs adapt their facial expressions to interpersonal relations. We end this section by motivating our goal.

3.1 Models of facial expressions

Several models of facial expressions have been proposed to enrich the agent's facial behaviour. The existing solutions usually compute new expressions "averaging" the values of the parameters of the expressions of "basic" emotions (Ekman and Friesen (1975); Ekman (2003b)). 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, while the neutral expression is represented by the centre of the circle. 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*.

Models of Tsapatsoulis et al. (2002) and Albrecht et al. (2005) use the expressions of two “neighbouring” 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 Ekman’s expressions: anger, disgust, fear, happiness, sadness, and surprise. More precisely, in Tsapatsoulis et al. (2002) a new expression can be derived either from a basic one by “scaling” it or by looking for the two spatially closest basic emotions in the multi-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.

Bui (2004) uses a set of fuzzy rules to determine the blending expressions of six basic emotions based on findings by Ekman and Friesen (1975). A subset of rules is attributed to each pair of emotions. The fuzzy inference determines the degrees of muscles contractions of the final expression as a function of the input emotions intensities. Finally, different types of facial expressions were considered by Rehm and André (2005). In a study on deceptive agents, they showed that users were able to differentiate between the agent displaying an expression of felt emotion and an expression of fake emotion (Rehm and André (2005)). For this purpose they manually defined facial expressions according to Ekman’s description of fake expressions. These expressions are more asymmetric and they miss reliable features.

In this section we described models that generate various facial expressions. In the next section we will present how agents apply them in different social contexts.

3.2 Social context in embodied agents

Prendinger and Ishizuka (2001) modelled “social role awareness” in animated agents. Their *social filter programs* are rules for facial expression management. To define them Prendinger and Ishizuka considered both social conventions (politeness) and personality of interlocutors. The social filter program defines the intensity of an expression as the function of social threat (power and distance), user’s personality (agreeableness, extroversion), and the intensity of emotion. As a result, it can either increase or decrease the intensity of a facial expression, or even totally inhibit it. De Carolis et al. (2001) built a reflexive agent able to adapt its expressions of emotions according to the situational context. Emotional displays of the agent depend on emotional factors (i.e. valence, social acceptance, emotion of the addressee) and scenario factors (i.e. personality, goals, type of relationship, and type of interaction). The agent uses *regulation rules* that define for which values of these parameters the “felt” emotion can (or cannot) be displayed. Despite the large set of parameters considered in this model, it uses only one type of complex facial expressions, namely inhibition (De Carolis et al. (2001)).

3.3 A new model of facial expressions in interpersonal relations

In our model we introduce the diversification of facial expressions in relation to their meaning, role, and appearance. The novelty of our system is that our agent is able to express different types of facial expressions like inhibited, masked or fake expressions. Complex facial expressions are computed by composing areas of facial expressions; that is the final expression is a combination of facial areas of input expressions. Concerning the inclusion of the social context in ECAs systems the solutions presented in the previous section do not use complex facial expressions. They do not exploit fully the communicative functions of facial expressions. Moreover, they may generate displays that are inadequate in a given social context. We build an agent that modifies its expressions of emotions depending on its relation with its interlocutors. It is expected that these changes will be perceivable and interpretable by human interlocutors.

4 Overview of our model

In this section we present the model of an ECA capable of modifying its facial behaviour in order to adapt it according to its relation with its interactant. Our model follows a 3-step process (as shown in Figure 1) to model an ECA that is capable of displaying socially adequate expressions. In a situation S the agent (step 1) **in a given emotional state $\boldsymbol{E}$** (step 2) **must know** whether it can display its felt emotion E; if not, it must decide which facial expression is appropriate in S and (step 3) it **must be able to display** the adequate (complex) facial expression resulting from the previous steps.

We implement the 3-step process as follow:

(1) We use an agent called Greta (Bevacqua et al. (2007)). The emotional state of our agent has to be defined explicitly in an input file either manually or computationally (Ochs et al. (2005)),

Fig. 1. Three stages allowing an ECA to manage its emotional expressions.

(2) Our agent uses a set of rules to compute which management type to use to modulate its default ("spontaneous") facial expressions (see section 7).
(3) A model of complex facial expressions allows our agent to display the appropriate facial display (see section 6).

In step 3, our model of complex facial expressions generates different displays for fake, masked, and inhibited expressions. Complex facial expressions for the six basic emotions: anger, disgust, fear, happiness, sadness, and surprise are described in the literature (Ekman and Friesen (1975); Ekman (2003b)) (see section 2). From this literature, for each type of expressions, we have extracted a set of rules that describes its characteristic features. For any input emotion label for which the complex expression is not explicitly defined by these rules (e.g. expressions of contempt or disappointment) the fuzzy similarity based algorithm (see the next section for details) is used to establish the degree of similarity between the expression of the input emotion label and the expressions whose complex facial expressions are described by our rules. Once the most similar expression is known, the corresponding rules are applied to the expression of input emotional state (see sections 6.1.2 and 6.2.3 for details). Then a model of facial expression management, described in section 7, is used to adapt the default facial expressions of the agent to interpersonal relation. In the current stage of our model, we are focusing on two factors characterising the relationship between interactants: their degree of familiarity and their power of relationship. In the remaining of this paper when mentioning either 'situation S' or 'social context', we refer to these two factors characterising relationships.

5 Similarity of facial expressions

For comparing computer generated facial expressions we use *fuzzy similarity* (Bouchon-Meunier et al. (1996)) which is used to describe objects characterized by loose description. In this approach each object or feature that does not have a precise definition can be described by a fuzzy set. Fuzzy similarity allows for the comparison of two fuzzy sets. The attributes of an expression are defined with fuzzy sets instead of using precise values. On the other hand, according to many researchers (e.g. Ekman and Friesen (1975); Izard (1977)) each "distinct and labelled expression of emotion" like "expression of anger" or “expression of contempt” is rather a class or a set of different but similar configurations of facial muscle actions (or a set of different *facial displays*). Indeed, there is not one precise smile or frown. Each smile is somehow unique but all smiles have some characteristics in common. The boundary between “smiling” and “not smiling” is imprecise. Different facial displays of different intensities are classified as smiles. In many experiments (e.g. Bartneck and Reichenbach (2005); Etcoff and Magee (1992)) different facial displays involving the same group of muscle contractions were described by subjects with the same label e.g. “expression of anger”. It has an imprecise “fuzzy” definition (see also Tsapatsoulis et al. (2002)). On the other hand, all facial displays that belong to one category like “happiness”, “anger”, or “embarrassment” have some common features (e.g. “smile” or “frown”). Therefore, any emotional category can be defined by a set of fuzzy sets that corresponds to these features. Thus we can compare facial expressions by comparing fuzzy sets that describe them.

5.1 Fuzzy similarity

Fuzzy similarity offers a set of methods to compare two objects. Each feature of an object is represented by a fuzzy set. Two fuzzy sets can be compared using *M-measure of comparison* (Bouchon-Meunier et al. (1996)). It expresses the strength of the relationship between the features of two objects. There are different types of the M-measures of comparison (see Bouchon-Meunier et al. (1996); Rifqi (1996) for details). For our application we chose the M-measure of resemblance (Bouchon-Meunier et al. (1996)). It is used for comparing objects of the same level of generality. Using this M-measure it is possible to check whether two objects “have many characteristics in common” (Bouchon-Meunier et al. (1996)). For our application we chose the measure of resemblance S defined by:

$$S(A,B) = \frac{(M(A \cap B))}{(M(A \cup B))} \qquad (1)$$

where A and B are two fuzzy sets (μ_A is the membership function of A) and M is the fuzzy measure on Ω:

$$M(A) = \int_{\Omega} \mu_A(x)dx \qquad (2)$$

As a result we obtain the value of comparison $x_i \in$ [0,1] for each pair of attributes. Following the approach proposed by Rifqi (1996) we use *Ordered Weighted Averaging* (OWA) operator to aggregate all values $x_1, ..., x_n$. The

OWA, $h_W : [0,1]^n \rightarrow [0,1]$, is defined as:

$$h_W = \sum_{i=1}^{n} w_i b_i \qquad (3)$$

where b_i is the i-th biggest value between $x_1, ..., x_n$ and $W = \{w_1, ..., w_n\}$ is a set of weights with $w_i \in [0,1]$ and such that $\sum_{i=1}^{n} w_i = 1$ (Rifqi (1996)). Finally, we use trapezoid fuzzy sets to describe the features of facial expressions as shown in Figure 2. This shape corresponds to the experimental results about the perception of facial expressions (Bartneck and Reichenbach (2005); Young et al. (1997)).

Each M-measure of resemblance S also has two other properties:

- reflexivity: $S(A, A) = 1$,
- symmetry: $S(A, B) = S(B, A)$.

Niewiadomski (2007) found that the perception of similarity between unlabelled facial expressions of an agent is symmetrical, i.e. expression A is similar to expression B to the same degree as B is similar to A. Obviously, similarity between unlabelled facial expressions is also reflexive.

5.2 Algorithm

Facial expressions of Greta agent are described in terms of facial animation parameters (FAPs) according to the MPEG-4 standard (Ostermann (2002)). Originally, facial expressions used by the agent were defined using precise values. They needed to be fuzzified. For each FAP of each expression we have defined the fuzzy set of plausible values (see Niewiadomski and Pelachaud (2007) for details). In our algorithm a membership function of a fuzzy set is represented by a symmetrical trapezoid with its centre in the point v, where v is the value of the original expression (see Figure 2).

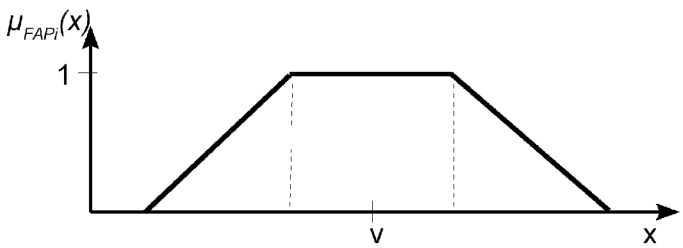


Fig. 2. A fuzzy set of FAP_i.

Using fuzzy definitions of facial expressions we calculate the value of similarity between them. Our algorithm works as follow: let E_i and E_j be two emotions whose expressions we want to compare. Thus we want to compute the fuzzy similarity $FS(Exp(E_i), Exp(E_j))$ between two expressions of the agent: $Exp(E_i)$ and $Exp(E_j)$ that are defined in terms of FAPs. Each $Exp(E)$ is associated with a number of fuzzy sets such that all plausible *facial displays* (in the sense of muscle contractions) for the emotion E are defined. That is, for each parameter k of an expression of E there is a fuzzy set FAP_k that specifies its range of plausible values. Then the value of fuzzy similarity for each parameter of $Exp(E_i)$ and $Exp(E_j)$ is established. The M-measure of resemblance S is used to find these similarity values. For each FAP_k of $Exp(E_i)$ and $Exp(E_j)$ we have:

$$fs_k = \frac{M(FAP_k(E_i) \cap FAP_k(E_j))}{M(FAP_k(E_i) \cup FAP_k(E_j))} \tag{4}$$

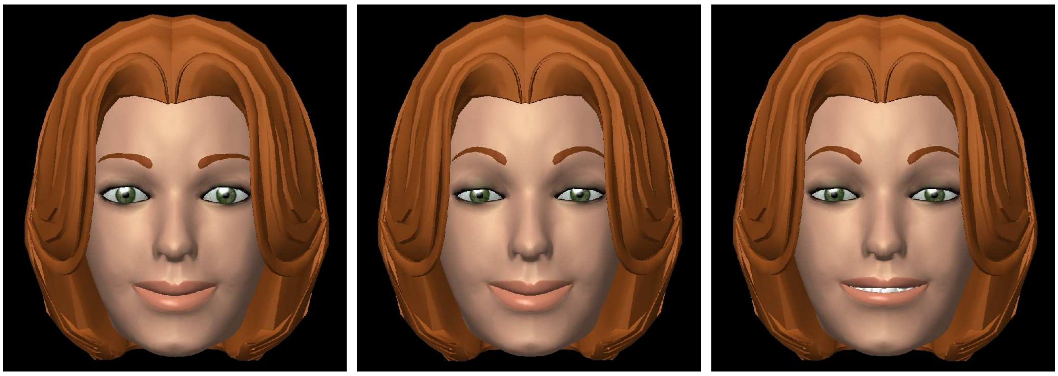
where $k = 1,\ldots,n$ and M is defined by equation (2). Finally, in the third step, all values are combined by means of the aggregation operator h_w (3):

$$FS(Exp(E_i), Exp(E_j)) = h_w(fs_1, ..., fs_n) \tag{5}$$

where h_w is OWA operator with the weights $w_k = \frac{1}{n}$ (see section 5.1).

