Communicating emotional states with the Greta agent

R. Niewiadomski[1], M. Mancini[2], S. Hyniewska[1], C. Pelachaud[1]

CNRS / Télécom ParisTech
37 rue Dareau
75014, Paris - France

InfoMus Lab, DIST – University of Genova
via Causa 13
I-16145, Genova - Italy

Chapter to be published in Scherer, K.R., Bänziger, T., & Roesch, E. (Eds.) A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing. Oxford: Oxford University Press, in press

Summary:

Recent technological progress made the creation of a humanoid interface to computer systems possible. An embodied conversational agent (ECA) is a computer-generated animated character that is able to carry on natural, human-like communication with users. For this purpose agent systems have been developed to simulate verbal and nonverbal communicative behaviours. ECAs can dialog with users using synthesised speech, gestures, gaze and facial expressions.

In this Chapter we present our embodied conversational agent called Greta and focus on its capabilities of generating emotional expressive behaviours. Our 3D agent is able to communicate using the verbal and nonverbal channels like gaze, head and torso movements, facial expressions, and gestures. It follows the SAIBA framework that defines functionalities and communication protocols for ECA systems. The system generates the output animation in MPEG-4 standard. Our system is optimised to be used in interactive applications. It has a rich repertoire of expressive emotional behaviours.

Several studies demonstrate that the application of facial expressions in embodied agents is justified. The agents that use emotional displays are perceived as more engaging (Walker et al. 1994), increase the user's interest level and influence the perception of the agent as the same ECA with different facial expressions is perceived as more credible and trustworthy (Rehm and André 2005). No wonder many models of emotional expressive behaviours for an ECA were proposed (see chapter by Hyniewska *et al.*). In this Chapter we present the algorithms that enrich the emotional expressive behaviours of the Greta agent. First we describe an algorithm that extends a set of facial expressions in the repertoire of the ECA, introducing complex facial expressions like superposition or masking. Complex expressions are composed of the facial areas of its constituent expressions but they can be distinguished from them and they can signal different meanings. There were used in our agent to show socially acceptable behaviours (Niewiadomski and Pelachaud 2007) or to display empathy (Niewiadomski et al. 2008). Then we show how our agent may display emotional states using different modalities. These expressions are composed of different signals partially ordered in time and belonging to different nonverbal communicative channels. Multimodal expressions are particularly useful to communicate by an ECA subtle emotions like pride or relief (Niewiadomski et al. 2009) that are difficult to signal using only facial channel.

In the second part of this chapter we describe also how full-body expressive behaviours are modulated in our agent using a small set of parameters. These parameters may express certain constant characteristics of the agent as well as modulate its single behaviours in particular emotional states. In the latter case they may modify any nonverbal behaviour such as multimodal or complex expressions of emotion. We show also how to extract the values of these parameters automatically from visual data.

Introduction

In recent years there has been a growing interest in developing embodied agents with expressive capabilities has. This is motivated by the effort to enhance human-machine interaction. To be able to express internal states, intentions or motivations our agent needs to be endowed by nonverbal communication skills and to access data on how to communicate in a human-like manner.

The nonverbal emotional behaviours of our agent can be conveyed through: the choice of the nonverbal signals and their realisation. Although humans communicate through several modalities at the same time, it is the face that is a privileged place for the expression and the decoding of emotions, as suggested by interdisciplinary theorists (e.g. Kaiser and Wehrle 2001; Ekman 1972). Facial expressions do not always correspond to felt emotions: they can be fake (showing an expression of an unfelt emotion), masked (masking a felt emotion by an unfelt emotion), superposed (showing a mixed of 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). 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. In this chapter we describe a model of complex facial expressions which is based on partitioning approach proposed by Ekman (Ekman and Friesen 1975; Ekman 2003a; Ekman 2003b). With this algorithm our agent is able display superposed, masked, fake and inhibited expressions of emotions. The novelty of our system comparing to the solutions presented in Chapter by Hyniewska *et al.* 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 facial areas of facial expressions; that is the final expression is a combination of facial areas of input expressions.

