An expressive ECA showing complex emotions

Elisabetta Bevacqua[1]
Maurizio Mancini[1]
Radoslaw Niewiadomski[2]
Catherine Pelachaud[1]

Abstract. Embodied Conversational Agents (ECAs) are a new paradigm of computer interface with a human-like aspect that allow users to interact with the machine through natural speech, gestures, facial expressions, and gaze. In this paper we present an head animation system for our ECA Greta and we focus on two of its aspects: the expressivity of movement and the computation of complex facial expressions. The system synchronises the nonverbal behaviours of the agent with the verbal stream of her speech; moreover it allows us to qualitatively modify the animation of the agent, that is to add expressivity to the agent's movements. Our model of facial expressions embeds not only the expressions of the set of basic emotions (e.g., anger, sadness, fear) but also different types of complex expressions like fake, inhibited, and masked expressions.

1 Introduction

Embodied Conversational Agents (ECAs) are a new paradigm of computer interface with a human-like aspect that are being used in an increasing number of applications for their ability to convey complex information through verbal and nonverbal behaviours like voice, intonation, gaze, gesture, facial expressions, etc. Their capabilities are useful in scenarios such as a presenter on the web, a pedagogical agent in tutoring systems, a companion in interactive settings in public places such as museums, or even a character in virtual storytelling systems. Our system provides control over the animation of a virtual agent head. It computes realistic behavior for the head movement (nods, shakes, direction changes, etc), gaze (looking at the interlocutor, looking away) and facial expression (performing actions like raising eyebrows, showing an emotion, closing eyelids, and so on). During conversation the agent moves her head according to what she is saying. Moreover eye movements are computed depending on the gaze intention. Since eyes and head are physically linked these two communicative modalities cannot be computed separately, so our system exhibits head and eye coordination to obtain a realistic gaze behaviour.

In this paper we present an ECA animation system, called Greta, focusing on two of its aspects: the expressivity of movement and the computation of complex facial expressions.

The *expressivity* of behaviour is "How" the information is communicated through the execution of some physical behaviour. Expressivity is an integral part of the communication process as it can provide information on the state of an agent, his current emotional state, mood, and personality [47]. Section 3 gives an overview of our head animation system architecture while Section 4 explains the implementation of the expressive animation computation.

There is a large amount of evidence in psychological research that human's repertoire of facial expressions is very large [13, 14, 34]. We call *complex facial expressions* the expressions that are different from the spontaneous facial displays of simple emotional states (e.g. display of anger or sadness). They can be displays of some combinations of emotions as well as expressions of emotions which are modified according to some social rules. It was shown [17, 24] that an expressed emotion does not always reveal a felt emotion. People may, for example, decide not to express the emotion they feel because of some socio-cultural norms called *display rules* [14]. When display rules are applied, a set of procedures of emotional displays management [42] is used. These procedures leads to different facial expressions [15].
It was proved that these facial expressions can be distinguished by humans (i.e. there are different facial signals) [16, 19] and have different role and meaning [14, 34]. This is why we have introduced the Complex Facial Expression Computation module which is detailed in Section 5. In section 2 we discuss some of the previous works on ECAs focusing on gaze, head and facial expression models. Then we give a detailed explanation of our head animation system in sections 3, 4 and 5. Finally we conclude the paper in section 6.

2 State of the art

Overviews of recent ECA implementations have been described by Cassell et al. and Prendinger et al. [8, 38]. K. R. Thórisson developed a multi-layer multimodal architecture able to generate the animation of the virtual 2D agent 'Gandalf' during a conversation with a user [43]. Gandalf has been created to communicate with users also through head movements (nods) and gaze direction. 'Rea' [7] is a humanoid agent able to understand the user's behaviour and respond with appropriate speech, facial expressions, gaze, gestures and head movements.

A number of studies have underlined the importance of gaze and head behaviour in the communication process. Vertegaal et al. [45] found that gaze is an excellent predictor of conversational attention in multiparty situations and placed special consideration on eye contact in the design of video conference systems [46]. Peters et al. [31] proposed a model of attention and interest using gaze behaviour, defining the capabilities an ECA requires to be capable of starting, maintaining and ending a conversation. Head movements hold an impor-

[1] University of Paris8, IUT de Montreuil, 140 rue de la Nuovelle France, 93100, Montreuil, France,
email: e.bevacqua,m.mancini,c.pelachaud@iut.univ-paris8.fr
[2] Università degli Studi di Perugia, Dipartimento di Matematica e Informatica, Via Vanvitelli 1, 06123, Perugia, Italy,
email: radek@dipmat.unipg.it

tant role in conversation and researches have been done to determine their pattern in order to enrich ECAs with more believable head animation. Heylen analyzed head patterns to define their properties and functions [21] useful to implement ECAs behaviour.

In all of these systems the final animation is obtained by interpolating between pre-determined body and facial configurations. One of the novelty of our system is that the agent movements can be qualitatively modified (changing their amplitude, speed, fluidity, etc) applying some parameters to add expressivity to the ECA.

Most of animated agents are able to display a small number of emotions (e.g., [3, 10, 26, 43]). Only few works implement models of mixed emotional expressions. The existing solutions usually compute new expressions in which single parameters are obtained by "averaging" the values of the corresponding parameters of expressions of certain "basic" emotions. Among others, the model called "Emotion Disc" [41] uses bi-linear interpolation between two closest 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. In Tsapatsoulis et al. [44] two different approaches are used: the new expression can be derived from basic one by "scaling" it. The second approach uses interpolation between facial parameters values of two closest basic emotions. A similar model of facial expressions was proposed by Albrecht et al. [1].