5.3 Example

The value of similarity between completely different expressions such as the one of sadness and one of happiness is 0. Much more interesting is the case of facial expressions that are similar. Let us compare the three slightly different facial expressions shown in Figure 3.



A B C

Fig. 3. Example of similar facial expressions: B is more similar to C than to A.

In Figure 3c, the lips are extended with greater intensity than in Figure 3b. When comparing Figure 3a and Figure 3b, the eye aperture in Figure 3b is more closed than in Figure 3a. Moreover, in these two images, the eyebrows have different shapes. Thus, Figure 3a differs from 3b in more details than 3c from 3b. Because of that it can be expected that A is less similar to B than B to C. Indeed, the values of similarity, outputted by our algorithm are: $FS(A,B) = 0.37$ and $FS(B,C) = 0.45$. That is, the expression C is more similar to B than A is to B.

Figure 4 presents a set of facial expressions ordered according to the gradual increase of fuzzy similarity between each of them and the reference object which is placed on the right (the values of fuzzy similarity decrease from right to left).

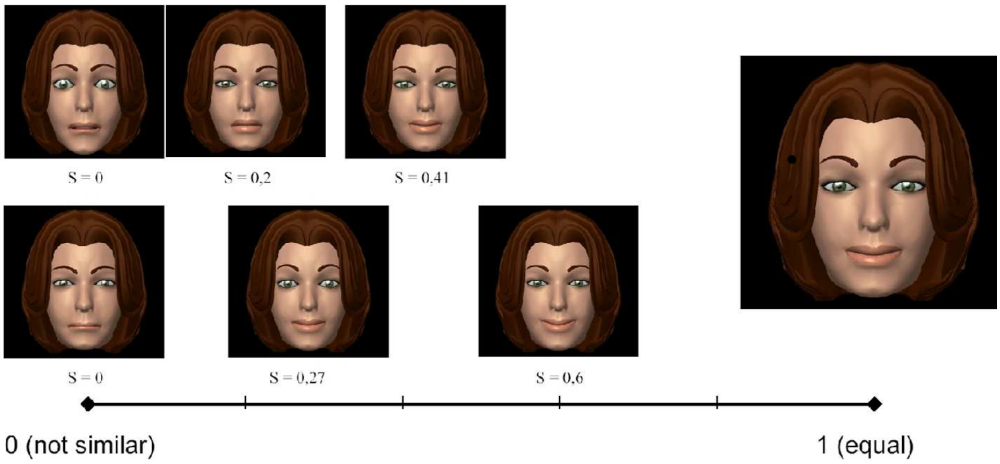


Fig. 4. The gradual increase of fuzzy similarity of facial expressions with the reference image on the far right.

5.4 Evaluation

We have conducted an evaluation study to verify if the values of the similarity established by our algorithm are consistent with human perception. In this experiment we focused on the process of comparison of any two facial expressions (i.e. the perception of the common features and the differences between them). Our hypothesis was that values of fuzzy similarity would be aligned with those found by human participants. In particular, we expected to find that our algorithm and human perception are concordant not only in evaluating whether any two expressions are similar to each other or not, but also that different degrees of resemblance perceived are adequately modelled in our algorithm.

5.4.1 Objects of Comparison

Our objects of comparison are images of facial expressions of the Greta agent. Each image depicting facial expressions follows the same setting: each image presents one facial expression, only the face is visible in the image, the face is directed towards the observer, a black background is used. Each jpeg file is a colour image of Greta's face of the size 240x219 pixels. Some examples of the images used in the experiment are presented in Figure 4.

We were interested in measuring the similarity between facial expressions at their apex. We were not concerned, here, with similarity between emotions. The images did not have the quantitative and qualitative description in the terms of emotional labels or emotional intensity. To capture the concept of similarity of synthetic facial displays we used 22 images presenting different eyebrows and lips shapes generated using MPEG-4 standard (Ostermann (2002)). Certain images were obtained only by scaling the values of animation parameters of other expressions in the set; others by using different facial shapes. More precisely the following images were used [1] :

- six images presenting various smiles; three of them where accompanied by some eyebrows movements,
- three images of lips pressed and three others where lips pressing is accompanied by various frowns,
- four images of lowered lips corners, two of them with eyebrows raising,
- two other images were composed of eyebrows and upper lip raised,
- an image of widely open mouth and raised eyebrows,
- an image of raised eyebrows and mouth tensed and opened,
- an image of slightly raised eyebrows,
- the neutral expression.

5.4.2 Procedure

We ascribed the 22 images to ten sets. Each set s_l, $l = 1,...,10$, is composed of one *reference expression* and six facial expressions that have to be *compared* with the reference one. It means that each experiment session consists of 60 operations of comparison (i.e. ten sets of six comparison pairs each). To have access to a greater number of participants, we set up our experiment on the web. The users had access to the evaluation study through a web browser. The participation to the experiment was anonymous. The sets of images s_l as well as the images in the set were displayed in a random order. For each pair of images (i.e. reference object, compared object) subjects had to choose the

[1] The set of images used in the experiment can be found at (27 May 2010): http://perso.telecom-paristech.fr/~niewiado/evaluations/similarity_images/index.html

degree of similarity by using five-point Likert scale, ranging from "1 - Not at all" to "5 - Equal".

5.4.3 Results

Forty six persons participated in the experiment. 23 participants were women, the other 18 - men. The remaining 5 persons did not specify their gender. The total number of answers was 2760. Participants were from different countries mainly from Italy (46 %), Poland (26 %), UK (4 %) and France (4 %). They were between 20 - 40 years old and none of them had worked with a synthetic face before.

First of all, we found that labels were used by subjects with different frequency. The first label: "1 - Not at all" that corresponds to the lowest degree of similarity occurred in nearly half of all answers (46%). Other labels occurred from 10% to 16% of all responses.

In order to interpret subjects' answers we compared them with the values returned by our algorithm. For this purpose we changed the Likert values into continuous scale. Then, we compared them with the values of fuzzy similarity. More formally, for the purpose of measuring the answers of participants we introduced the *average similarity index*. Let *(A, B)* be a pair of expressions in which A is the reference and B is the compared object. We call u_i the number of answers using i-th label, i.e. u_1 corresponds to the label "1 – Not at all" and u_5 to the "5 – Equal". The average similarity index, asi_{AB}, is :

$$asi_{AB} = \frac{\sum_{i}^{5}(w_i u_i) - w_1 \sum_{i}^{5} u_i}{(w_5 - w_1)\sum_{i}^{5} u_i} \tag{6}$$

where $w_i = i$ is the weight that corresponds to u_i (we assumed that labels are laid out evenly along the interval of possible values). Let us notice that the values of asi_{AB} and the values of fuzzy similarity FS are in the interval [0,1]. Let the vector $[a_i]$ contains the values of our fuzzy similarity FS such that: $a_i = FS(A_i, B_i)$ and let the vector $[b_i]$ be such that: $b_i = asi_{A_iB_i}$. The correlation (r) between $[a_i]$ and $[b_i]$, *i=1,..,60*, is 0.89. The average similarity index for 80% of the considered pairs is different from the perfect value (represented by the main diagonal) by 0.2 at most.

On the other hand, certain pairs were evaluated significantly higher by the participants of the experiment than by our algorithm. For this reason we also measured the discrepancy between values b_i and a_i. The mean difference

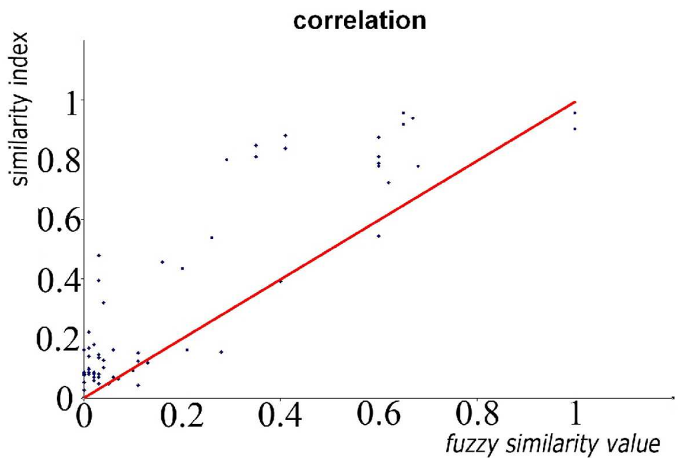


Fig. 5. Correlation between the fuzzy similarity and the average similarity index.

between b_i and a_i, $\frac{\sum_i^n (b_i - a_i)}{n}$, is 0.09.

5.4.4 Discussion

The aim of our experiment was to verify if the degrees of the similarity of computer generated facial expressions established by our algorithm are consistent with human perception. Firstly, we compared the weighted average of the subjects' answers with the values of our algorithm. We found that the subjects' answers and our algorithm results are positively correlated and that the correlation coefficient is high (89%). Thus we can say that *average similarity index*, asi_{AB} (corresponding to subjects' answers) tend to be proportional to the fuzzy similarity values (see Figure 5). The higher the index value is, the higher the fuzzy similarity value is as well. The mean difference between subjects' responses and our algorithm results is small (0.09).

We can notice some limitations in this study. For example, its result is influenced by the choice of the method for ranking the subjects' answers (i.e. asi_{AB}). In particular, we assumed that the distance between any two degrees of similarity was constant. We noticed that the pairs that receive average score using our fuzzy similarity are evaluated as relatively more similar by participants. Indeed the mean difference between subjects' responses and our algorithm results is positive. It means that the algorithm has a tendency to evaluate certain pairs of expressions as less similar in comparison with the subjects' choices. As shown in Figure 5 more points in this interval are situated above the diagonal than under it. Other limitation of this study is related to the quality of synthesized images. To exclude the possibility that imperfections of Greta's display influence people's judgement one may check if similarity is perceived in the same way for plausible or exaggerated facial expressions. Last but not least this method could also be extended to dynamical facial displays. Nevertheless, the algorithm based on the notion of fuzzy similarity meets our expectations and we use it to create complex facial expressions.