Nevertheless, some studies show that emotions are expressed also through several other modalities, such as for example body movements (Wallbott 1998; Pollick et al. 2001). Some observational studies have explored the complexity of emotional expressions in terms of their dynamics (Keltner 1995; Shiota, et al. 2003; Harrigan and O'Connell 1996). Among others, Keltner (1995) studied the sequence of facial and body movement in embarrassment. The typical expression of embarrassment starts from gaze downing or gaze shifts which are followed by “controlled” smiles. These smiles are often accompanied by a pressure of the lips. At the end of the expression the movement of the head to the left was often observed as well as some face touching gestures (Keltner 1995). We called these expressions *multimodal sequential expressions* of emotions. They might be composed of nonverbal behaviours displayed over different modalities that change dynamically over time. Our agent is able to display *multimodal sequential expressions.* For this prupose we developed a language for the formal description of such expressions from real data and of an algorithm that uses this description to automatically generate emotional displays. In comparison to the exisitng solutions (see Chapter by Hyniewska *et al.*) this algorithm system generates automatically a variety of multimodal emotional expressions. It is based on a high-level symbolic description of nonverbal behaviours. Contrary to many other approaches which use captured data for behaviour reproduction, in this approach the observed behaviours are interpreted by a human who creates constraints. The sequences of nonverbal displays are independent behaviours that are not driven by the spoken text. The system allows for the synthesis of any number of emotional states and is not restricted by the number of modalities. While it uses a discrete approach in its use of labels to refer to emotions, it is also linked to the componential approach by the underlined importance of the sequence of difference elements in the perception of emotions. Its advantage is to generate a variety of animations for one emotional label avoiding the repetitiveness in the behavior of a virtual character.

What is more, any expression can be performed slowly or quickly, jerkily or fluently etc. With the term *expressivity* we identify the ways the single nonverbal behaviours are synthesised. They are the external, visible qualities of a movement, like its speed, amplitude, fluidity and so on. The *expressivity of behaviour* is an important aspect of the nonverbal behaviour. Expressivity is an integral part of the communication process as it can provide information on the emotional state, mood and personality of a person (Wallbott 1998). The nonverbal behaviour in Greta may be modulated by the use of 6 expressivity parameters. Moreover the expressivity parameters in Greta are related to the current emotional state as well as to the characteristics of an overall individuality of the agent.

The remainder of this Chapter is organised as follows. In the next Section we present a general overview of the architecture of our agent. Then in Section 3 we describe our model of complex facial expressions while in Section 4 the algorithm that generates multimodal sequential expressions in Greta is presented. In Section 5 we explain how we model the expressivity of nonverbal behaviours. We conclude this Chapter in Section 6.

GRETA the interactive embodied conversational agent

This Section presents an overview of the architecture of our embodied conversational agent. Our agent's architecture follows the design methodology proposed by Thórisson et al. (2005) and is compatible with the SAIBA framework (Vilhjálmsson et al. 2007). The architecture is modular and distributed. Each module exchanges information and data through a central message system. We use the concept of a whiteboard (Thórisson et al. 2005) that allows internal modules and an external software to be integrated easily. The system is designed to be used in interactive applications working in real-time.

Generic architecture for embodied conversational agent

SAIBA (Vilhjálmsson et al. 2007) is an international research initiative which main aim is to define a standard framework for the generation of virtual agent behaviour. It defines a number of levels of abstraction, from the computation of the agent's communicative intention, to behaviour planning and realisation. The **Intent Planner** module decides of the agent's current goals, emotional state and beliefs, and encodes them into the Function Markup Language (FML) (Heylen et al. 2008). To convey the agent's communicative intentions, the **Behaviour Planner** module schedules a number of communicative signals (e.g., speech, facial expressions, gestures) which are encoded with the Behaviour Markup Language (BML). It specifies the verbal and nonverbal behaviours of ECAs (Vilhjálmsson et al. 2007). Finally the task of the third element of the SAIBA framework, **Behaviour Realizer**, is to realise the behaviours scheduled by the Behaviour Planner. It receives input in the BML format and it generates the animation. There exist several implementations of the SAIBA standard, among others SmartBody (Thiebaux et al. 2008, Lee et al. 2006), BMLRealizer (Árnason and Þorsteinsson 2008), RealActor (Cerekovic et al. 2009) and EMBR (Heloir and Kipp 2009).

Our general purpose embodied conversational agent system is almost a full implementation of the SAIBA framework offering solution for the Behaviour Planner and the Behaviour Realizer layers and a partial implementation of the Intent Planner for the listener. In our implementation of the SAIBA standard we focus on the Behaviour Planner rather than on the Behaviour Realizer as the other implementations do. Our Behaviour Planner generates a large set of nonverbal behaviours from a single communicative intention. We also provide an advanced Intent Planner for the listener that includes reactive, cognitive as well as mimicry backchannels.

Architecture

Figure 1 illustrates the architecture of our agent. It follows the SAIBA framework and is composed of 3 main modules. In SAIBA the Intent Planner is dedicated to the speaker. To be able to control a Listener agent, we have introduced the **Listener Intent Planner**. This module generates automatically the communicative intentions of the listener. On the other hand, in the current state of our system, the intentions of the speaker are defined manually in an FML-APML input file. In the future they will be generated by the Speaker Intent Planner. The **Behaviour Planner** module receives as input the agent's communicative intentions, be a speaker or a listener, written in FML-APML and generates as output a list of signals in the BML language. These signals are sent to the **Behaviour Realizer** that generates the MPEG-4 FAP and BAP frames. Finally, the animation is played in the **FAP-BAP Player**. All modules are synchronized by the Central Clock and communicate with each other through the Psyclone whiteboard (Thórisson 2005).