Different approach was proposed by Duy Bui [5]. She introduced the set of fuzzy rules to determine the blending expressions of six basic emotions. In this approach a set of fuzzy rules is attributed to each pair of emotions. The fuzzy inference determines the degrees of muscles contractions of the final expression in function of the input emotions intensities. Blending expressions of six basic emotions are also used in [23].

Different types of facial expressions were considered by Rehm and André [39]. by In a study on deceptive agents, they show that users are able to differentiate between the agent displaying an expression of felt emotion versus an expression of fake emotion [39]. Prendinger et al. [37] implement a set of procedures called *social filter programs*. In a consequence their agent is able to modulate the intensity of the expression according to the social context.

Comparing with other models we introduce the diversification of facial expressions in relation to their meaning, role, and appearance. Thus, another novelty of our system is that our agent is able to express different types of facial expressions (like inhibited, masked or fake expressions). Moreover, following the psychological evidence [15] complex facial expressions are computed by composing whole facial areas of any facial expression. Thus the final expression is combination of facial areas of other expressions. Finally we can create complex facial expressions not only in a case of six basic emotions but for any expression that was described by the researchers (e.g., embarrassment [22] or contempt [13]).

3 The Greta head animation system

Greta is an Embodied Conversational Agent (ECA) that communicates through her face and gestures while talking to the user. The head animation system, topic of this paper, is a process that computes the low-level animation of the agent head. For example it has to precisely determine which horizontal angle the head should rotate in order to perform a head-shake, or to determine which facial points have to be moved to show a particular facial expression. Figure 1 shows the general architecture of the system. The input data of the system

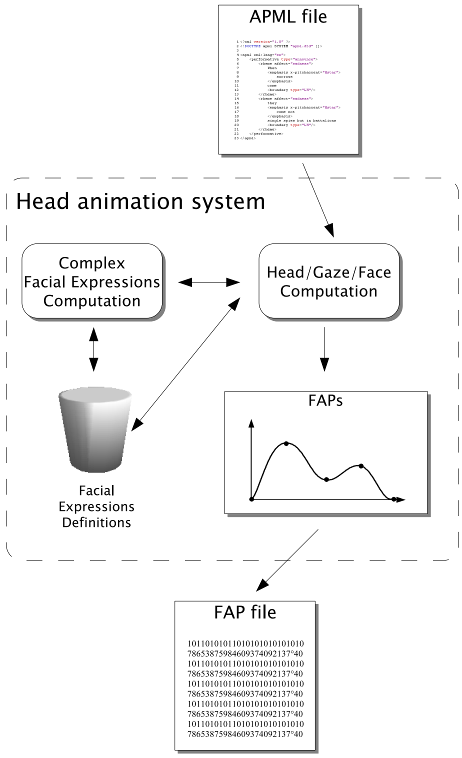

Figure 1. Low-level representation of the Greta's face engine.

is a file with an high-level description of communicative acts that the agent aims to communicate. The input file follows the format of the Affective Presentation Markup Language APML [33] (see Figure 2 for an example of an APML input file). APML is an XML-based language whose tags represent communicative acts. In the example of Figure 2 the APML tags surrounding the text specify that the agent is going to *announce* something (line 5) while showing a *sad* emotional face (lines 6 and 14). The APML tags give information about the speaker's goals of conversation. That is, the enclosed sentences could be translated into a facial expression and/or head movements and/or gaze change [35]. The animation corresponding to APML tags is computed by the *Head/Gaze/Face Computation* module, explained in detail in Section 4. In some cases, for some values of the *affect* tag (for instance a complex emotion), this module yields the generation of the facial expression to the *Complex Facial Expressions Computation* module, described in detail in Section 5.

The output of the system is an animation file, that is a sequence of frames, and a wav file. In particular, our system produces an animation file following the MPEG4/FAP format [29, 32]. The standard defines some activation points on the agent's face, called FAPs, and the way each FAP contributes to the deformation of the face area underneath it. A FAP file is a sequence of FAP frames, one frame for each time unit, and each FAP frame is a sequence of FAP values. Since this is a standard format, every talking head player implementing FAPs can playback the animation files generated by our engine.

```
<?xml version="1.0" ?>
<!DOCTYPE apml SYSTEM "apml.dtd" []>

<apml xml:lang="en">
    <performative type="announce">
        <rheme affect="sadness">
            When
            <emphasis x-pitchaccent="Hstar">
                sorrows
            </emphasis>
            come
            <boundary type="LH"/>
        </rheme>
        <rheme affect="sadness">
            they
            <emphasis x-pitchaccent="Hstar">
                come not
            </emphasis>
            single spies but in battalions
            <boundary type="LH"/>
        </rheme>
    </performative>
</apml>
```

Figure 2. Example of an APML input file.

4 Expressive computation of head/gaze/face

4.1 Expressivity

Many researchers (Wallbott and Scherer [47], Gallaher [18], Ball and Breese [2], Pollick [36]) have investigated human motion characteristics and encoded them into categories. Some authors refer to body motion using dual categories such as slow/fast, small/expansive, weak/energetic, small/large, unpleasant/pleasant. The expressivity of behaviour is "How" the information is communicated through the execution of some physical behaviour.