6 Complex facial expressions model

In the previous section we have presented an algorithm which allows us to compare any two facial expressions of an embodied agent. We use it to generate different types of facial expressions. Previous models (see section 3.1) of facial expressions deal with the display of emotional states. They are based on the assumption that emotions which are similar (for instance in terms of valence or arousal values) have similar expressions. On the contrary, we propose that the visual resemblance between two facial expressions is the measure that can be used to generate a new expression. Our algorithm computes the facial expression of the agent when it superposes, masks, fakes, and inhibits its emotional expressions. Our model of facial expressions is based on studies by Ekman and Friesen (Ekman and Friesen (1969, 1975); Ekman (2003a,b)). In conformity to this theory (see section 2.2) we define facial expressions using a face partitioning approach. Each facial expression is defined by a set of eight facial areas F_i, $i = 1,..,8$ (i.e., F_1 - brows, F_2 - upper eyelids, F_3 - eyes direction, F_4 - lower eyelids, F_5 - cheeks, F_6 - nose, F_7 - lips, F_8 - lips tension). An expression is the composition of these facial areas, each of which can display signs of emotion. For complex facial expressions, different emotions (e.g. expression of disappointment masked by happiness) can be expressed on different areas of the face (e.g. disappointment is shown on the eyebrows area while happiness is displayed on the mouth area).

The main task of our algorithm is to assign expressions of emotions to the different parts of the face. For this purpose we define a set of rules for each type of complex facial expression that describes the composition of facial areas. These rules, based on the description proposed by Ekman (see section 2.2), refer to six emotions, namely: anger, disgust, fear, happiness, sadness, and surprise. In the case of an input expression for which the complex facial expression is not explicitly defined by our rules (e.g. expression of contempt or disappointment) our algorithm chooses the closest solution. This closeness is computed by analysing *visual resemblance* between expressions. For this purpose we use the algorithm presented in section 5. Each facial expression is described by a set of fuzzy sets. Then fuzzy similarity is used to compute the degree of similarity between them. Finally an algorithm computes the composition of a new expression using these values and rules.

Before considering each type of complex facial expressions separately let us introduce the notation used in following sections. Let $Exp(E_i)$ be the simple expression of an emotion E_i, while $Exp(N)$ is the neutral expression. $Int(E_i)$ is the intensity of emotion E_i. EXP is a set of all simple expressions and $BASEEXP$ is the set composed of the six expressions analysed by Ekman. Only expressions from $BASEXP$ have explicit descriptions of complex facial expressions. Finally $F_k^{Exp(E_i)}$ is k-th facial area of expression of emotion E_i.

6.1 Superposition

Superposition happens when two emotions are felt at the same time. The resulting expression has some features of the expressions of both felt emotions, like in a "bitter-sweet" expression of superposition of happiness and sadness described by Paul Ekman where the raised brows of sadness are accompanied by a smile (Ekman and Friesen (1975); Ekman (2003b)). The superposition of two emotions is usually expressed by a combination of the upper part of one expression with the lower part of the other one (Ekman and Friesen (1975); Ekman (2003b)).

6.1.1 Superposition schemes

The six emotions analysed by Ekman give us 30 different ordered pairs of emotions. We classify them according to their face partition. It enabled us to distinguish 10 different superposition schemes. By *superposition scheme* SS_i we mean a particular division of the eight facial areas between any two emotions e.g. the facial areas F_1, F_2, F_3, F_4 (forehead, brows, eyelids, and eyes) belong to the expression of the *first* emotion and the facial areas F_5, F_6, F_7, F_8 (nose, cheeks, and lips) to the *second* one.

By Z we denote a set of superposition schemes of Ekman's expressions. Obviously, two different pairs of emotions can share the same superposition scheme. It means that two different ordered pairs of emotions divide the face in the same way e.g. both pairs: sadness and fear as well as anger and happiness create the superposition expression in which the F_1, F_2, F_3, F_4 are taken from the first expression (sadness or anger respectively) while the F_5, F_6, F_7, F_8 are taken from the second element of the ordered pair (fear or happiness respectively).

6.1.2 Algorithm

Our algorithm generates the expression of superposition $Exp_{Sup}(E_i, E_j)$ for any two expressions $Exp(E_i)$ and $Exp(E_j)$ by choosing one superposition scheme SS_i from a set Z of superposition schemes. In the first step, for each input (i.e. simple) expression $Exp(E_i)$ we establish its values of similarity with the expressions in the *BASEXP* set. Any simple expression can be represented by a vector with values in the interval [0,1] that corresponds to the degrees of similarity between that expression and the ones from *BASEXP*. A set of rules SFR_{sup} is used to create an expression of superposition. For each pair of expressions from *BASEXP* we define a rule that associates it with one SS_i. In a second step the values of fuzzy similarity *FS* between the input expression and elements of *BASEXP* are used to select an adequate superposition scheme from the rules of SFR_{sup}. Let us present our algorithm in more detail.

The input to our system consists of two emotion labels E_i and E_j. Our algorithm classifies the ordered pairs of expressions among the elements of *Z*. Each such pair can be described in our algorithm by 12 features: $[a_1, ..., a_6, b_1, ...b_6]$ $\in [0, 1]^{12}$. The parameters a_i (resp. b_i) correspond to the similarity values of the *Exp*(E_i) (resp. *Exp*(E_j)).

The fuzzy inference is used to model the superposition of facial expressions. We have defined 12 input fuzzy variables that correspond to the features of an input pair. Each variable expresses the resemblance to one element of *BASEXP*. We can think, for example, about "anger-likeness" as a scale of resemblance to the expression of anger, or "happiness-likeness" in the case of happiness. Formally, the membership value is the degree of similarity. It means that for every *Exp*(E_i) from *BASEXP*:

$$\mu_{Exp(Ei)}(Exp(E_j)) = FS(Exp(E_i), Exp(E_j)) \tag{7}$$

For example: $\mu_{Exp(anger)}(Exp(E_i)) = FS(Exp(anger), Exp(E_i))$ expresses the degree of similarity between the expression of anger and E_i.

Ten output variables of SFR_{sup} correspond to elements of *Z*. Each output variable corresponds to exactly one SS_i. The affiliation to the SS_i is represented by two singleton fuzzy sets: YES, NO. The algorithm uses 30 fuzzy rules. Each rule uses the subset of input variables presented above and all the output variables. Each rule associates a pair of emotions: E_u, E_w, where *Exp*(E_u), *Exp*(E_w) $\in$ *BASEXP*, with an element of *Z*. The rules are defined according to the following pattern: "if the input expression of E_i is (similar to) the expression of E_u and the input expression of E_j is (similar to) the expression of E_w then expression the $Exp_{Sup}(E_i, E_j)$ is of the type SS_i and is not of the type SS_1 and ... and is not of the type SS_{i-1} and is not of the type SS_{i+1} and ... and is not of the type SS_{10}", where *Exp*(E_u), *Exp*(E_w) $\in$ *BASEXP*, $SS_i \in Z$. Let us present an example of such a rule. The ordered pair for the emotions sadness and happiness is described in SFR_{sup} by the following rule:

SUP_{26}: if *Exp*(E_i) is *sadness* and *Exp*(E_j) is *happiness* then
S_1 is NO and S_2 is NO and S_3 is NO and S_4 is NO and S_5 is YES and

S_6 is NO and S_7 is NO and S_8 is NO and S_9 is NO and S_{10} is NO.

It means that if the first input expression is similar to sadness and the second input expression is similar to happiness, then the final expression is of the type SS_5, and at the same time, is not of the type SS_i, where $i = 1,..,10$, $i \neq 5$. SS_5 corresponds to the situation in which the facial areas F_1, F_2, and F_3 belong to $Exp(E_i)$ while the other facial areas belong to $Exp(E_j)$. It means that the final expression contains brows, upper eyelids, and eyes of the first expression (i.e. sadness) and the rest of the second one (i.e. happiness).

As proposed by Kuncheva (2000) for SFR_{sup} we use the product operator for conjunction and the max for aggregation. Finally, we use the maximum defuzzification method. As a result, we obtain a ten-element vector $[g_1, .., g_{10}]$, $g_i \in [0,1]$ that is used to decide which superposition scheme has to be applied. Then:

- if there is only one index k such as $g_k = 1$ and $g_i = 0$ for $i \neq k$ then SS_k is chosen;
- if there are k and l, $k \neq l$ such that $g_k = g_l > 0$ then the superposition scheme is randomly chosen between SS_k and SS_l.

6.1.3 Example

Figure 6 presents an example of the superposition expression computed by our model. Figures 6a and 6b show the expressions of anger and fear respectively. In Ekman and Friesen (1975) the expression of anger is characterised by the brows that are lowered and drawn together in a frown. The lids are tensed and the eyes have a hard glare. The upper lids can be lowered and the lower lids maybe raised. The lips are pressed tightly together with the corners straight. The expression of fear is characterised by brows and the upper eyelids that are raised. The lower eyelids are tensed and drawn up. The mouth is usually open with the lips that are drawn back. They can be tensed slightly or stretched. Figures 6c and 6d show the superposition as a composition of facial areas of both input expressions. In Figure 6d we can see which parts of the face correspond to fear and which ones to anger. In that image the areas F_5, F_6, F_7, and F_8 (expressing fear) are marked by the yellow circles while areas F_1, F_2, F_3, and F_4 (expressing anger) by a blue colour.

Even if the intensity of the emotional state is not an explicit factor of our model, it is implicitly considered as it affects the specification of the facial expression of emotion. Figure 7 shows the superposition of two different happiness expressions with the sadness expression. According to Ekman, in the expression of happiness the lower lids are raised without tension while the cheeks are raised and the corners of lips are raised and drawn back. It is also characterised by the wrinkles which are visible around the outer corners of the

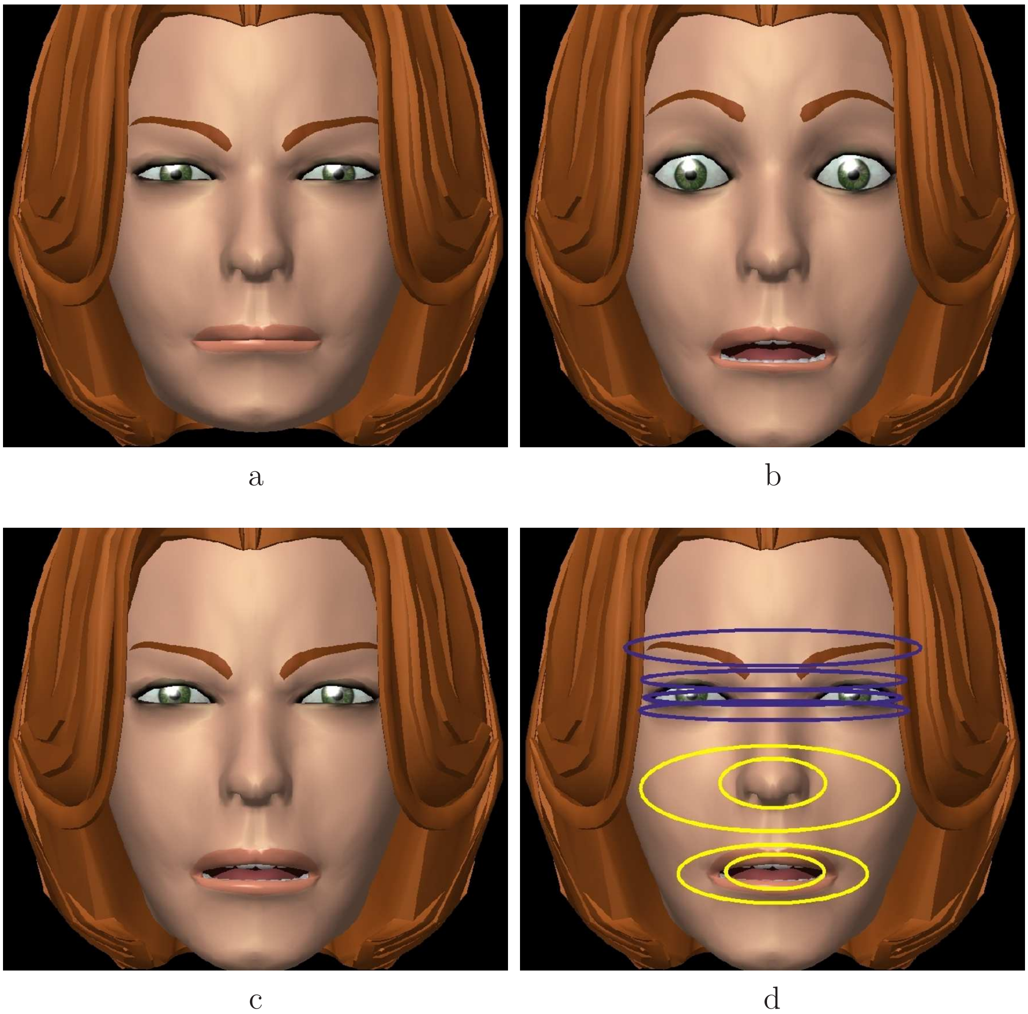


Fig. 6. Superposition of anger and fear. From the left to right: anger a), fear b), superposition of anger and fear c) superposition of anger and fear with significant areas marked d).

