Contextual Factors and Adaptative Multimodal Human-Computer Interaction: Multi-level Specification of Emotion and Expressivity in Embodied Conversational Agents

Myriam Lamolle, Maurizio Mancini, Catherine Pelachaud, Sarkis Abrilian, Jean-Claude Martin, Laurence Devillers

Contextual Factors and Adaptative Multimodal Human-Computer Interaction: Multi-level Specification of Emotion and Expressivity in Embodied Conversational Agents

Conference Paper *in* Lecture Notes in Computer Science · July 2005

DOI: 10.1007/11508373_17 · Source: DBLP

CITATIONS
6

READS
116

6 authors, including:

Myriam Lamolle
Université de Vincennes - Paris 8
109 PUBLICATIONS **451** CITATIONS
SEE PROFILE

Maurizio Mancini
University College Cork
124 PUBLICATIONS **2,015** CITATIONS
SEE PROFILE

Catherine Pelachaud
French National Centre for Scientific Research
453 PUBLICATIONS **9,566** CITATIONS
SEE PROFILE

Jean-Claude Martin
Computer Sciences Laboratory for Mechanics and Engineering Sciences
264 PUBLICATIONS **3,686** CITATIONS
SEE PROFILE

Contextual Factors and Adaptative Multimodal Human-Computer Interaction: Multi-Level Specification of Emotion and Expressivity in Embodied Conversational Agents

Myriam Lamolle[1], Maurizio Mancini[1], Catherine Pelachaud[1], and Sarkis Abrilian[2], Jean-Claude Martin[2], Laurence Devillers[2]

[1] LINC, IUT de Montreuil, University Paris8,
140, rue de la Nouvelle France, 93100 Montreuil
{m.lamolle, m.mancini, c.pelachaud}@iut.univ-paris8.fr
[2] LIMSI-CNRS, BP 133, 91403 Orsay
{sarkis, martin, devil}@limsi.fr

Abstract. In this paper we present an Embodied Conversational Agent (ECA) model able to display rich verbal and non-verbal behaviors. The selection of these behaviors should depend not only on factors related to her individuality such as her culture, her social and professional role, her personality, but also on a set of contextual variables (such as her interlocutor, the social conversation setting), and other dynamic variables (belief, goal, emotion). We describe the representation scheme and the computational model of behavior expressivity of the Expressive Agent System that we have developed. We explain how the multi-level annotation of a corpus of emotionally rich TV video interviews can provide context-dependent knowledge as input for the specification of the ECA (e.g. which contextual cues and levels of representation are required for enabling the proper recognition of the emotions).

1 Introduction

Multimodal Human-Computer Interfaces aim at enabling the combined use of several communication modalities between the user and the computer. Amongst them, Embodied Conversational Agents make use of a wide range of "natural" modalities such as speech, gesture, facial expressions. This rich set of modalities provides the user with different non-verbal behaviors depending on the current application context. Yet, the definition of the dynamics of these various modalities still remains to be done. For example, emotional behavior and expressivity of animated agents play a central role for the user, namely in Intelligent Tutoring Applications. But how to define the dynamics of each modality and their combination? at which level? how to select them by considering contextual factors?

We aim at creating an Embodied Conversational Agent (ECA) that would exhibit a consistent behavior with her personality and with contextual environment factors. The behavior of an agent depends not only on factors defining her

individuality (such as her culture, her social and professional role, her personality and her experience), but also on a set of contextual (such as her interlocutor, the social conversation setting), and dynamic variables (belief, goal, emotion). These factors may act at different levels: they may act on what to say and when as well on how to say it and to express it. Thus they may act not only on the selection of a non-verbal behavior to convey a meaning (i.e. on the choice of the signals) but also on its expressivity (e.g. on their intensity level), in order to qualify it or to accentuate it.

To achieve such a goal we took a two-steps approach: 1) elaborate rules by analysis; 2) animate by copy synthesis. In the first phase we analyze and annotate a video corpus. We have elaborated an annotation scheme. Annotation of communicative behavior in social settings in extremely complex due to the large amount of variables acting in the communication process. Several annotation schemes of gesture [1] [2] [3], face [4], gaze [5], emotion [6] [7] exist. Each of these schemes are extremely rich in the data they encode and complex to use. When we have developed our annotation scheme, we had in mind the aim our study. Thus our annotation scheme encodes multimodal behaviors and complex emotions. Complex emotion may be defined as the combination of two affective states. Our annotation scheme encodes not only the signals being displayed but also their temporal evolution. Our second phase of study consists in animating an ECA. The ECA system takes as input the annotation made in the first phase and computes the face and gesture animation of the ECA.

Our expectation from this work is manifold. On one hand we aim at studying which perceptual cues are used to perceive a given emotion. The use of an ECA allows one to turn on and off given signals. By studying if subjects perceive from the synthesized animation, we can circumscribe which cues are the most salient to convey a given emotion. On the other hand, the copy synthesis method allows us to refine our animation model, in particular in relation to the modelling of gesture expressivity.