Greta is an expressive ECA, that is her animation can be qualitatively modified by a set of expressivity parameters affecting the physical characteristics of movements (like speed, width, strength, etc.). Starting from the results reported in [47] and [18], we have defined the expressivity by 6 dimensions:

- *Overall Activity* models the general amount of activity (e.g., passive/static or animated/engaged);
- *Spatial Extent* modifies the amplitude of movements (e.g., expanded versus contracted);
- *Temporal Extent* changes the duration of movements (e.g., quick versus sustained actions);
- *Fluidity* influences the smoothness and continuity of movement (e.g., smooth, graceful versus sudden, jerky);
- *Power* represents the dynamic properties of the movement (e.g., weak/relaxed versus strong/tense);
- *Repetitivity* models the tendency of the agent to replicate the same movement with short and close repetitions during time. Technical details on the implementation of these parameters can be found in [20].

Let us describe how each part of the *Head/Gaze/Face Computation* (see Figure 1) works.

4.2 Head model

The head model generates the animation of the head: a single movement corresponds to a change in head direction (up, down, left, etc.) while a composed movement is obtained by the repetition of a single movement (as in the case of head nod and shake). The quality of the head movement can be modified by varying the expressivity parameters, for example by increasing the *Spatial Extent* Greta's head movement will be wider. Variation in the *Temporal Extent* parameter changes the rotation speed: the smaller is such expressivity parameter the smaller is the rotation angle of the head. *Repetitivity* can cause one or more repetitions of the same movement; for example, it will increase the frequency of head nods/shakes.

Our agent follows the standard MPEG-4/FAP, so a head position is given by specifying the value of 3 FAPs, one for each axis, through a rotation vector:

$$RV = (HRx, HRy, HRz).$$

We define RV_{RP} the rotation vector that moves the head back to its *reference position*. A head movement HM is described by a sequence of keyframes where each keyframe is a couple (T, RV) containing a time label T and the rotation vector RV that specifies the head position at time T:

$$HM = ((T_0, RV_{RP}), (T_1, RV_1), ..., (T_{n-1}, RV_{n-1}), (T_n, RV_{RP})).$$

By default, a head movement starts and ends with the *reference position*, that is the first and last key frame correspond to the head position RV_{RP}. When two successive movements are computed we check if the first head movement needs to coarticulate into the next head movement or if it has time to go back to its reference position. The decision is based on the duration between successive head movements. If two head movements are too close to each other, the key frames to the *reference position* are deleted to avoid unnatural jerky movement. Let us consider two consecutive head movements:

$$HM_1 = ((T_{1_0}, RV_{RP}), (T_{1_1}, RV_{1_1}), (T_{1_2}, RV_{1_2}), (T_{1_3}, RV_{RP})),$$

$$HM_2 = ((T_{2_0}, RV_{RP}), (T_{2_1}, RV_{2_1}), (T_{2_2}, RV_{2_2}), (T_{2_3}, RV_{RP})).$$

For sake of simplicity, both movements perform rotations only around the x axis. Figure 3(a) shows the curve of the FAP HRx representing both movements HM_1 and HM_2. We calculate their temporal distance TD as:

$$TD = T_{2_1} - T_{1_2}.$$

If such a temporal distance is less than a given threshold, we consider both movements to be too close to each other and, in order to avoid jerky movements of the head, we delete the last key frame in HM_1 and the first key frame in HM_2 to obtain a smoother curve and then a better animation of the head. The new curve is shown in Figure 3(b). As explained in before the head movements can be modulated by the value of the expressivity parameters affecting the amplitude of their movement, as well as their speed and acceleration. Once all the key frames have been calculated they are interpolated to obtain the whole head movement. Further computation may be necessary to ensure correlation between head and eye movement (see Section 4.3.1).

4.3 Gaze model

The gaze model generates the animation of the eyes. It is based on statistical data obtained from the annotation of behaviour (smile, gaze direction, speaking turn, etc.) of dyads [30].

A belief network, embedded both types of information, is used to compute the next gaze direction. Personalized gaze behaviour is obtained by specifying temporal parameters of the belief network. Maximal and minimal time for mutual gaze, look at the other participant

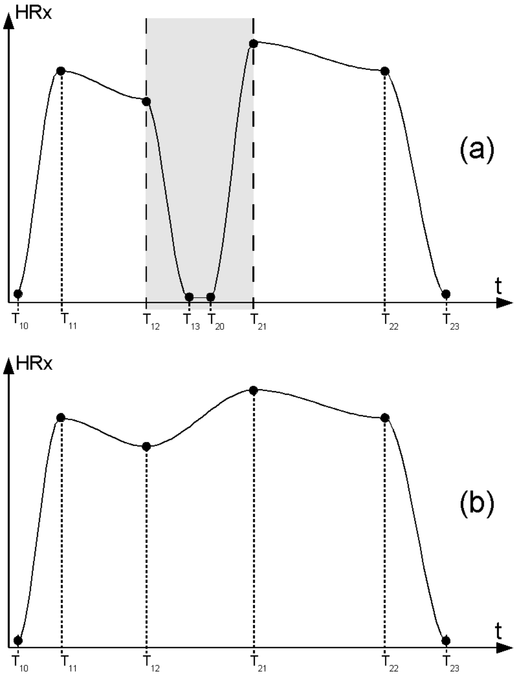

Figure 3. (a) Curves of two very close head rotations around axis x. The grey area shows the jerk in the head movement. (b) Key frames in T_{1_3} and in T_{2_0} have been deleted to obtain a smoother animation.