eyes. In the case of intense smile the mouth may be open so that teeth are exposed. Figures 7a and 7b present two expressions of happiness of different intensities. One can see that in Figure 7c the happiness component is also less intense (comparing to Figure 7d) as it comes from the less intense expression of happiness (Figure 7a).

6.2 Masking

Masking occurs when a felt emotion should not be displayed for some reason; it is preferred to display a different emotional expression. The masking expression is influenced by deception clues described in section 2.2. The felt emotion leaks over the masked one according to the inhibition hypothesis. On the other hand, the fake expression is not complete as it lacks the reliable

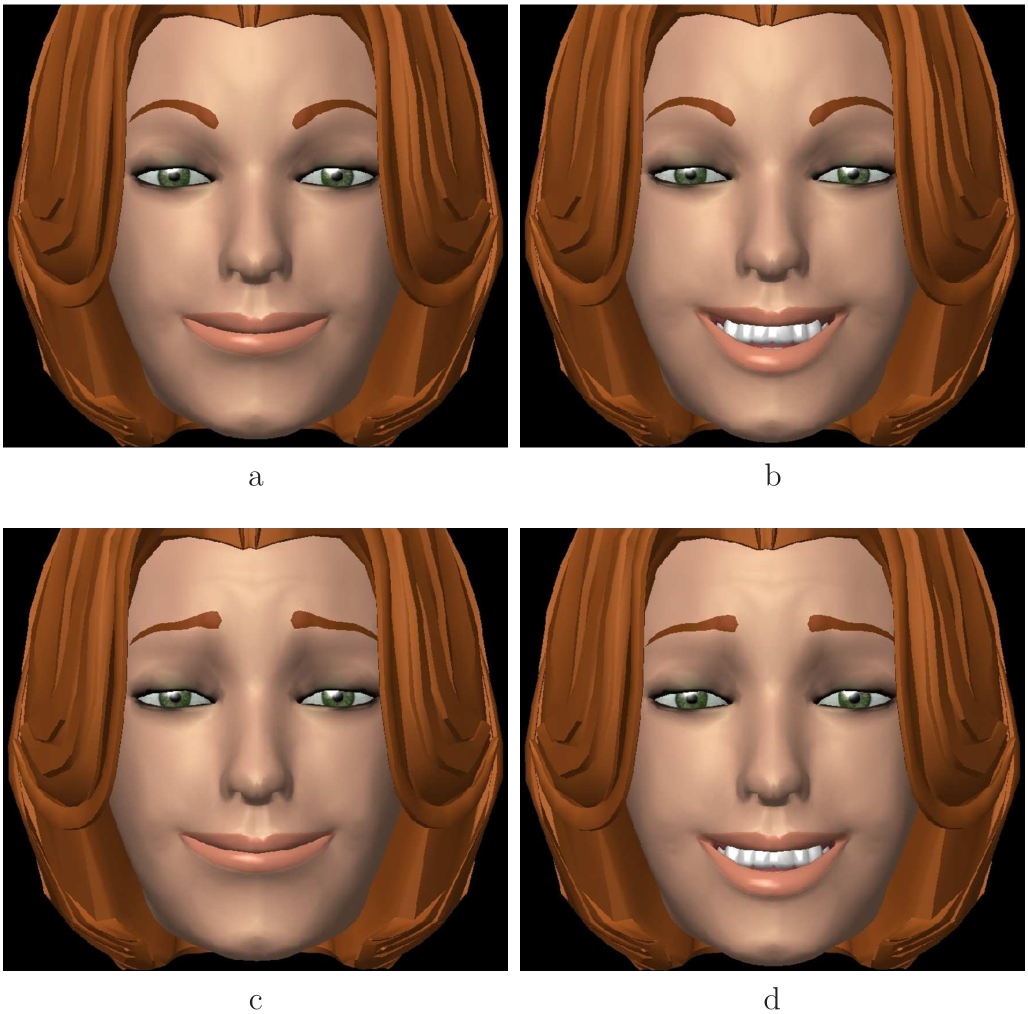


Fig. 7. Superposition of happiness and sadness. From the left to right: happiness a), open-mouth happiness b), superposition of happiness and sadness c) superposition of open-mouth happiness and sadness d).

features (Ekman and Friesen (1975); Ekman (2003b)).

For each deception clue we define in our model a separate set of rules. The SFR_{inh} describes which elements of the felt facial expression are expected to appear even if the expression is concealed. Then the SFR_{rf} specifies the facial areas that do not occur in a fake expression. In order to create the facial expression of masking we use both sets of rules.

6.2.1 Reliable features

SFR_{rf} is a set describing rules for reliable features of expressions from *BASEXP*. Each rule indicates the reliable features of an expression. We map each reliable feature to the facial areas F_1,..., F_8. Each input variable of SFR_{sup} corresponds to one expression from $BASEXP$ and each output variable corresponds to one facial region F_k of the resulting expression. We define six input fuzzy variables. Each variable expresses the resemblance to one of the expressions from the $BASEXP$ set. We also define eight output variables, each of them related to a particular facial area. The output variable expresses "the possibility of the occurrence" (POS_{F_k}) of a facial area F_k of a given expression $Exp(E_i)$. We use the fuzzy sets to describe the gradation of POS_{F_k} in terms of linguistic labels: high, may_occur, low. Each rule RF_i says: "the more the input expression of E_i is (similar to) the expression of E_u, the more the possibility of occurrence of area F_k in the final expression is $\{high, low, may\_occur\}$". For example the rule RF_4 is: "the more the input expression is (similar to) the expression of happiness, the more the possibility of occurrence of the lower_eyelids area of the input expression is *low*, the possibility of occurrence of brows and upper_eyelids areas is *may_occur*, and the possibility of occurrence of other areas is *high*". Reliable features of happiness *are located* in the area of lower_eyelids, while the facial areas F_1 and F_2 *may* betray the felt emotion (Ekman and Friesen (1975)). It corresponds to the following rule of SFR_{rf}:

RF_4: the more $Exp(E_i)$ is *happiness*,
the more POS_{F_1} is *may_occur* and POS_{F_2} is *may_occur* and
... and POS_{F_4} is *low* and ... ... and POS_{F_7} is *high* and POS_{F_8} is *high*.

The min operator for conjunction and the max for aggregation are used for SFR_{rf}. As a result, we obtain a set of fuzzy values. Each of them expresses the possibility that certain F_k of the input expression is displayed.

6.2.2 Inhibition hypothesis

Another set of rules, SFR_{inh}, is defined on the basis of the inhibition hypothesis. Each rule of SFR_{inh} indicates the elements of facial expressions that leak over the mask. For this purpose we map each leaking feature to facial area from the set F_1,...,F_8. Activity in these facial areas can be observed even if the expressions are inhibited.

In the previous section we described how we can predict the absence of some parts of expression in any voluntary display. We apply such an approach for the inhibition hypothesis. The input and output variables are identical to the ones in the previous case. Thus each input variable corresponds to one expression from $BASEXP$, and each output variable corresponds to one facial area F_k of the resulting expression. Each rule of SFR_{inh} describes the features which appear even if the facial expression of E_i is hidden or masked. Thus, each rule INH_i says: "the more the input expression of E_i is (similar to) the expression of E_u, the more the possibility of occurrence of area F_k in the final expression is $\{high, low, may\_occur\}$". For example, the following rule is applied for sadness: “the more the input expression is (similar to) sadness, the more the possibility of occurrence of the brows and upper_eyelids areas is high and the possibility of occurrence of other areas is low”. Formally, INH_5 is:

INH_5: the more $Exp(E_i)$ is *sadness*,
the more POS_{F_1} is *high* and POS_{F_2} is *high* and
and POS_{F_3} is *low* and ... and POS_{F_8} is *low*.

6.2.3 Algorithm

The input to our system consists in specifying two emotion labels: the felt one (E_i) and the fake one (E_j). First, the values of fuzzy similarity FS are established for their expressions and the elements of $BASEXP$. Similarly to superposition case each facial expression $Exp(E_i)$ (resp. $Exp(E_j)$) is represented by a 6-element vector $[a_1,..., a_6]$ (resp. $[b_1,..., b_6]$) of real values in the interval [0,1]. Then the elements of the final expression are processed separately. The vector $[a_i]$ of the felt expression of emotion E_i is processed by SFR_{inh}, while the vector $[b_i]$ of the fake expression of emotion E_j is processed by SFR_{rf}.

SFR_{inh} and SFR_{rf} are complementary. Both return the predictions about the occurrence of $F_k, k = 1..8$. For each F_k, the results of SFR_{rf} and SFR_{inh} are combined in order to obtain the masked expression. In particular, for each facial area F_k the following can happen:

- C1) POS_{F_k} of the felt expression $Exp(E_i)$ is high and the POS_{F_k} of the fake expression $Exp(E_j)$ is also high. It results in a case of conflicting predictions. It means that both $F_k^{Exp(E_i)}$ and $F_k^{Exp(E_j)}$ are candidates to be shown. In this case the felt expression should be expressed as it is difficult to inhibit it voluntarily.
- C2) POS_{F_k} of the felt expression $Exp(E_i)$ is high and POS_{F_k} of the fake expression $Exp(E_j)$ is low. Then $F_k^{Exp(E_i)}$ is used.
- C3) POS_{F_k} of the felt expression $Exp(E_i)$ is low and POS_{F_k} of the fake expression $Exp(E_j)$ is high. Then $F_k^{Exp(E_j)}$ is used.
- C4) POS_{F_k} of the felt expression $Exp(E_i)$ is low and POS_{F_k} of the fake expression $Exp(E_j)$ is low. It means that neither F_k of $Exp(E_i)$ nor of $Exp(E_j)$ can be used. In this situation $F_k^{Exp(N)}$ (neutral expression) is used instead.
- C5) The possibilities of occurrence of both: F_k of $Exp(E_i)$ and F_k of $Exp(E_j)$ are somewhere between high and low. It means that both F_k may occur. F_k is chosen randomly between $Exp(E_i)$ and $Exp(E_j)$.