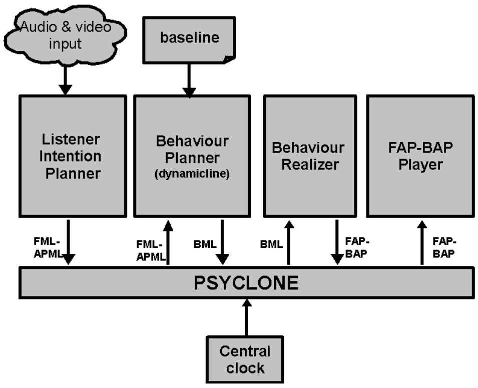

Figure 1. The architecture of the Greta agent.

Listener Intent Planner. The Listener Intent Planner module is in charge of the computation of the agent's behaviours while being a listener when conversing with a user (or another virtual agent). This component encompasses three modules called reactive backchannel, cognitive backchannel, and mimicry.

The *reactive backchannel module* prompts a backchannel signal when our system recognises certain speaker's behaviours; for example, a head nod or a variation in the pitch of the user's voice will trigger a backchannel with a certain probability. On the other hand, the *cognitive backchannel module* computes when and which backchannel should be displayed using information about the agent's beliefs towards the speaker's speech. The cognitive module selects which signals to display from the lexicon depending on the agent's reaction towards the speaker's speech. The third module is the *mimicry module*. This module determines when and which signals the agent is to mimic.

Behaviour Planner. The Behaviour Planner takes as input both the agent's communicative intentions specified by the FML-APML language and some agent's characteristics (i.e. *baseline*). The main task of this component is to select, for each communicative intention, the adequate set of behaviours to display. The output of the Behaviour Planner is described in the BML language. It contains the sequence of behaviours with their timing information to be displayed by our virtual agent.

Behaviour Realizer. This module generates the animation of our agent following the MPEG-4 format (Ostermann 2002). The input of the module is specified by the BML language. It contains the text to be spoken and/or a set of nonverbal signals to be displayed. Facial expressions, gaze, gestures, torso movements are described symbolically in the repository files. Each BML tag is instantiated as a set of key-frames that are then smoothly interpolated. The Behaviour Realizer synchronizes the behaviours across modalities. It solves also eventual conflicts between the signals that use the same modality. The speech is generated by an external TTS and the lips movements are added to the animation.

When the Behaviour Realizer receives no input, the agent does not remain still. It generates some idle movements. Periodically a piece of animation is computed and is sent to the FAP-BAP Player. This avoids unnatural “freezing” of the agent.

FAP-BAP Player. The FAP-BAP Player receives the animation generated by the Behaviour Realizer and plays it in a graphic window. The player is MPEG-4 compliant. Facial and body configurations are respectively described through FAP and BAP frames.

Synchronization. The synchronisation of all modules in the distributed environment is ensured by the Central Clock which broadcasts regularly timestamps through the whiteboard. All other components are registered in the whiteboard to receive timestamps.

Model of complex facial expressions

We base our model of complex facial expressions on Paul Ekman's studies (Ekman and Friesen 1975; Ekman 2003a; Ekman 2003b). We define complex facial expressions using a face partitioning approach. Each facial expression is defined by a set of eight facial areas F_k (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 a composition of these facial areas, each of which can display signs of emotion. For complex facial expressions, different emotions can be expressed on different areas of the face; e.g., when sadness is masked by happiness, sadness 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 emotion to different areas of the face. For this purpose we define for each type of complex facial expressions a set of rules that describe the composition of the facial areas. These rules, based on the description proposed by Ekman, refer to six emotions, namely: anger, disgust, fear, joy, sadness, and surprise. For an input emotion for which the facial expression is not defined explicitly by our rules (e.g., expression of contempt or disappointment) our algorithm chooses the most appropriate solution. For this purpose we use an algorithm (Niewiadomski and Pelachaud 2007) based on fuzzy similarity. In this approach each facial expression is described by fuzzy sets. Fuzzy similarity is used to compute the degree of visual similarity between them. Our algorithm compares any two facial expressions and produces an index of similarity in the interval [0..1], where 0 means “not similar at all" while 1 means identical expressions. Once the most similar expression, chosen among the 6 is known the algorithm applies the corresponding rules to the input expression. The rules determine which elements of the input expression are used in the complex facial expression. (More details about the algorithm and its evaluation can be found in Niewiadomski and Pelachaud (2007, 2007b) and Buisine et al. (2006).