and gaze away can be specified. This model computes the agent's gaze pattern as a temporal sequence of two possible states: *LookAt* and *LookAway*. *LookAt* means that the ECA gazes at the other participant (the user or an other agent in the virtual environment), whereas *LookAway* implies that the agent moves away her gaze. The result of the gaze model is a sequence of couples:

$$GAZE = ((t_0, S_0)...(t_n, S_n)),$$

where t_i and S_i are respectively the start time and the value of the i^{th} state ($S_i = 1$ means *LookAt* whereas $S_i = 0$ means *LookAway*). The gaze state *LookAt* corresponds to a precis direction while the gaze state *LookAway*is defined as negation of *LookAt*. In our algorithm the space is divided into 8 regions related to the user's head (up, up right, down, down left, etc.). Some communicative functions specifies the gaze should be direct to one of these regions; if no specification exists a region is chosen casually. Once a region is determined the exact eye direction is computed randomly. To ensure spatial coherency (the eyes do not move in every direction during a *LookAway*) a region is fixed for a certain duration.

4.3.1 Correlation between head and gaze movements

The result of the gaze model could be inconsistent with the animation of the head. Such inconsistency shows up when the directions of the head and of the gaze are too different causing an unnatural rotation of the eyes in the skull. Figure 4 shows such inconsistency. In Figure 4(a) the gaze of the agent is away (look down) and the head is down. The expression of sadness generates this gaze/head pattern. Figure 4(b) shows the next frame where the head is still down but the direction of the eyes changes because of a *LookAt*. Since the rotation of the head was quite wide, the iris of the eyes is no more visible

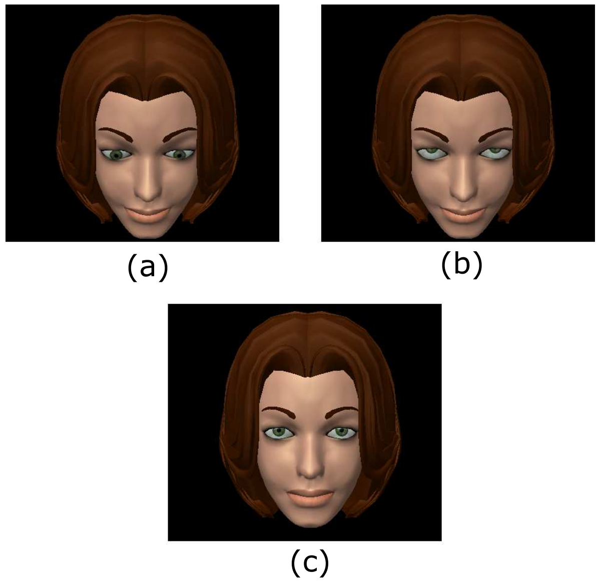

Figure 4. Example of an inconsistency between head and gaze. (a) Frame 1: head down and gaze away. (b) Frame 2: the head is still down but the eyes must perform a *LookAt* disappearing in the skull: inconsistency. (c) New Frame 2: the inconsistency is deleted forcing a rotation of the head.

creating an awkward animation. To remove all the inconsistencies between the gaze and the head movement we analyse the sequence $GAZE$ (deriving from the gaze model) and for each couple (t_i, S_i) we check the validity of the head position for each frame in the interval of time $[t_i, t_{i+1}]$, where t_{i+1} is the start time of the $(i+1)^{th}$ element of the sequence. A head position $RV = (HRx, HRy, HRz)$ (see Section 4.2) is *valid* if:

$$-th_x < HRx < th_x,$$

$$-th_y < HRy < th_y,$$

$$-th_z < HRz < th_z,$$

where th_x, th_y and th_z are respectively the threshold of the rotation around the axes x, y and z. When a *not-valid* position is found, the nearer key frames are modified (moved nearer to the *reference position*) and the interpolation recomputed to generate the new animation of the head. Figure 4(c) shows the same frame in Figure 4(b) where the inconsistency between the gaze and the head has been deleted. As we can see the head position has changed to allow the eyes to reach the direction defined by the *LookAt* and remain visible.

4.4 Face model

Depending on APML tags, the face model decides which facial expressions have to be performed by the agent. As explained in the introduction, a facial expression can be either a simple or a complex one. Simple expressions are directly retrieved from a static definition library (the *Facial Expressions Definitions* object in Figure 1). On the other hand, complex expressions are dynamically calculated by the *Complex Facial Expressions Computation* module which is presented in detail in section 5. In both cases, the simple or complex expressions are converted into a sequence of FAP values that are inserted into a data structure and will be interpolated afterwards.

As we explained before, our agent follows the standard MPEG-4/FAP, so a facial expression is specified by the value of the FAPs on the face. The first step to compute a facial animation is to define a sequence of keyframes. A keyframe is defined as a couple (T, FS) containing a time label T and facial shape FS that specifies the values of the FAPs of the face at time T. By default, each facial expression starts and ends with the *neutral expression* and it is characterized by four temporal parameters [25]:

- *attack*: is the time that, starting from the neutral face $FS_{neutral}$, the expression takes to reach its maximal intensity FS_1;
- *decay*: is the time during which the intensity of the expression lightly de-creases, usually to reach a stable value FS_2;
- *sustain*: is the time during which the expression is maintained, usually it represents the more visible part of the expression;
- *release*: is the time that the expression takes to return to the neutral expression $FS_{neutral}$.