Thus, the final expression is composed of facial areas of the felt emotion, the fake, and the neutral ones. More precisely, SFR_{inh} and SFR_{rf} return two different sets of fuzzy values. Each of these fuzzy values expresses the potential occurrence of certain F_k. The output of SFR_{inh} (resp. SFR_{rf}) is a set of eight fuzzy values corresponding to eight different facial areas of $Exp(E_i)$ (resp. $Exp(E_j)$). We use them as input to the system of rules SFR_{mask}. The outcome of these rules corresponds to the four possible choices described in five conditions C1 – C5: $F_k^{Exp(E_i)}$, $F_k^{Exp(E_j)}$, $F_k^{Exp(N)}$, and "random". SFR_{mask} has nine fuzzy rules based on conditions C1 – C5. Each of them describes which facial area F_k should occur in the final expression.

For each input variable we use trapezoid fuzzy sets that correspond to the terms "high", "low", and "may occur" of the possibility of occurrence. The output of the SFR_{mask} is composed of four variables. Three of them correspond to three facial expressions that can potentially be used: felt, fake, and neutral. The fourth one ("called random") corresponds to the situation in which we need to choose randomly between felt and fake expression as the information is not sufficient to decide anything. For each output variable we introduce two singleton fuzzy sets corresponding to the possible decisions: {YES, NO}. As a consequence, our composition rules are defined according to the following pattern: if the possibility of occurrence of F_k of E_i is A_l and the possibility of occurrence of F_k of E_j is B_l, then F_k of the felt expression is (YES/NO) and F_k of the fake expression is (YES/NO), and F_k of the neutral expression is (YES/NO), and the random case is (YES/NO)". For example the case POS_{F_k} being high for both E_i and E_j (condition C1) is represented by the following rule:

if POS_{F_k} of E_i is *HIGH* and POS_{F_k} of E_j is *HIGH* then
FELT is *YES* and *FAKE* is *NO* and
NEUTRAL is *NO* and *RANDOM* is *NO*.

To maintain consistency with other systems we use the min operator for conjunction and the max for aggregation. Finally, we use the maximum defuzzification method. As a result, we obtain a four-element vector $[\mathrm{g}_k]$ of values, where:

- g_1 corresponds to the $Exp(E_i)$,
- g_2 corresponds to the $Exp(E_j)$,
- g_3 corresponds to the $Exp(N)$,
- g_4 corresponds to the "random case". The choice is made between $Exp(E_i)$ and $Exp(E_j)$.

Two situations are possible for the values g_1 - g_4:

- usually there is only one index k such that $g_k = 1$ and $g_i = 0$ for $i \neq k$. Then the facial area of the corresponding facial expression is used. If $k = 4$ we choose randomly between the facial areas of $Exp(E_i)$ and $Exp(E_j)$.
- if there are two indices k and l, $k \neq l$ such that $g_k = g_l > 0$ it means that

the facial areas of both expressions can be used. Then a facial area F_k is randomly chosen between the expressions corresponding to k and l.

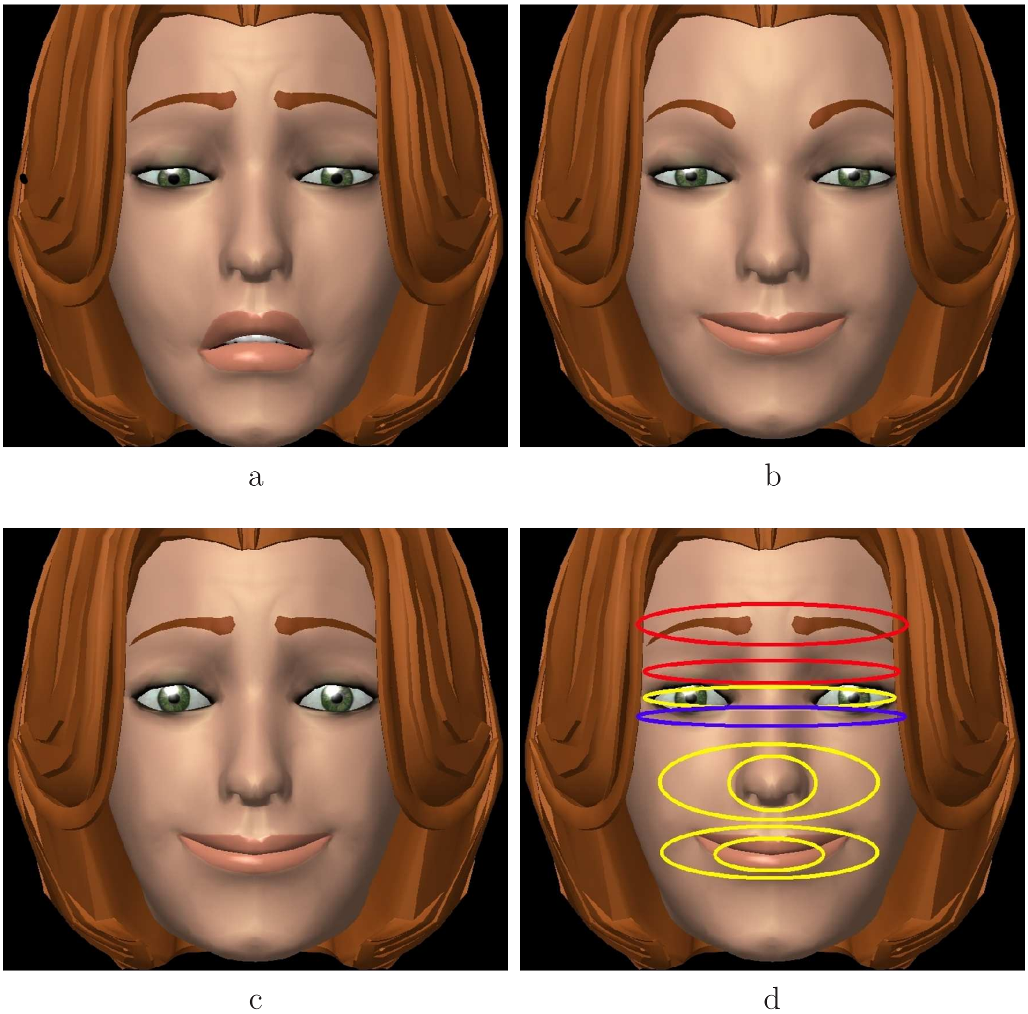


Fig. 8. Disappointment masked by happiness. From the left to right: disappointment a), happiness b), disappointment masked by happiness c) disappointment masked by happiness with significant areas marked d).

6.2.4 Example

Figure 8 presents the agent displaying the expression of disappointment masked by a fake happiness. Let us explain how we obtain the complex expression displayed by the agent. We applied our similarity algorithm and found that facial expression of disappointment is very similar to the one of sadness. In our model the features of felt sadness that leak over the masking expression are: forehead, brows, and upper eyelids. These elements are represented by the facial areas F_1 (forehead and brows) and F_2 (upper eyelids). According to the inhibition hypothesis, they can be observed in masked sadness. The expression of disappointment (Figure 8a) is very similar to the expression of sadness according to the similarity algorithm. So the rules of sadness will be applied for the disappointment expression. In the expression of disappointment masked by fake happiness (Figure 8c) we can notice the movement of brows, which is a characteristic of disappointment. On the other hand, the mouth area displays a smile (sign of happiness).

6.3 Fake and inhibited expressions

Two other types of facial expressions occur often in real life. Inhibition takes place when an individual avoids to express his emotions. Instead, he tries to show no emotion at all. The inhibited expression can still leak over the neutral face. On the contrary, a fake expression takes place when an individual tries to express emotions that he does not feel at the moment. This expression is different from the spontaneous one as people are not able to control all their facial muscles and thus do not display the full expression of emotion (see section 2).

Similarly to other cases of facial expression management, the fake and inhibited expressions can be detected by deception clues. The felt emotion leaks over the facial mask according to the inhibition hypothesis while a fake expression is incomplete as it lacks reliable features. It means that SFR_{inh} needs to be used for inhibited expressions, while SFR_{rf} - for fake expressions. The expression of inhibition can be seen as hiding the felt emotion under the "mask" of the neutral expression. Similarly, making a fake expression means "masking" the neutral facial expression by some fake expression. Thus we can use the same procedure that we used in the case of masking for fake or inhibited expressions. We introduce a slight modification to the algorithm presented in the previous section: we add rules for the neutral expression $Exp(N)$ to the sets SFR_{rf} and SFR$_{inh}$. We assume that "false neutral expression" can be easily made deliberately. Thus we add to SFR_{rf} a new trivial rule (RF_7): "the more the input expression is (similar to) the neutral expression, the more the possibility of occurrence of any area is high". Then we also add a new trivial rule (INH_7) to SFR_{inh}: "the more the input expression is (similar to) the neutral expression, the more the possibility of occurrence of any area is low". It is so as the neutral expression does not involve any particular facial movement.

For fake and inhibited expressions the input to the system is an ordered pair of emotion labels (one of them corresponds to the neutral expression). As for any other expression type, the values of similarity between the input expression and the seven facial expressions (six $BASEXP$ expressions and the neutral expression) is computed. $Exp(E_i)$ is represented as a seven-element vector $[a_1,...,a_7]$ of real values in the interval [0,1]. Then, we use the algorithm presented in section 6.2.3.

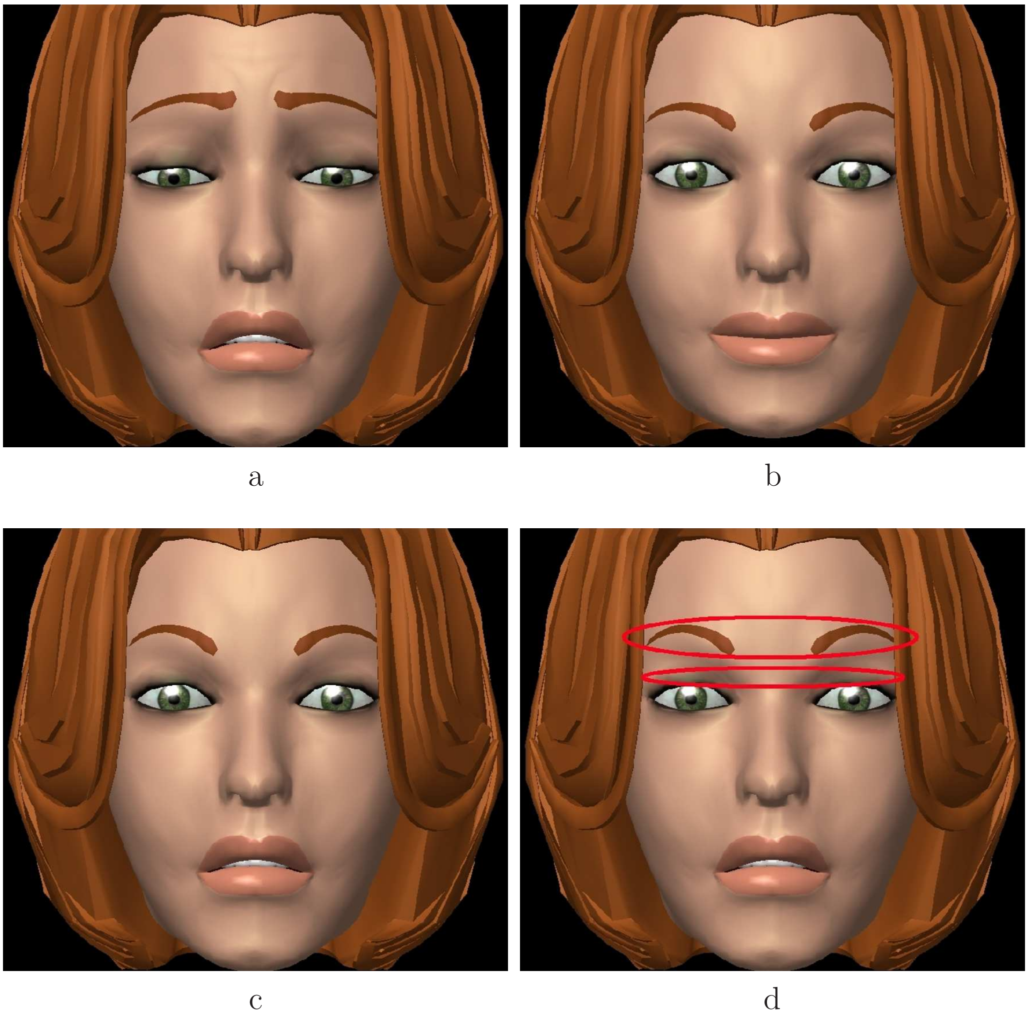


Fig. 9. Fake distress. From the left to right: distress a), neutral expression b), fake distress c), fake distress with significant areas marked d).