For different types of complex facial expressions: superposition, masking, fake or inhibited expressions can be generated with our algorithm. For example, the expression of superposition happens when two emotions are felt at the same time. Figure 2 shows an example of the superposition expression computed by our model. Figures 2a and 2b show the expressions of anger and sadness respectively. Figures 2c and 2d show the superposition as a composition of face areas of both input expressions. In the Figure 2d we see which parts of the face correspond to sadness and which ones to anger. In that image the areas F_5, F_6, F_7, and F_8 (expressing sadness) are marked out with by the light grey circles while areas F_1, F_2, F_3, and F_4 (expressing anger) by a dark grey colour.

a b c d

Figure 2. Superposition of anger and sadness. From the left to right: anger a), sadness b), superposition of anger and sadness c) superposition of anger and sadness with significant areas marked d).

Figure 3 shows the agent displaying the expression of disappointment masked by a fake happiness. We applied our similarity algorithm and found that disappointment has a facial expression very similar to 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. On the other hand, the mouth area displays a smile (sign of happiness).

a b c d

Figure 3. 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).

Sequential expressions of emotions

From the analysis of expressions of emotions like embarrassment (Keltner 1995), awe, amusement, pride (Shiota, et al. 2003) or anxiety (Harrigan and O'Connell 1996) (see Chapter by Hyniewska *et al.*) it was shown that certain emotions are expressed by a set of signals which are arranged in a certain interval of time rather than by a static facial

expression (i.e. an expression at its apex). The expressions of emotional states are dynamic and they can be displayed over different modalities like face, gaze and head movement, gestures, or even posture. Interestingly, these signals do not have to occur simultaneously (Keltner 1995).
To go beyond agents showing simply static facial expressions of emotions, we have defined a representation scheme that encompasses the dynamics of facial expressions of an emotion. The main task of our algorithm is to generate multimodal sequential expressions of emotions, i.e., expressions that are composed of different signals partially ordered in time and with the use of different nonverbal communicative channels. These multimodal expressions can be of any duration while the duration of constitutive signals is limited, e.g., facial expressions of emotions usually are not longer than four seconds (Ekman and Friesen 1975) and gestures often have at least a minimum duration. We define for each emotional state a *behaviour set*, a set of signals through which the emotion is displayed, and a *constraint set* that defines logical and temporal relations between the signals in the behaviour set. These two sets are defined from literature and from manual annotation conducted on a video-corpus (Niewiadomski et al. 2009).

The single signals like a *frown*, *head nod* or *self-touch gesture* are described in the repositories of the agent's nonverbal behaviours and are grouped in the behaviour sets. Each behaviour set associates one emotional state with a set of plausible signals that might be displayed by the agent. For each signal in a behaviour set one may define the probability of occurrence, the minimum and maximum duration and the number of repetitions. All the signals that belong to one behaviour set may occur in the displays of the emotion associated with it, but their occurrence is not random. We developed an XML-like language to describe a set of relations, that is, the constraint set, between the signals of one behaviour set. Two types of constraints are possible:

- *temporal constraints, that* define arithmetic relations on the start time and end time of a signal, e.g., “signal s_i cannot start at the beginning of animation“ or “signal s_i starts immediately after the signal s_j finishes“. The temporal constraints are defined using arithmetical relations: <, >, =;
- *appearance constraints, that* describe more general relations between signals like inclusion or exclusion, e.g., “signals s_i and s_j cannot co-occur“ or “signal s_j cannot occur without signal s_i“.

The constraints of both types can be composed using the logical operators: *and, or, not*.

From the single label of an emotional state *e* (e.g., anger or embarrassment) our system generates a sequence of multimodal expressions, i.e., the animation *A* of a given duration *t* composed of a sequence of signals $\{s_{i(j)}\}$ on different modalities. It chooses a coherent subset of signals from the behaviour set BS_e, computes their durations, and their display order. Thus, at time-step t_i (1..n-1, t_n=t) it chooses one signal s_j from the behaviour set BS_e; it decides its duration and checks if the constraints CS_e are satisfied against the partial animation A (0...t_{i-1}). Consequently, either it adds s_j to *A*, or it repeats the procedure with the different signal.
Our algorithm is a part of *Behaviour Planner* module of the Greta agent. It can generate several sequences of signals each of them satisfying the constraint sets. (More details

about the algorithm and its evaluation can be found in Niewiadomski *et al.*, (2009, 2009b).

In Figure 4 an example of the animation for the expression of embarrassment is shown. The images present the frames of the animation of Greta displaying respectively the signals: a) *look_right*, b) *head_down* and *gaze_down,* c) *gaze_left,* d) *gaze_left* and *non-Duchenne_smile,* e) *gaze_left.*

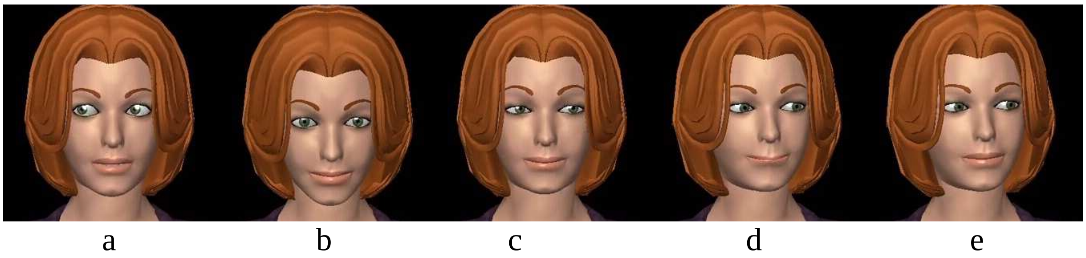

Figure 4. An example of an expression of embarrassment.