A keyframe is computed for each temporal parameter and so, a facial expression animation FA can be defined as follows:

$$FA = ((T_{attack}, FS_1), (T_{decay}, FS_2),$$
$$(T_{sustain}, FS_2), (T_{release}, FS_{neutral})).$$

The final animation is obtained by interpolating between the resulting keyframes. Like for the head, when two consecutive facial expressions are computed we need to check their temporal distance. If such a distance is less than a given threshold, it means that the facial expressions are too close to each other and we need to delete the last keyframe of the first expression and the first keyframe of the second expression in order to avoid an abrupt return to the neutral face in between.

The facial animation depends also on the expressivity parameters. While computing the keyframes, the FAP values are modified according to the parameters. For example *Spatial extent* scales the FAP values of the expression; that is it changes the amplitude of the displacement of FAPs on the agent's face. *Temporal extent* increases (resp. decreases) the speed by which the expression appears: low (resp. high) values will make the expressions appear faster (resp. slower).

5 Complex Facial Expressions Computation

Our model of complex facial expressions is based on Paul Ekman's results [12–15]. We model complex facial expressions using a face partitioning approach. It means that different emotions are expressed on different areas of the face. More precisely, 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 etc.). Then the complex facial expressions are composed of the facial areas of input expressions.
While analysing human facial expressions of emotions, Ekman distinguished between: modulating, falsifying, and qualifying an expression [15]. One modulates expressions by *de-intensifying* or *intensifying* them. For example, to intensify an expression one can change the intensity or duration of the expression. Falsifying a facial expression means to *simulate* it (to show a fake emotion), *neutralize* it (to show neutral face) or *mask* it. Masking occurs when one tries to hide "as much as possible" an expression by simulating another one. Finally, *qualification* means to add a fake expression (usually a smile) to a real one in order to express combination of both. In this case, the felt expression is not inhibited.
Using the model presented in this section we can generate the facial expressions of masking, as well as fake and inhibited expressions. The model generates different displays for these different types of expression. Complex facial expressions are obtained from the six basic emotions: anger, disgust, fear, joy, sadness, and surprise are described in the literature [13, 15]. Basing on it we have defined for each type of expression a set of fuzzy rules that describe its characteristic features in terms of facial areas. Each rule correspond to one basic emotion.
In the case of an input expression for which the complex facial expression is not defined explicitly by our rules (e.g. expression of contempt or disappointment) our algorithm chooses the most appropriate solution. This appropriateness is measured by analysing *visual resemblance* between expressions. For this purpose we introduced an innovative approach to compare two facial expressions. It is based on the notion of fuzzy similarity. In our approach any facial expression is described by a set of fuzzy sets. The main advantage of this approach is that slightly different expressions can be described by one label (like "joy" or "sadness"). Our algorithm compares two facial expressions attribute-after-attribute and then it composes single results into one value in the interval [0,1]. Finally, the values of similarity and the rules mentioned above are used to generate the complex facial expressions. Let us present in detail of our model.

5.1 Comparing Two Facial Expressions

The first step of the algorithm consists in establishing the degree of similarity between the input expression (i.e. the expression for which we want to find the complex facial expression) and the expressions of basic emotions. Let E_u and E_w be two emotions whose expressions we want to compare. Thus we want to establish the degree of similarity between *Exp(E_w)* and *Exp(E_u)*. In our approach each expression *Exp(E_i)* is associated with a set of fuzzy sets in terms of which all plausible expressions of emotion E_i are defined. That is, for each numerical parameter (FAP) of an expression of emotion E_i there is a fuzzy set that specifies a range of plausible values. Firstly, the value of similarity for each parameter (FAP) of *Exp(E_w)* and *Exp(E_u)* is established independently. The M-measure of resemblance *S*:

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

where *A* and *B* are two fuzzy sets [4] is used in this case. Finally all values are combined by means of *Ordered Weighted Averaging (OWA)* operator (see [40] for detailed discussion).

5.2 Rules For Creation of Complex Facial Expressions

Several researchers have proposed a list of *deception clues* i.e. the features of expressions that are useful in distinguishing between fake and felt expressions [11, 12, 15]. At the moment, two of them are implemented in our model: *reliable features* and the *inhibition hypothesis*.
First of all humans are not able to control all their facial muscles. In a consequence expressions of felt emotions may be associated with specific facial features like: sadness brows [15] or orbicularis oculi activity in the case of joy [12]. Such *reliable features* lack in

fake expressions as they are difficult to do voluntarily. For each basic emotion the features which are missing in fake expressions are known [12, 15].
On the other hand, people are not able to fully inhibit felt emotions. According to the *inhibition hypothesis* [12], 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. Finally, Ekman enumerates all facial areas that leak over the mask during the emotional displays management [15].
For each type of deception clues considered by us a separate set of rules has been developed. The first one - SFR_{fake} - describes the features of a fake expression, while SFR_{felt} - of a felt one.
In a case of the SFR_{fake} the meaning of each rule is as follows: the more the input expression of E_i is similar to the expression of E_u, the more possible is that facial areas of $\text{Exp}(\text{E}_i)$ corresponding to reliable features of $\text{Exp}(\text{E}_u)$ should not be used in the final expression. For example, in the case of sadness the following rule is applied: "the more the input expression is similar to sadness, the more possible is that the brows of the input expression should not be visible". Similarly, each rule of SFR_{felt} describes the features which occur even in a covered expression of a felt emotion.