6.3.1 Example

In Figure 9 and 10 we present examples of fake expressions and inhibited expressions of non-basic expressions. The non-basic expressions used in the examples were created from the description in the literature (e.g. Matsumoto and Ekman (2004)) or from the multilevel annotation worked out by Jean-Claude Martin et coll. (e.g. Abrilian et al. (2005)). Figure 9 shows the fake expression of distress (Figure 9c and 9d). We can compare it with the felt distress expression (Figure 9a) and the neutral expression (Figure 9b). The expression of distress is the most similar to the expression of sadness as returned by the similarity algorithm. Thus the complex facial expressions algorithm applies the rules of sadness. The facial areas F_1 and F_2 (eyebrows and upper eyelids) cover the reliable features of felt sadness. As a consequence,

F_1 and F_2 are missing in fake sadness and thus they are also missing in fake distress. In Figure 9c only the mouth of distress is visible (this facial area is signalled in Figure 9d) while the eyebrows and the upper eyelids display neutral expression.

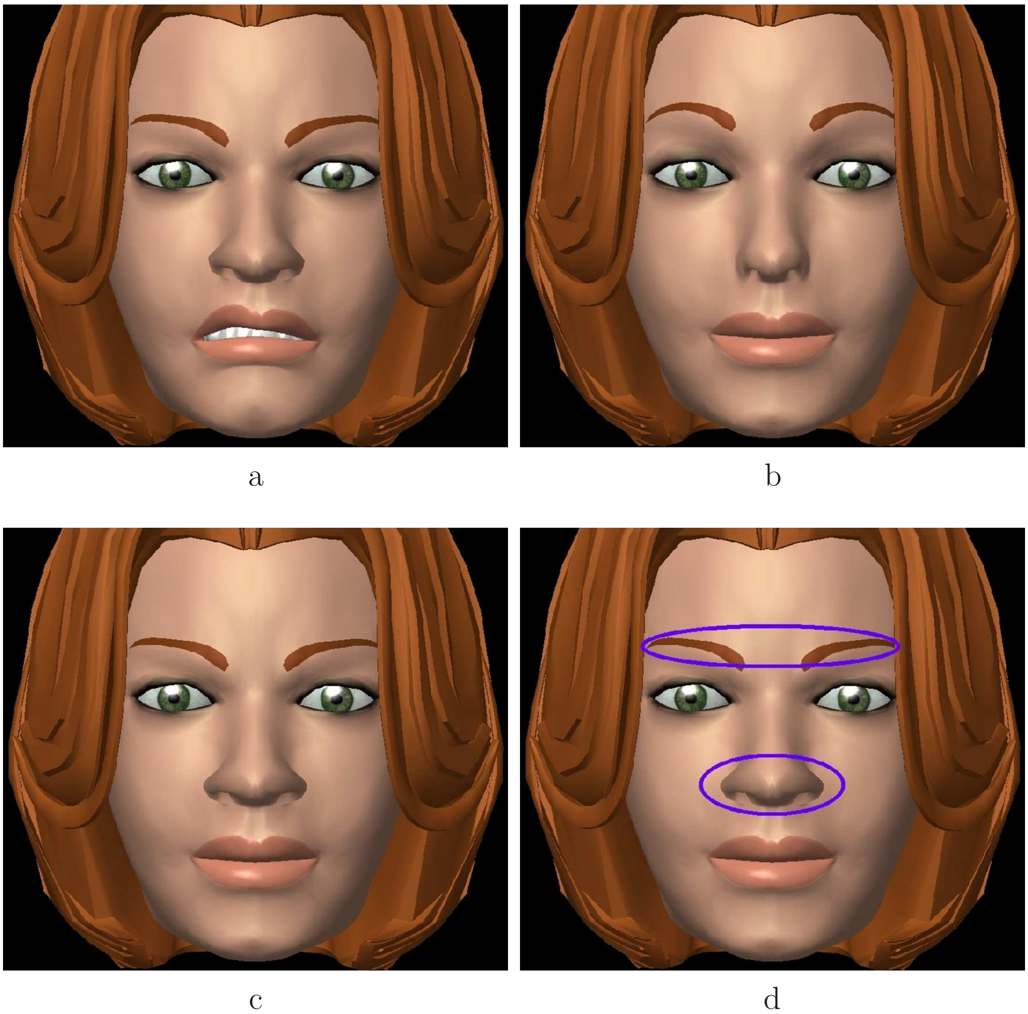


Fig. 10. Inhibited contempt. From the left to right: contempt a), neutral expression b), inhibited contempt c), inhibited contempt with significant areas marked d).

In Figure 10 we can see another complex facial expression of non-basic expression, i.e. the inhibited expression of contempt. The expression of contempt is expressed by the unilateral lip corner raise and tighten (Action Units 12 and 14) (Matsumoto and Ekman (2004)) which might be accompanied by a slight frown and nose wrinkling (Ekman (2003b)). We can compare the inhibited expression of contempt (Figure 10c and 10d) with the felt expression of contempt (Figure 10a) and the neutral expression (Figure 10b). The expression of contempt is very similar to the expression of disgust. Then the facial areas F_1 (eyebrow) and F_5 (nose) cover the features of felt disgust that leak over the mask. Consequently, in Figure 10c only frown and nose wrinkles are visible (these facial areas are signalled in Figure 10d) while lip corner raise is masked by the neutral expression.

6.4 Evaluation [2]

Complex facial expressions generated with our model were evaluated. For this purpose two videos from EmoTv corpus (Abrilian et al. (2005)) were annotated with different types of complex facial expressions. Four different animations of Greta agent were generated for each video. In the first two animations the agent displayed one of the two emotions indicated by the annotators. The third animation used complex emotion label and our complex facial expression algorithm to compute the associated expression. The fourth video was obtained from the manual annotation of the behaviour i.e. a description of the facial expressions.

We evaluated the quality of the animations by asking subjects to compare them with the real videos in two conditions (with and without audio). Participants were asked to order the four animations from the most similar to the least similar one in comparison with the original video. In both conditions the rate of being the most similar animation was measured.

Forty subjects (23 males, 17 females), French native speakers, aged between 19 and 36 (average 24) participated in the experiment. Each of them had to compare the original videos with Greta's four animations. After the experiment they also filled in a questionnaire in which they had to annotate the emotions they perceived in the animation they ranked as the most similar to the original video. They could select one or several labels from a closed list. The subjects were rigorously different from the participants of the experiment described in section 5.4.

The first video (17 sec. long) was recognised by annotators as an example of masking. In both conditions (audio/without audio) facial expressions created by our model were rated very high (see Table 1). In the audio condition this animation was considered as the most similar to the original video (38%) and in the second condition it took the second place (33%). In both conditions our animation was considered at least as good as the animation generated from manual annotation (38% vs. 27%, and 33% vs. 20%, respectively).

The second video (15 sec. long) was intended to present an example of superposition of emotions. However the animation displaying a simple expression was perceived as the closest to the original by 33% of the participants in the audio condition, and by 61% of the participants in the no audio condition (see Table 2). Comparing two animations of complex facial expressions the one

[2] Work conducted together with Jean-Claude Martin and Stephanie Buisine

facial expression	audio condition	no audio condition
one emotion displayed		
disappointment	11%	7%
happiness	24%	40%
two emotions displayed		
manual definition	27%	20%
our algorithm	38%	33%

Table 1
The animation most similar to the first video (Buisine et al. (2006)).

generated with our model of complex facial expressions received 17% (audio) and 9% (no audio condition) in turn while the manually defined complex facial expressions were better in this test (24% and 20%)(see Buisine et al. (2006) for detailed results).

In summary, according to the results of this evaluation study our complex facial expressions are perceived by the observers. The animations created with our model obtained a satisfactory result when compared with the manually created animations of complex expressions. In the first video (masking case) the automatically generated expression was evaluated as good as the manually defined complex expression in audio and no audio conditions. The low result in the second video (superposition case) may be due to some multimodality issues. It may be that subjects perceived despair mainly from the voice. It could explain the strong preference for the anger display and the very low result for the despair animation in the no audio condition. The effect of multimodality (i.e. combination of audio and visual data) should be studied more in depth in the future.

facial expression	audio condition	no audio condition
one emotion displayed		
anger	33%	61%
despair	26%	9%
two emotions displayed		
manual definition	24%	20%
our algorithm	17%	9%

Table 2
The animation most similar to the second video (Buisine et al. (2006)).

7 Complex facial expression applied to management display

In this section we analyse certain factors that influence the display of facial expressions in interpersonal relations, and we describe a model of facial behaviour management derived from this analysis for an ECA. We model interpersonal relations within Brown and Levinson's politeness theory. We consider two out of the three factors of Brown and Levinson's theory, namely social power and social distance. Thus, in our work, interpersonal relations are characterized by these two factors. The third factor we take into account is the interlocutor's emotional state. To find the relations between these three factors and the occurrence of a particular type of complex facial expressions, we analyzed a video corpus where we looked for the co-occurrence of complex facial expressions and politeness strategies. Our rules of facial behaviour management are based on the results of the annotation of this video-corpus. Our algorithm computes whether the virtual agent can display its felt emotions; if not, it outputs the type of facial behaviour management (masking a felt emotion by another one, inhibition, displaying a fake emotion) the agent should show.

7.1 Politeness strategies

Brown and Levinson (1987) proposed a computational model of politeness in language. According to this theory, any linguistic act like request or promise can threaten the "face" of the speaker and/or the hearer. Politeness consists in taking remedial actions to counterbalance the negative consequences of these face acts. Brown and Levinson classified all polite verbal behaviours i.e. actions that prevent eventual negative consequences of acts. They defined five different strategies of politeness: bald, positive politeness, negative politeness, off-record, and "don't do the action". The strategies are ordered according to the impact they have on avoiding threatening situations. The bald strategy does nothing to minimise threats to the face. The speaker's message is clear and non-ambiguous, and the speech act preserves the maximum efficiency. The positive politeness strategy is used to protect the positive image of the addressee and to give him the impression that the speaker supports some of the hearer's goals (or that the speaker likes the hearer). The negative politeness strategy consists in assuring that the hearer's freedom is respected. It is characterised by formality or self-restraint. In the off-record strategy the message is communicated in an ambiguous way in order to avoid face threatening act. Finally, the fifth strategy - "don't do the action" - allows the speaker to avoid negative consequences but, at the same time, it precludes the communication of his intentions.