Figure 5 shows another example of animation of the same emotion. In Figure 5 the following signals are displayed: a) *neutral expression,* b) *smile,* c) *smile* and *gaze_right,* d) *gaze_left*, e) *gaze_down* and *head_down,* f) *face touching gesture.*

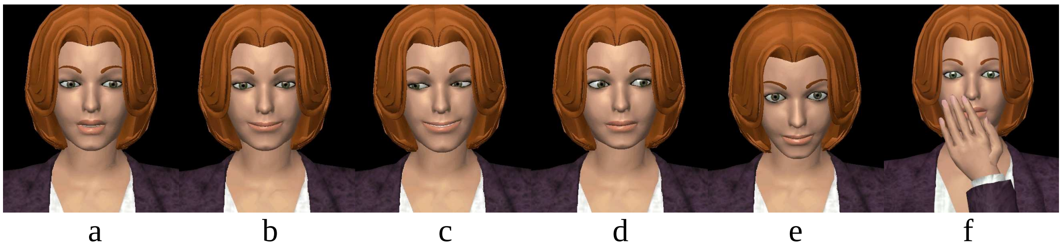

Figure 5. An example of multimodal expression of embarrassment.

Behaviour expressivity model

In this Section we describe a set of parameters that we use to represent the behaviour expressivity. We will show how these parameters may be used to define some global characteristics of the agent as well as how they modulate the synthesis of any emotional signal. Finally, we explain the process of capturing the expressivity values automatically from video corpora.

Expressivity parameters

In order to increase the life-likeness of the Greta ECA, Hartmann *et al.* have defined and implemented a set of parameters that allow one to alter the way the agent expresses its actual communicative intention (Hartmann et al. 2005):

- Overall Activity - OAC: amount of activity (e.g., passive/static versus animated/engaged). As this parameter increases (resp. decreases), the number of head movements, facial expressions, gestures and so on, increases (resp. decreases).
- Spatial Extent - SPC: amplitude of movements (e.g., expanded versus contracted). It determines the amplitude of, for example, head rotations and hand positions.
- Temporal Extent - TMP: duration of movements (e.g., quick versus sustained actions). This parameter modifies the speed of execution of movements. They are slow if the value of the parameter is negative, or fast when the parameter is positive.
- Fluidity - FLD: smoothness and continuity of movement (e.g., smooth, graceful versus sudden, jerky). Higher values allow smooth and continuous execution of movements while lower values create discontinuity in the movements.
- Power - PWR: dynamic properties of the movement (e.g., weak/relaxed versus strong/tense). Higher (resp. lower) values increase (resp. decrease) the acceleration of the head or limbs rotation, making the overall movement look more (resp. less) powerful.
- Repetitivity - REP: this parameter permits the generation of rhythmic repetitions of the same rotation/expression/gesture. For example, a head nod with a high repetitivity becomes a sequence consisting of very fast and small nods.

Baseline and Dynamicline

People vary in their behaviours, depending on their personality, the environmental situation, social rules and so on. Argyle (1988) and Gallaher (1992) state that there is an *underlying tendency* which is constantly present in each person's behaviour. People that tend to look more and perform wide gestures will continue to do so in most situations.
This is the idea we are capturing with the concept of a *Baseline* for ECAs. The Baseline represents the agent's general, underlying expressive behaviour tendencies. We define the Baseline as a set of expressivity parameters *Expr* that represent the agent's general behaviour expressivity, as explained in the previous Section: for example an agent could tend to perform slow and smooth gestures (low Temporal and high Fluidity) while another agent could tend to move in a fast and jerky manner (high Temporal and low Fluidity).