5.3 Generation of Complex Facial Expressions

Using our model different types of expression can be generated. Let us present the process of generation of a complex facial expression on the example of masking. Masking occurs when a felt emotion should not be displayed for some reason; it is preferred to display a different emotional expression. The expression of masking is composed from a fake expression that covers the expression of the real emotional state. Thus, both sets of rules SFR_{felt} and of SFR_{fake} should be applied in this case.
Let *B* be the set of the basic emotions (including neutral state) and $\text{Exp}(\text{E}_u)$ be the expression corresponding to one of these emotions, $\text{E}_u \in B$.

In the case of masking the input to the system consists in specifying two emotion labels: the felt one E_i and the fake E_j. Both, E_i and E_j are specified in the APML input file.
In the *first step* our algorithm establishes the degrees of similarity between $\text{Exp}(\text{E}_i)$, $\text{Exp}(\text{E}_j)$ and all expressions of emotions that belongs to the set *B*. In a consequence we obtain two vectors $[\text{a}_k]$ and $[\text{b}_k]$, $1 \leq$ a,b $\leq$ —B—, a_k, $\text{b}_k \in [0, 1]$ of the degrees of similarity.
In the *second step* the deception clues for input expressions $\text{Exp}(\text{E}_i)$, $\text{Exp}(\text{E}_j)$ are established. For this purpose the sets of rules SFR_{felt} and SFR_{fake} are used. The vector $[\text{a}_k]$ of felt expression E_i is processed by SFR_{felt}, while the vector $[\text{b}_k]$ of the fake expression E_j is processed by SFR_{fake}. The SFR_{felt} and SFR_{fake} returns certain predictions about which parts of the face will (not) be visible in the masking expression.
The fake and felt parts of the final expression are considered separately. Finally, in the *last step* of the algorithm, for each facial area, the results of SFR_{felt} and of SFR_{fake} are composed in order to obtain the final expression. It is realized using another set of rules that takes as an input the outputs of precedent systems. The crisp output indicates the part of which expression (felt, fake or neutral) will be used in the final expression. The main task of this system is to resolve the eventual conflicts (i.e. the situation in which according to results of SFR_{fake} and SFR_{felt} different expressions should be shown in the same facial region). At the contrary, in the case in which neither felt nor fake emotion can be shown in a particular region of the face, the neutral expression is used instead.

Figure 5 presents the agent displaying the expression of a disappointment, that is masked by fake happiness. In the image on the right the parts of expression copied from the expression of disappointment are marked with blue and of happiness with red circles. We can notice that the absence of *orbicularis oculi* activity as indicator of fake happiness is visible on both images. Also the movement of brows can be observed, which is characteristic of disappointment. It is so because the expression of disappointment is very similar (according to the procedure described in section 5.1) to the expression of sadness. The facial areas F_1(forehead and brows) and F_2 (upper eyelids) cover the features of felt sadness that leak over the mask. As a consequence, they can be observed in inhibited sadness and thus they can be also observed in covered disappointment.

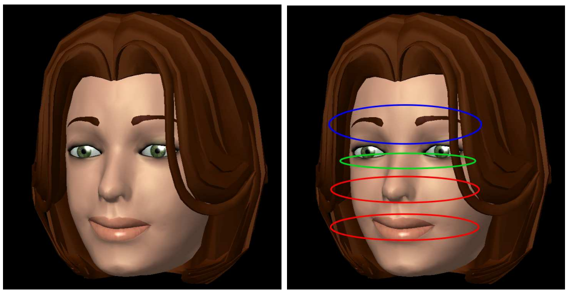

Figure 5. Example of a disappointment masked by a joy.

Similarly we can generate different complex facial expressions. Figure 6 shows two other examples of our algorithm's output. In the first row one can see: on the left the expression of contempt; on the right the same expression is inhibited. In the second row the expression of sadness is presented on the left, while the fake expression of sadness - on the right.

5.4 Evaluation

Complex facial expressions generated with our model were evaluated in a study based on the "copy-synthesis" method [6]. According to this approach the human behaviour is analysed and annotated by means of a high-level annotation schema. The animation of the agent is then obtained from the annotation. In one of such studies that is called EmoTV [9] different types of complex facial expressions were observed.
We generated a set of animations starting from two videos of the EmoTv video-corpus [9] that were annotated with different types of complex facial expressions. More precisely, four different animations were compared with each original video. The first two animations used simple facial expressions and body movements. Each of them displayed one of the two emotions indicated by the annotators. The two other animations used complex facial expressions that were created in two different ways: in the first one we used our model; in the second one the low-level annotation was used instead.
Then we evaluated the quality of the animations by asking subjects to compare them with the original videos.

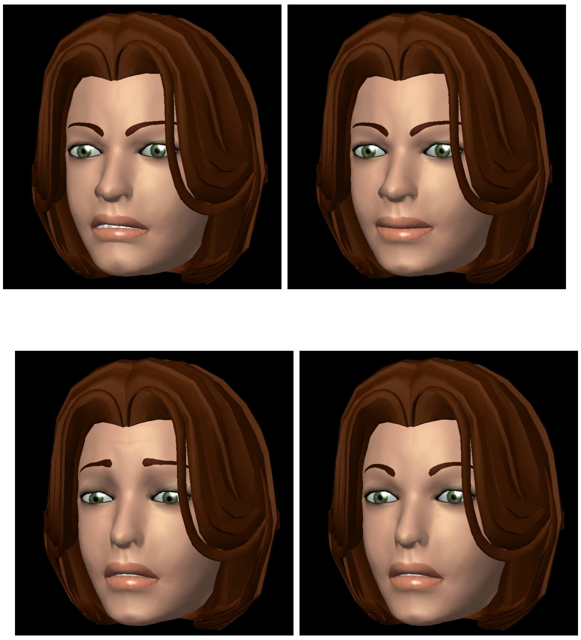

Figure 6. Examples of inhibited contempt (first row) and simulated sadness (second row).