The decision about the strategy to be used depends on the value of Face

Threat Act (FTA). Brown and Levinson proposed to estimate FTA by using three variables: the *social distance*, the *power relation*, and the *absolute ranking of imposition* of an act. Social distance refers to the degree of intimacy and the strength of the relation, while social power expresses the difference in status and the ability to influence others. The last parameter depends on the objective importance of an act in a specific culture or situation. FTA value is calculated as the sum of these three values. Finally, the more antagonistic a given act is (higher FTA value), the more likely a high ordered strategy is to be chosen (Brown and Levinson (1987)).

7.2 Video-corpus

Brown and Levinson's theory was used by Rehm and André (2005) who aimed to analyse the relationship between different gestures types and politeness strategies in verbal acts. Rehm and André (2005) built a video-corpus called SEMMEL that contains various examples of verbal and nonverbal behaviours during face threatening interactions. They found that gestures types are indeed related to politeness strategies. However, facial expressions had not been considered in their study. Inspired by their results we decided to analyse the SEMMEL video-corpus in order to find any relations between politeness strategies and facial expressions.

7.3 Annotation scheme and results

We used 21 videos of the SEMMEL corpus involving eight different protagonists. The overall duration of the analysed clips is 6 minutes and 28 seconds. In this study we used the original annotation of the politeness strategies done by Rehm and André. They considered four politeness strategies: bald, positive politeness, negative politeness, and off-record strategy (Rehm and André (2005)).

To the existing annotation schema we added the annotation of facial expressions. We considered four types of facial expressions: felt, inhibited, masked, and fake expression. So far the videos were annotated by one person. We annotated only one feature of emotion: its valence, i.e. we distinguished between positive, negative emotions, and neutral state. Instead of annotating separate emotional states and their corresponding expressions, we annotated the videos with *patterns* of facial behaviours. Each pattern corresponds to one of the three values of valence (positive, neutral, negative) and one of the four types of (complex) facial expressions. For example, a pattern called "positive masked" describes any facial expression that occurs in a situation in which any positive emotion is masked by another one. Nine patterns of facial expressions were considered in the annotation process: negative masked, negative inhibited, negative expressed, fake negative, neutral expression, fake positive, positive masked, positive inhibited, positive expressed.

Pattern	Strategy				All
	bald	positive	negative	off-record	
negative masked	0	0	1	4	5
negative inhibited	0	0	1	2	3
negative expressed	0	2	0	2	4
fake negative	0	0	0	0	0
neutral expression	4	8	34	7	53
fake positive	0	5	16	6	27
positive masked	0	0	0	0	0
positive inhibited	0	0	2	0	2
positive expressed	2	3	1	2	8
All	6	18	55	23	102

Table 3
Occurrence of the different patterns of facial expressions.

The detailed results of our annotation are presented in Table 3. It shows the link between politeness strategies and facial expression patterns. We can see that the patterns of facial expressions are not evenly distributed along the strategies of politeness. An emotional state was observed in 20% of the facial expressions, while 80% of the expressions were annotated as being displayed without being in any emotional state. Indeed, the "neutral expression" pattern was the most often observed (52% of all cases). "Fake positive" pattern was observed 26.5%. Some patterns were not observed at all. None of "positive masked" expressions or "fake negative" expressions was annotated. The occurrence of certain patterns was correlated with the occurrence of a particular politeness strategy. The negative masked pattern co-occurs with the off-record strategy (4/5), the neutral expression with the negative strategy (38/53), and the positive inhibited with the negative strategy (all observed cases).

Patterns of facial expressions are connected to politeness strategies. Moreover from Brown and Levinson's theory (see section 7.1) the choice of the strategy depends on the values of the social power and social distance factors. We can conclude that there is a link between the facial patterns and these factors. The type of link is further explained in section 7.4. On the other hand, the annotated data is quite small. The observed cases, linking strategies and facial patterns, are not equilibrated. Some politeness strategies were not seen at all in the data. While the conclusion derived from the results of our study cannot be used to draw a complete model of facial management in interpersonal relation, it does show the emergence of some display patterns between politeness strategies and facial expressions.

7.4 Facial Management Algorithm

In this section we explain how our embodied agent adapts its expressive behaviour depending on the social power and social distance factors. We introduce a set of rules of facial behaviour management for our agent. For each strategy of politeness we have chosen the most characteristic pattern of facial expressions according to the annotation results. The pairs (politeness strategy, pattern of facial expressions) were used to define the rules that our agent will apply in interactions.

7.4.1 Variables

In our model, we consider three variables. Two of them encompass the characteristics of the interaction, namely: social distance (SD) and social power (SP). The third one, valence of emotion (VAL), describes the emotional state of the displayer. Social distance (SD) and social power (SP) are important factors that describe interpersonal relations. According to Wiggins et al. (1988) all personality traits relevant to social interaction can be located in a two dimensional space defined by the orthogonal axes of dominance and affiliation. So two variables: dominance (corresponding to SP) and affiliation (corresponding to SD) are sufficient to describe interpersonal relations. Moreover Brown and Levinson include SP and SD in their theory of politeness (see section 7.1). Power (SP) and social distance (SD) are two factors that influence human's expressions according to various studies on facial behaviour (Buck et al. (1992); France and Hecht (2005); Wagner and Smith (1991), see also section 2).

Facial behaviour management is also conditioned by emotional factors. In particular, it depends on the valence (VAL) of emotion. Negative emotions are more often masked or inhibited, while positive emotions are often pretended (Buck et al. (1992); Manstead et al. (2005)).

7.4.2 Rules

Let us consider three different classes of emotional states: negative, positive, and neutral. For each of them we looked for the pattern of facial behaviour that coincides the most with each politeness strategy. The choice is based on the frequency of co-occurrence for the strategy j and the pattern i in the annotated video clips (see Table 3). In more detail, for each strategy of

face threat	bald	positive	negative	off-record
positive emotion	positive expressed	positive expressed	positive inhibited	positive expressed
neutral state	neutral expressed	fake positive	neutral expressed	fake positive
negative emotion	negative expressed	negative expressed	negative inhibited	negative masked

Table 4
Facial behaviour and strategies of politeness.

politeness j ($j = 1, .., 4$) and each emotional state k ($k = 1, .., 3$) we choose the pattern i ($i = 1, .., 10$) such that the value *a(i,j,k)*:

$$a(i, j, k) = \frac{x_{ijk}}{\sum_{z=1}^{4} x_{izk}} \tag{8}$$

is maximal (the value x_{ijk} expresses the co-occurrence of the i-th pattern of a facial behaviour and the strategy j in the emotional situation k). In the situations in which the data gathered in the annotation study was insufficient to make a choice, we used the conclusions from the experiments of Buck et al. (1992), France and Hecht (2005), and Manstead et al. (2005). In Table 4 we report which pattern of facial expression i will be used for each type of emotion (positive, neutral, negative) and each strategy of politeness.

7.4.3 Generation of socially adequate expressions

The values of social power (SP), distance (SD) and the label of an emotional state E_i are the inputs of our model. SP and SD take values in the interval [0,1]. The emotional state is described by an emotional label from a finite set of labels. This set contains, apart from neutral state, emotions whose expressions can be displayed by the agent. The valence $VAL(E_i)$ of an emotion E_i is computed using the dimensional model of emotions. Thus, in our model any emotional state can be either *positive* or *negative*, while the neutral emotional state is *neutral*.

In Brown and Levinson's theory the choice of strategy depends on the sum of the values of social power and distance. In our model since we reverse the direction of the power variable (we look at who has the power rather than who is dominated), we use the difference operator. Let: w be the difference between social power and social distance: $w = SD - SP$ ($w \in$ [-1,1]). We define an interval of acceptable values for each strategy. For this purpose we split the interval of all possible values of w into four equal parts: $w \in$ [-1, -0.5] (very low) is associated with the bald strategy, $w \in$ (-0.5,0] with positive politeness,

$w \in (0,0.5]$ with negative politeness, while $w \in (0.5,1]$ (very high) with the off-record strategy. Our facial management rules (see Table 4 for details) are of the type:

$$\begin{gathered}\text{if } VAL(E_i) \text{ is } \{\text{positive} \mid \text{negative} \mid \text{neutral}\} \text{ and} \\ w \text{ is } \{\text{very low} \mid \text{low} \mid \text{high} \mid \text{very high}\} \text{ then} \\ \text{the expression of } E_i \text{ is } \{\text{expressed} \mid \text{fake} \mid \text{inhibited} \mid \text{masked}\}.\end{gathered}$$

For fake positive pattern the system uses the expression of fake joy while for negative masked pattern the input expression is masked by an expression of joy.

Recapitulating, our algorithm works as follow (see Figure 11): given an emotional state E_i and values of social distance SD and of social power SP, the algorithm generates an adequate facial expression using the facial management rules generated with the model presented in section 6.

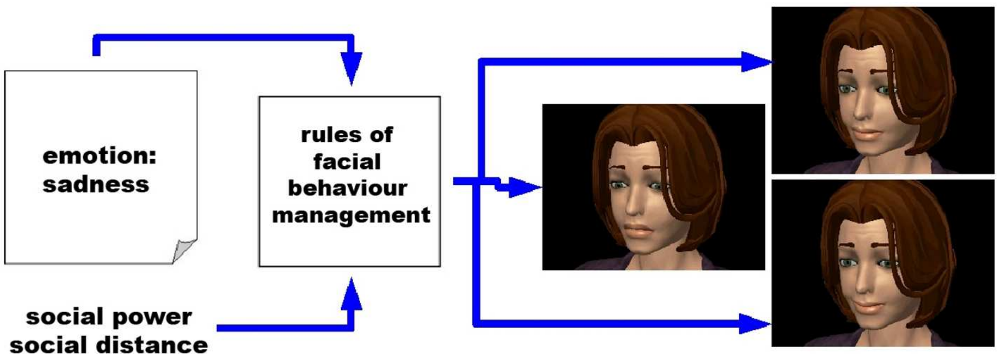


Fig. 11. Facial behaviour management in Greta.

In Figure 1 (see section 4) we presented the process of facial behaviour management. A person, in an emotional state E, adapts his facial expressions of E taking into account social and cultural rules. Our algorithm presented above (and illustrated in Figure 11) follows the flow illustrated in Figure 1. Being in the emotional state E our agent modifies its default expression of emotion E according to rules which are related to the type of interpersonal relations it has with the interlocutor.

8 Limitation of our work

This paper addresses the problem of expressive skills of an embodied agent. It results from the assumption that the communicative role of facial expressions in embodied agents has not been sufficiently explored so far. We tackled only a small part of this problem. A number of issues needs to be resolved in the future.

First of all, our model of complex facial expressions is not complete. We did not implement all features (deception clues) that are used to distinguish between different types of facial expressions e.g. time-related deception clues such as synchronisation or the clues related to the duration of apex, offset, and onset. Secondly, we have based our work on emotional expression view, we need to compare our model with models based on different theoretical assumptions. Scherer's componential theory leads to various expressions that are counterparts to our complex facial expressions. It will be interesting to compare our complex expressions with displays generated according to Scherer's theory and corresponding to the same internal state. The shortcomings of our method of similarity were discussed in section 5.4.4. We also need to define the fuzzy definitions of facial expressions from empirical data.