When communicating an emotional state or an intention, one may vary one's general tendencies, e.g., one can gesture in a greatly different way. For example a person that is always slow in performing hand gestures while talking may change her behaviour quality when being very angry at someone. One of the contributions of our work is the creation of a system in which the agent general expressive behaviour tendencies can be modulated depending on the agent's emotional state and communicative intention. We aim to model such influences as "an agent *A* with general expressive tendencies *T*, with an emotional state *E* tends to behave as T_E", where T_E is a particular "version" of *T* obtained by a modulation induced by *E*. In this way the agent's general behaviour tendencies *T* become *local* to the agent's actual emotional state. We embody the agent's local behaviour tendencies with the concept of a *Dynamicline*. The Dynamicline structure is the same as the Baseline, they differ by their meaning and by the way they are computed. The Dynamicline is a set of expressivity parameters *Modulated-Expr* that represent the agent local behaviour expressivity, that is, the Dynamicline is a set of parameters that derive both from the agent's Baseline and current communicative intention (e.g., its emotional state). It is the Dynamicline that represents the agent's current expressivity, the one that influences the execution of the agent behaviour. Figure 6 outlines how the Dynamicline is computed from the Baseline at runtime.

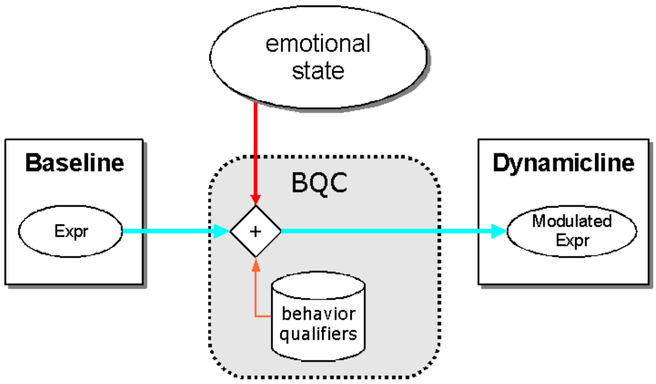

Figure 6. Behaviour Quality Computation: the agent's Baseline and communicative intention are used in the computation of the agent's Dynamicline.

This process is executed by the *Behaviour Quality Computation* (BQC) module which is a part of the *Behaviour Planner* module. The input of the BQC module is the agent's Baseline, which is a constant, and the emotional state the agent aims to communicate. Each time the emotional state varies the module computes a new Dynamicline for the agent. During the BQC process, the expressivity parameters contained in the Baseline are modulated depending on the agent's actual emotional state and the resulting expressivity values are stored in the Dynamicline.

This means that different emotional states have different impacts on the Dynamiclines of two agents having different Baselines. For example, if an agent has a general tendency (Baseline) to perform many movements with an average speed/amplitude then in a sad state it could simply stop moving, as the sad state reduces its activation parameter; on the other hand, an agent with a general tendency of moving a lot with fast

and large movements continues to make movements even when being sad, although with a lower frequency and with a lower expressivity (average speed and amplitude).
As shown in the Figure 6, the way in which the agent Baseline is modulated into its Dynamicline is determined by the agent *behaviour qualifiers* which are described in detail in the next Section.

Behaviour qualifiers

We call a *behaviour qualifier* a set of *modulations* that, given an emotional state, act on the behaviour tendencies of a conversational agent. By *modulation* we mean a variation over one of the expressive parameters contained in the agent's Baseline.
As an example, the following description can be modelled by a behaviour qualifier:

"a hot anger state (i) increases the degree of bodily activation and at the same time (ii) the speed, amplitude and energy of movements are very high".

We can notice that the behaviour qualifier described in the sentence above has two kinds of influences: a *relative* and an *absolute*. In the above example, part (i) of the sentence indicates that we may increase the degree of the bodily activation. This is a relative variation because we give an indication of the current behaviour tendency (Dynamicline) in terms of the general tendencies (Baseline). Instead, part (ii) of the sentence indicates that speed, amplitude and energy of movement should be very high: in this case we talk about absolute values, that is, the current behaviour tendencies (Dynamicline) are explicitly defined, and we do not refer to a general tendency (Baseline).
So, the behaviour qualifier in the example will act on the *Overall activation* (OAC) expressivity parameter of the agent's *torso, face* and *gesture* modalities by increasing them by a relative value as described in part (i) of the example sentence. It will also act on the *Temporal*, *Spatial* and *Power* expressivity parameters of the agent's gesture modality by assigning high absolute values to them, as described by part (ii) of the example sentence.

The task of defining behaviour qualifiers to allow us to perform BQC in ECAs is neither obvious nor simple. There is no certain data about the expressivity modulations induced by emotional states in humans. Only some experimental data is available, and we used the results reported in (Argyle 1988; Gallagher 1992; Wallbott and Scherer 1986) to define the behaviour qualifier of the Greta ECA.
For example several researchers observed that in a sad emotional state people tend to move slowly, perform few gestures and exhibit a contracted posture. Anger and joy induce wide and powerful movements with differences in fluidity: angry movements exhibit low fluidity while happy movements are smoother.

Determining Dynamicline automatically

In order to compute the agent's Dynamicline, Behaviour qualifiers need to be specified. This specification can be done manually following data from the literature (Argyle

1988; Gallagher 1992; Wallbott and Scherer 1986) or can be extracted automatically from visual data. In this Section we present a system that takes video data as input, extracts movement expressivity, and finally determines a Dynamicline for the Greta agent.