The results are promising (see [6] for detailed results): The use of complex facial expressions created by our model has influenced the evaluation score significantly, especially in the case of animation in which facial expressions were easily observed. Animations created with our model obtained a satisfactory result in comparison with manually created animations of complex expressions. In one case (expression of masking) automatically generated expressions were evaluated even better than the manually defined complex expressions. In the second test the result was slightly worse, particularly in the no audio condition.

In another experiment we used different types of complex facial expressions in order to express different interpersonal relations between interlocutors (see [27] for details). We found different complex expressions generated using our model are recognized by humans and that these expressions comunicate different social signals [27].

6 Conclusions and Future

We have presented an expressive head animation system for ECAs. After giving a general overview of the system, we have focused on the implementation of two important aspects of behaviour: the expressivity of movement and the computation of complex facial expressions. Our head/gaze/face model generates facial expressions and coordinated head and gaze movements under the influence of some expressivity parameters.
Then we have described a model to compute complex facial expressions. Our model introduces the diversification of facial expressions. It builds different types of complex facial expressions. As a consequence, these different types of complex facial expressions can be distinguished by the user, because their appearance is different.
In the near future we are going to develop the head/gaze model to make the ECA pointing at objects in the environment with gaze. We will also integrate this model in a speaker/listener system for ECAs. We also plan to model other types of complex facial expressions and to implement other deception clues like *micro-expressions* and *time-related* deception clues. At the moment all expressions (basic and complex ones) are specified in the APML file. We aim at integrating our system with an Elicited-Emotion module which is responsible for the evaluation of an event and the emotion elicitation (see [28]).

REFERENCES

[1] I. Albrecht, M. Schroder, J. Haber, and H. Seidel, ‘Mixed feelings: Expression of non-basic emo-tions in a muscle-based talking head’, *Special issue of Journal of Virtual Reality on ”Language, Speech & Gesture”*, (2005).

[2] G. Ball and J. Breese, ‘Emotion and personality in a conversational agent’, in *Embodied Conversational Characters*, eds., S. Prevost J. Cassell, J. Sullivan and E. Churchill, MITpress, Cambridge, MA, (2000).

[3] C. Becker, S. Kopp, and I. Wachsmuth, ‘Simulating the emotion dynamics of a multimodal conversational agent’, in *Affective Dialogue Systems*, eds., E. Andr, L. Dybkjr, W. Minker, and P. Heisterkamp, 154–165, Springer Verlag, (2004).

[4] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel, ‘Towards general measures of comparison of objects’, *Fuzzy sets and systems*, **84**(2), 143153, (1996).

[5] T. Duy Bui, *Creating Emotions And Facial Expressions For Embodied Agents, PhD thesis*, University of Twente, Departament of Computer Science, Neslia Paniculata, Enschede, 2004.

[6] S. Buisine, S. Abrilian, R. Niewiadomski, J.-C. Martin, L. Devillers, and C. Pelachaud, ‘Perception of blended emotions: From video corpus to expressive agent’, in *The 6th International Conference on Intelligent Virtual Agents*, Marina del Rey, USA, (August 2006).

[7] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjlmsson, and H. Yan., ‘Embodiment in conversational interfaces: Rea’, in *Conference on Human Factors in Computing Systems*, Pittsburgh, PA, (April 15-20 1999).

[8] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, *Embodied Conversational Agents*, MIT Press, Cambridge, MA, 2000.

[9] L. Devillers, S. Abrilian, and J.-C. Martin, ‘Representing real life emotions in audiovisual data with non basic emotional patterns and context features’, in *Proceedings of First International Conference on Affective Computing & Intelligent Interaction*, pp. 519–526, Pekin, China, (2005).

[10] A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann, ‘Imparting individuality to virtual humans’, in *First International Workshop on Virtual Reality Rehabilitation*, Lausanne, Switzerland, (2002).

[11] P. Ekman, *Non i volti della menzogna: gli indizi dell’inganno nei rapporti interpersonali, negli affari, nella politica, nei tribunali*, Giunti-Barbera, 1989.

[12] P. Ekman, ‘Darwin, deception, and facial expression’, *Annals of the New York Academy of Sciences*, **1000**, 205–221, (2003).

[13] P. Ekman, *The Face Revealed*, Weidenfeld & Nicolson, London, 2003.

[14] P. Ekman and W.V. Friesen, ‘The repertoire of nonverbal behavior’s: Categories, origins, usage and coding’, *Semiotica*, **1**, 49–98, (1969).

[15] P. Ekman and W.V. Friesen, *Unmasking the Face. A guide to recognizing emotions from facial clues*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[16] M.G. Frank, P. Ekman, and W.V. Friesen, ‘Behavioral markers and recognizability of the smile of enjoyment’, in *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, eds., P. Ekman and E.L. Rosenberg, Oxford University Press, (1995).

[17] W.V. Friesen, *Cultural differences in facial expressions in a social situation: An experimental test of the concept of display rules*, University of California, 1972. Unpublished doctoral dissertation.