The model proposed in section 7 can be seen as a premise of a more complete model of facial behaviour management in interpersonal relations. It implements only a small subset of factors that influences facial behaviour. We use facial displays to express the relations with an interlocutor which is only a small part of a social context.

In the future we plan to consider inter-cultural differences, other types of facial expressions (like suppression or exaggeration), as well as other factors which influence the facial behaviour in interpersonal relations. So far, for sake of simplicity, we have considered neither the personality of the displayer, the circumstances of interaction (see section 2) nor the features of the emotional state of the displayer other than its valence (e.g. we have not considered the intensity or the dominance value). For instance, in our model, as sadness and anger have the same valence, the expression of sadness is processed in the same way as the expression of anger, even if anger is concealed in different situations than sadness. Moreover we do not consider the problem of the frequency of applications of our rules. In our model the default expression will be modified for any co-occurrence of a particular positive/negative expression and certain social circumstances. It has the tendency to generate behaviours that are too deterministic and which are not dynamically adaptable to the social context.

9 Conclusion

In this paper we describe an agent that adapts its facial expressions to interpersonal relations. The agent is able to mask, hide or even simulate the expression of emotions. Its facial expressions reflect the management of the display of its emotions. First, we introduce an algorithm to generate different types of facial expressions that we called complex facial expressions. Using the algorithm presented in section 6 we can generate superposed, masked, inhibited or fake expressions. We also introduce an innovative method of comparison between any two facial expressions based on fuzzy similarity. Our model was applied with success in the EmoTv project (Abrilian et al. (2005)) that uses the "copy-synthesis" methodology to study and generate multimodal behaviour. Our model of complex facial expressions can be used to express interpersonal relations. From the annotation of the SEMMEL video-corpus we construct a set of rules of facial behaviour management for our agent. Consequently depending on the values of social distance, social power, and valence of an emotional state our agent adapts its facial expressions accordingly. It uses certain complex facial expressions instead of spontaneous expressions of emotions.

We believe that our model of facial behaviour management has many applications. Agents can express their attitude, friendliness or authority. We believe that ECAs could take different social roles related to the activities they carry out. For instance, being a guide or a tutor an ECA can be more dominant, while being a virtual salesman it should respect the social distance with a new customer. Other applications of ECAs require close relations with the user (e.g., virtual friend or companion). Different relations between an ECA and a user can be expressed by appropriate patterns of facial behaviour. For example, an agent whose aim is to become a virtual companion cannot be emotionally suppressive if it wants to gain the sympathy of the user (as suppression can be perceived as a manifestation of hostility and may inhibit the development of the new relationships, Butler et al. (2003)). Having the same emotional state a virtual salesman would display different facial expressions than a virtual teacher. Last but not least, nowadays most ECAs applications are mainly used in short-term interactions. With our model, long-term relations can be created in which values of social distance and social power can dynamically change and this change will be reflected onto the agent's behaviour.

Acknowledgement We are very grateful to Giulianella Coletti and to Andrea Capotorti for their help on fuzzy methods, to Jean-Claude Martin, Stephanie Buisine and their collaborators for the evaluation of our model of complex facial expressions and Elisabeth André and Matthias Rehm for letting us use the SEMMEL video-corpus. We also thank Elisabetta Bevacqua and Maurizio Mancini for implementing the Greta system. Part of this research was supported by the EU FP6 Network of Excellence HUMAINE and by the EU FP6 Integrated Project CALLAS.

References

Abrilian, S., Devillers, L., Buisine, S., Martin, J.-C., 2005. EmoTV1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In: 11th International Conference on Human-Computer Interaction (HCII'2005). Las Vegas, Nevada, USA.

Albrecht, I., Schröder, M., Haber, J., Seidel, H., 2005. Mixed feelings: expression of non-basic emotions in a muscle-based talking head. Virtual Reality 8 (4), 201–212.

Bartneck, C., Reichenbach, J., 2005. Subtle emotional expressions of synthetic characters. International Journal Human-Computer Studies 62 (3), 179–192.

Bassili, J., 1979. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality and Social Psychology 37 (11), 2049–2058.

Bevacqua, E., Mancini, M., Niewiadomski, R., Pelachaud, C., 2007. An expressive ECA showing complex emotions. In: Proceedings of the AISB Annual Convention. Newcastle, UK, pp. 208–216.

Bouchon-Meunier, B., Rifqi, M., Bothorel, S., 1996. Towards general measures of comparison of objects. Fuzzy sets and systems 84 (2), 143–153.

Brown, B., Levinson, S., 1987. Politeness: Some universals on language usage. Cambridge University Press.

Buck, R., Losow, J., Murphy, M., Costanzo, P., 1992. Social facilitation and inhibition of emotional expression and communication. Journal of Personality and Social Psychology 63 (6), 962–968.

Bui, T. D., 2004. Creating emotions and facial expressions for embodied agents. Ph.D. thesis, University of Twente, Departament of Computer Science.

Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L., Pelachaud, C., 2006. Perception of blended emotions: From video corpus to expressive agent. In: The 6th International Conference on Intelligent Virtual Agents. Marina del Rey, pp. 93–106.

Butler, E., Egloff, B., Wilhelm, F., Smith, N., Erickson, E., Gross, J., 2003. The social consequences of expressive suppression. Emotion 3 (1), 48–67.

Constantini, E., Pianesi, F., Prete, M., 2005. Recognizing emotions in human and synthetic faces: The role of the upper and lower parts of the face. In: Proceedings of the 10th International Conference on Intelligent User Interfaces. San Diego, California, USA, pp. 20–27.

Darwin, C., 1872. The expression of the emotions in man and animals. John Murray.

De Carolis, B., Pelachaud, C., Poggi, I., de Rosis, F., 2001. Behavior planning for a reflexive agent. In: Proceedings of IJCAI 2001. Oporto, Portugal, pp. 1059–1066.

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

Ekman, P., 1985. Telling lies: Clues to deceit in the marketplace, politics, and marriage. W. W. Norton & Company.

Ekman, P., 2003a. Darwin, deception, and facial expression. Ann. N.Y. Acad. Sci. 1000, 205–221.

Ekman, P., 2003b. The Face Revealed. Weidenfeld & Nicolson, London.

Ekman, P., Friesen, W., 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., 1982. Felt, false, miserable smiles. Journal of Nonverbal Behavior 6 (4), 238–251.

Elliott, C., 1997. I picked up catapia and other stories: A multimodal approach to expressivity for 'emotionally intelligent' agents. In: Proceedings of the First International Conference on Autonomous Agents. Marina del Rey, CA, pp. 451–457.

Etcoff, N., Magee, J., 1992. Categorical perception of facial expressions. Cognition 44 (3), 227–240.

France, M. L., Hecht, M., 2005. Option or obligation to smile: The effects of power and gender and facial expression. In: Philippot, P., Feldman, R., Coats, E. (Eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press, pp. 45–70.

Frank, M., Ekman, P., Friesen, W., 1995. Behavioral markers and recognizability of the smile of enjoyment. In: Ekman, P., Rosenberg, E. (Eds.), What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, pp. 217–238.

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

Goleman, D., 1995. Emotional intelligence. Bantam Books, New York.

Gosselin, P., Kirouac, G., Doré, F., 1995. Components and recognition of facial expression in the communication of emotion by actors. In: Ekman, P., Rosenberg, E. (Eds.), What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, pp. 243–267.

Izard, C., 1977. Human emotion. New York: Plenum Press.

Kuncheva, L., 2000. How good are fuzzy if-then classifiers? IEEE Transactions on Systems, Man and Cybernetics 30 (4), 501–509.

Manstead, A., Fischer, A., Jakobs, E., 2005. The social and emotional functions of facial displays. In: Philippot, P., Feldman, R., Coats, E. (Eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press, pp. 287–316.

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

Matsumoto, D., Ekman, P., 2004. The relationship among expressions, labels, and descriptions of contempt. Journal of Personality and Social Psychology 87 (4), 529–540.

Niewiadomski, R., 2007. A model of complex facial expressions in interpersonal relations for animated agents. Ph.D. thesis, University of Perugia.

Niewiadomski, R., Pelachaud, C., 2007. Fuzzy similarity of facial expressions of embodied agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (Eds.), Proceedings of the 7th International Conference on Intelligent Virtual Agents. Springer, pp. 86–98.

Ochs, M., Niewiadomski, R., Pelachaud, C., Sadek, D., 2005. Intelligent expressions of emotions. In: Tao, J., Tan, T., Picard, R. W. (Eds.), First International Conference on Affective Computing and Intelligent Interaction (ACII). Springer, Pekin, China, pp. 707–714.

Ostermann, J., 2002. Face animation in MPEG-4. In: Pandzic, I., Forchheimer, R. (Eds.), MPEG-4 Facial Animation - The Standard Implementation and Applications. Wiley, England, pp. 17–55.

Pentland, A., 2004. Social dynamics: Signals and behavior. IEEE International Conference on Developmental Learning (ICDL), 63–70.

Plutchik, R., 1980. Emotions: A psychoevolutionary Synthesis. Harper & Row, New York.

Poggi, I., June 2005. Interacting bodies and interacting minds. In: 2nd International Society for Gesture Studies (ISGS) Conference Interacting Bodies. Lyon, France.

Prendinger, H., Ishizuka, M., 2001. Social role awareness in animated agents. In: Proceedings of the fifth international conference on Autonomous agents. Montreal, Quebec, Canada, pp. 270–277.

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. Informing the design of embodied conversational agents by analysing multimodal politeness behaviors in human-human communication. In: Workshop on Conversational Informatics for Supporting Social Intelligence and Interaction.

Rifqi, M., 1996. Mesures de comparaison, typicalité et classification d'objets flous : théorie et pratique. Ph.D. thesis, Université de Paris 6, France.

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.

Saarni, C., Weber, H., 2005. Emotional displays and dissemblance in childhood: Implications for self-presentation. In: Philippot, P., Feldman, R., Coats, E. (Eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press, pp. 71–108.

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, pp. 92–119.

Tsapatsoulis, N., Raouzaiou, A., Kollias, S., Crowie, R., 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 & Sons, pp. 141–168.

Vinciarelli, A., Salamin, H., Pantic, M., 2009. Social signal processing: Understanding social interaction through nonverbal behavior analysis. In: Proceedings of the International Workshop on Computer Vision and Pattern Recognition for Human Behavior. pp. 42–49.

Wagner, H., Lee, V., 2005. Facial behavior alone and in the presence of others. In: Philippot, P., Feldman, R., Coats, E. (Eds.), The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction). Cambridge University Press, Cambridge University Press, pp. 262–286.

Wagner, H., MacDonald, C., Manstead, A., 1986. Communication of individual emotions by spontaneous facial expressions. Journal of Personality and Social Psychology 50 (4), 737–743.

Wagner, H., Smith, J., 1991. Facial expression in the presence of friends and strangers. Journal of Nonverbal Behavior 15 (4), 201–214.

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, United States, pp. 85 – 91.

Wehrle, T., Kaiser, S., Schmidt, S., Scherer, K., 2000. Studying the dynamics of emotional expression via synthesized facial muscle movements. Journal of Personality and Social Psychology 78 (1), 105–119.

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, pp. 113–131.

Wiggins, J., Trapnell, P., Phillips, N., 1988. Psychometric and geometric characteristics of the revised interpersonal adjective scales (IAS-R). Multivariate Behavioral Research 23 (3), 517–530.

Young, A., Rowland, D., Calder, A., Etcoff, N., Seth, A., Perrett, D., 1997. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition 63 (3), 271–313.