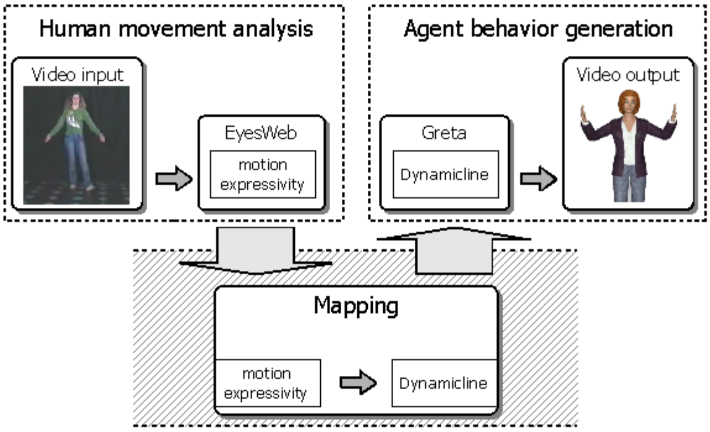

Figure 7. From video analysis to behaviour generation: system overview.

The EyesWeb system (www.eyesweb.org) (Camurri et al. 2004) is used for the analysis of human movement. A mapping between the extracted expressive parameters and the corresponding expressive parameters of the agent's Dynamicline is then established.

Figure 7 shows an overview of the system architecture and its different modules:

- *Human movement analysis*: we perform the automatic extraction of motion expressivity from a video source by using the EyesWeb platform and the EyesWeb Expressive Gesture Processing Library. More specifically, we extract the following motion expressive cues:
    - *Contraction Index* (CI): this is a measure, ranging from 0 to 1, of how contracted a movement or gesture is (i.e., performed near to the body) or expanded (i.e., performed with a use of the space surrounding the body). It is calculated by extracting the body *silhouette* from the background.
    - *Velocity*: we find the 2D coordinates of the actor's right or left hand in the image and from them we compute its velocity, taking into account its horizontal and vertical components.
    - *Acceleration*: by computing the derivative function of the velocity of the actor's hand we obtain its acceleration.
    - *Directness Index* (DI): this is a measure of how direct or flexible a trajectory between two points in space is. We obtain it by comparing the length of the minimal distance between two points with the length of the trajectory of the actor's right or left hand. We apply the DI to the coordinates of the hand to obtain an indicator of the fluidity of the movement.

- *Mapping*: the extracted motion expressive cues are mapped onto the agent's Dynamicline expressivity parameters and rescaled. We define the following (see Figure 8):
    - Contraction Index is mapped onto Spatial Extent, since they provide a measure of the amplitude of movements;
    - Velocity onto Temporal Extent, as they refer to the velocity of movements;
    - Acceleration onto Power, as both are indicators of the acceleration of movements;
    - Directness Index onto Fluidity, as they refer to the degree of the smoothness of movements.

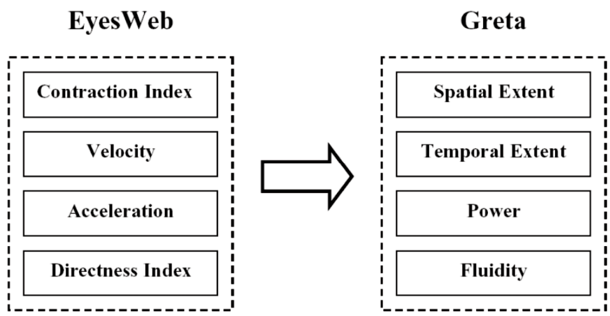

Figure 8. Mapping between video parameters extracted with EyesWeb and the Greta expressivity parameters.

- *Agent behaviour generation*: this module is part of the Greta agent animation system. It receives as input the agent Dynamicline and computes the animation of Greta.

This method allows us to extract qualitative information on behaviour execution in real-time. From these data we are able to elaborate behaviour qualifiers for different emotional states. When these qualifiers are integrated in the agent system, a Dynamicline can be computed. The system can model behaviour distinctiveness on the one hand and behaviour quality modulation on the other hand to convey specific messages.

Conclusion

In this chapter we have presented models of emotional expressive behaviours of our ECA. Each model captures an aspect of the complexity of emotion expressions. The algorithms described in this chapter aim at enriching the expressive capabilities of Greta by considering expressions that go the beyond stereotypical facial expressions in their apex. First of all, we presented the algorithm that generates superposed, masked, inhibited and fake expressions.

We also introduced to our ECA multimodal sequential expressions. They may be composed of nonverbal behaviours displayed over different modalities, of a sequence of behaviours, or of expressions within one modality that change dynamically.
Finally we have also presented an algorithm that modulates the behaviour expressivity of agents. Expressivity is an important cue of emotional expressions. It can also be used to define the general characteristics of the agent's behaviours. We have presented a system where the expressivity parameters are captured automatically from real human behaviour.