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

[19] P. Gosselin, G. Kirouac, and F.Y. Dor, ‘Components and recognition of facial expression in the communication of emotion by actors’, in *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression*

Using the Facial Action Coding System (FACS), eds., P. Ekman and E.L. Rosenberg, 243–267, Oxford University Press, (1995).

[20] B. Hartmann, M. Mancini, and C. Pelachaud, 'Implementing expressive gesture synthesis for embodied conversational agents', in *The 6th International Workshop on Gesture in Human-Computer Interaction and Simulation*, VALORIA, University of Bretagne Sud, France, (2005).

[21] D. Heylen, 'Challenges ahead. head movements and other social acts in conversation', in *AISB 2005 - Social Presence Cues Symposium*, (2005).

[22] D. Keltner, 'Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame', *Journal of Personality and Social Psychology*, **68**, 441 – 454, (1992).

[23] J. Ktsyri, V. Klucharev, M. Frydrych, and M. Sams, 'Identification of synthetic and natural emotional facial expressions', in *ISCA Tutorial and Research Workshop on Audio Visual Speech Processing*, pp. 239–244, St. Jorioz, France, (2003).

[24] M. LaFrance and M. A. Hecht, 'Option or obligation to smile: The effects of power and gender and facial expression', in *The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction)*, Cambridge University Press, (2005).

[25] M. Mancini, R. Bresin, and C. Pelachaud, 'From acoustic cues to an expressive agent', in *Gesture Workshop*, pp. 280–291, (2005).

[26] D.W. Massaro, M.M. Cohen, J. Beskow, S. Daniel, and R.A. Cole, 'Developing and evaluating conversational agents', in *First Workshop on Embodied Conversational Characters*, Lake Tahoe, CA, (1998).

[27] R. Niewiadomski, *A model of complex facial expressions in interpersonal relations for animated agents*, Ph.D. dissertation, University of Perugia, 2007.

[28] M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek, 'Intelligent expressions of emotions', in *Proceedings of First International Conference on Affective Computing & Intelligent Interaction*, Pekin, China, (2005).

[29] C. Pelachaud, 'Visual text-to-speech', in *MPEG4 Facial Animation - The standard, implementations and applications*, eds., Igor S. Pandzic and Robert Forcheimer, John Wiley & Sons, (2002).

[30] C. Pelachaud and M. Bilvi, 'Modelling gaze behavior for conversational agents', in *International Working Conference on Intelligent Virtual Agents*, Germany, (September 15-17 2003).

[31] C. Pelachaud, C. Peters, M. Mancini, E. Bevacqua, and I. Poggi, 'A model of attention and interest using gaze behavior', in *International Working Conference on Intelligent Virtual Agents*, Greece, (2005).

[32] E. Petajan, 'Facial animation coding, unofficial derivative of MPEG-4 standardization, work-in-progress', Technical report, Human Animation Working Group, VRML Consortium, (1997).

[33] I. Poggi, 'Mind markers', in *Gestures. Meaning and use.*, ed., N. Trigo M. Rector, I. Poggi, University Fernando Pessoa Press, Oporto, Portugal, (2003).

[34] I. Poggi, 'Interacting bodies and interacting minds', in *International Society for Gesture Studies - Interacting Bodies*, (2005).

[35] I. Poggi and C. Pelachaud, 'Performative facial expressions in animated faces', 155–188, (2000).

[36] F. E. Pollick, 'The features people use to recognize human movement style', in *Gesture-Based Communication in Human-Computer Interaction*, eds., Antonio Camurri and Gualtiero Volpe, number 2915 in LNAI, 10–19, Springer, (2004).

[37] H. Prendinger and M. Ishizuka, 'Social role awareness in animated agents', in *Proceedings of the fifth international conference on Autonomous agents*, pp. 270–277, Montreal, Quebec, Canada, (2001).

[38] *Life-Like Characters*, eds., H. Prendinger and M. Ishizuka, Cognitive Technologies, Springer, 2004.

[39] M. Rehm and E. Andr, 'Catch me if you can - exploring lying agents in social settings', in *International Conference on Autonomous Agents and Multiagent Systems*, pp. 937–944, (2005).

[40] M. Rifqi, *Mesures de comparaison, typicalit et classification d'objets flous : thorie et pratique*, Thse, 1996.

[41] Z. Ruttkay, H. Noot, and P. ten Hagen, 'Emotion disc and emotion squares: tools to explore the facial expression face', *Computer Graphics Forum*, **22**, 49–53, (March 2003).

[42] C. Saarni and Hannelore Weber, 'Emotional displays and dissemblance in childhood: Implications for self presentation', in *The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction)*, Cambridge University Press, (2005).

[43] K. R. Thrisson, *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*, Ph.D. dissertation, MIT Media Laboratory, 1996.

[44] N. Tsapatsoulis, A. Raouzaiou, S. Kollias, R. Crowie, and E. Douglas-Cowie, 'Emotion recognition and synthesis based on mpeg-4 faps', in *MPEG4 Facial Animation - The standard, implementations and applications*, eds., Igor S. Pandzic and Robert Forcheimer, John Wiley & Sons, (2002).

[45] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, 'Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes', in *Conference on Human factors in computing systems*, pp. 301–308, New York, NY, USA, (2001). ACM Press.

[46] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, 'Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction', in *Conference on Human factors in computing systems*, pp. 521–528, New York, NY, USA, (2003). ACM Press.

[47] H. G. Wallbott and K. R. Scherer, 'Cues and channels in emotion recognition', *Journal of Personality and Social Psychology*, **51**(4), 690–699, (1986).