In the future we will continue the research on multimodal emotional signals. Our work will focus on the synchronisation between different modalities (including the verbal channel) as well as on the perception of emotions from different modalities. We also plan to work also on full-body emotional expressions e.g. including the postures. We would like to improve our expressivity model to be more related to the emotional state of the agent. This can be achieved through perceptive studies.

ACKNOWLEDGMENTS

We are very thankful to Björn Hartmann for defining Greta's set of expressivity parameters, and for implementing them in the Greta's gesture generation engine. We are very grateful to Ginevra Castellano for her help. The works presented in this chapter have been supported by the EU funded Human-Machine Interaction Network on Emotion Network of Excellence (HUMAINE: http://emotion-research.net/) and by the IP-CALLAS project IST-034800 (http://www.callas-newmedia.eu).

References

Argyle, M. (1988). Bodily Communication. Methuen & Co., London 2nd edition.

Árnason, B.P. and Porsteinsson, A. (2008). The CADIA BML Realizer, http://cadia.ru.is/projects/bmlr/.

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: Proceedings of the 6th International Conference on Intelligent Virtual Agents, Marina del Rey, USA, 93-106.

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

Cerekovic, A., Pejsa, T., Pandzic, I.S. (2009). RealActor: Character Animation and Multimodal Behaviour Realization System, In Proceedings of 9th International Conference on Intelligent Virtual Agents, IVA 2009, Amsterdam, Holland, 486-487.

Ekman, P (1972). Universals and Cultural Differences in Facial Expression of Emotion. In R. Cole (ed.), Nebraska Symposium on Motivation, Lincoln, Nebraska, (pages 207-283). University of Nebraska Press.

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

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

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

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

Harrigan, J.A. and O'Connell, D.M (1996). How do you look when feeling anxious? Facial displays of anxiety. Personality and Individual Differences, volume 21, number 2 205-212.

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

Heloir, A, Kipp, M. (2009). EMBR - A Realtime Animation Engine for Interactive Embodied Agents, In Proceedings of 9th International Conference on Intelligent Virtual Agents, IVA 2009, Amsterdam, Holland, 393-404.

Heylen, D., Kopp, S., Marsella, S., Pelachaud, C. and Vilhjálmsson, H. (2008). Why conversational agents do what they do? Functional representations for generating conversational agent behaviour. The First Functional Markup Language workshop.

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

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. In: P. Ekman & E.L. Rosenberg (eds.), What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Series in affective science, (pages 133-160). New York: Oxford University Press.

Lee, J. and Marsella, S. (2006). Nonverbal behaviour generator for embodied conversational agents. In Proceedings of 6th International Conference on Intelligent Virtual Agents, volume 4133 of LNCS, Marina Del Rey, CA, USA 243-255.

Mancini, M. and Pelachaud, C. (2008). Distinctiveness in multimodal behaviors. In Proceedings of 7th Conference on Autonomous Agents and Multi-Agent Systems, Estoril, Portugal.

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

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

Niewiadomski, R., Hyniewska, S., Pelachaud, C. (2009). Modeling emotional expressions as sequences of behaviours, in: Proceedings of the 9th International Conference on Intelligent Virtual Agents, Amsterdam, Holland.

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

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

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

Pollick, F., Paterson, H.M., Bruderlin, A., Sanford, A.J. (2001). Perceiving Affect from Arm Movement, Cognition, vol. 82, 51-61.

Rehm, M., André, E. (2005). Catch Me If You Can - Exploring Lying Agents in Social Settings, In: Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS'2005), Utrecht, Holland, 937-944.

Shiota, M.N., Campos, B., Keltner, D. (2003). The faces of positive emotion: prototype displays of awe, amusement, and pride. Ann. N. Y. Acad. Sci. 1000 296-299.

Thiébaux, M., Marsella, S., Marshall, A.N., and Kallmann, M. (2008). SmartBody: behaviour realization for embodied conversational agents. In Proceedings of 7th Conference on Autonomous Agents and Multi-Agent Systems 151-158.

Thórisson, K.R., List, T., Pennock, C. and Dipirro, J. (2005). Whiteboards: Scheduling blackboards for semantic routing of messages & streams. In AAAI-05 Workshop on Modular Construction of Human-Like Intelligence, 8-15.

Vilhjálmsson, H.H., Cantelmo, N., Cassell, et al. (2007). The Behaviour Markup Language: Recent developments and challenges. In Proceedings of 7th International Conference on Intelligent Virtual Agents, volume 4722 of LNCS, Paris, France, 99-111.

Walker, J. H., Sproull, L., Subramani, R. (1994). Using a human face in an interface, Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, Boston, Massachusetts, United States, 85-91.

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

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