2 State of the Art

There has been a lot of psychological researches on emotion and nonverbal communication of facial and vocal expressions of acted basic emotions: anger, disgust, fear, joy, sadness, surprise [8], and also on expressive body movements [9] [10] [11] [12]. Indeed, research in non-verbal communication has already studied the relations between movements and emotions [13] [14] [15]. Yet, these studies were based mostly on acted basic emotions. Annotation of communicative multimodal behaviors in TV videos has also been addressed but without a focus on emotion [16] [17] or with the use of any annotation tool [18]. Thus, real-life multimodal corpora are indeed very few despite the general agreement that it is necessary to collect database that highlight naturalistic expressions of emotions [19]. These results from the literature in Psychology are very useful for the specification of Embodied Conversational Agents, but yet provide few details, nor do they study variations about the contextual factors of multimodal emotional behavior.

Several systems have been developed aiming at creating agents whose behaviors may be modulated by different factors: culture, emotion, social relationship, personality and so on. A first attempt was done by Barbara Hayes Roth [20] that developed a detailed and complex scheme to describe the characteristics of an ECA. Her model takes into consideration factors such as personality, habits, past memory, tastes. She elaborates a dialog and behavior model that uses this information to compute the animation of the agent. We are aware of very few other attempts. The role of social context in an agent's behavior have been considered. Poggi et al. [21] propose a model that decides whether an agent will display or not her emotion depending on several contextual and personality factors. Prendinger et al [22] integrate contextual variables, such as social distance, social power and threat, in their computation of the verbal and nonverbal behavior of an agent. They propose a statistical model to compute the intensity of each behavior. Rist and Schmitt [23] modelled how social relationship and attitudes toward others affect the dynamism of an interaction between several agents. Ruttkay and Noot [24] aim at creating agents with style. They developed a very complex representation language based on several dictionaries that reflect an aspect of the style (e.g. cultural or professional characteristics or personality) and that define the association between meanings and signals. The authors modelled explicitly how factors such as culture and personality affect behaviors.

But very few researchers have been using context specific multimodal corpora for the specification of an ECA [17]. In [25], the multimodal behaviors of subjects describing a house were annotated and used for informing the generation grammar of the Rea agent.

We distinguish our work from previously mentioned work in the sense that we do not model cultural and contextual factors per se, rather we modelled the different types of influences that may occur and how these ones may modulate an agent's behaviors at several levels.

3 Example Description

In this section we describe shortly an example for illustrating our approach. More details are provided in the following sections. The frame provided in figure 1 is from a video sample of a TV interview. The woman is reacting to a recent trial in which her father was kept in jail. As revealed by the manual annotation of such a video by 3 persons, the behavior displayed by this woman is perceived as a complex combination of anger and despair with temporal variation during the video clip. Furthermore, such emotional behavior is perceived in speech and in several visual modalities (head, eyes, torso, gestures).

Figure 2(b) shows a corresponding behavior displayed by an ECA thanks to a combination of manual specifications and automatic mapping between emotional tags and multimodal signs of emotion.

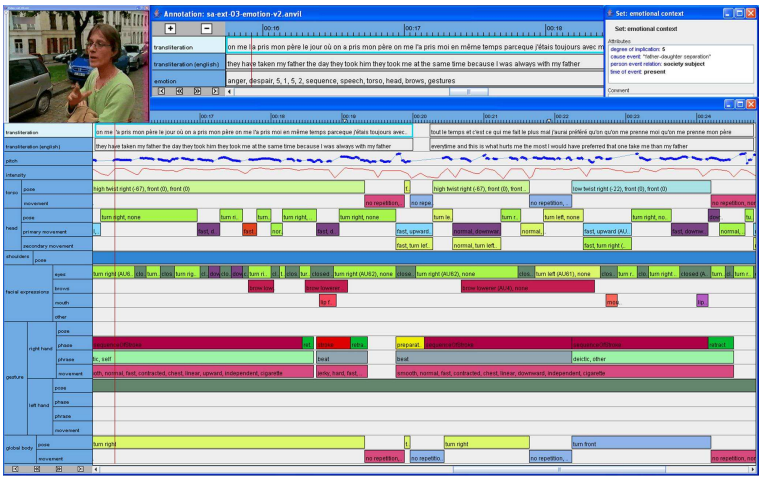

Fig. 1. Example of multi-level annotation with the Anvil tool: annotation of emotions, context, and multimodal behaviors.

4 Annotation and Modelling Emotional Behaviors

The annotation and modelling of emotional behaviors require representing multiple levels involved in the emotional process: the emotional context, the emotion itself and the corresponding multimodal behavior.

4.1 Emotion Labels

Three types of emotion annotations are generally used in research on emotion: appraisal dimensions, abstract dimensions and most commonly verbal categories. These verbal categories include both "primary" labels (anger, fear, joy, sadness, etc. [8]) and "secondary" labels for social emotions (e.g. love, submission). Plutchik [26] also combined primary emotions to produce other labels for "intermediate" emotions. For example, love is a combination of joy and acceptance, whereas submission is a combination of acceptance and fear.

The number of labels required for annotating real-life emotions might be very high when compared to basic emotions. Actually, most of the emotion modelling studies have used a minimal set of labels to be tractable [27]. Instead of using these limited number of categories, some researchers define emotions using continuous abstract dimensions: Activation-Evaluation [19], Intensity-Evaluation [28]. But, these dimensions do not allow precise emotion representation as it is, for example, impossible to distinguish between Fear and Anger. Finally, the appraisal model is useful for describing the perception / production of emotion. The major advance in this theory is the detailed specification of appraisal dimensions that are assumed to be used in evaluating emotion-antecedent events (pleasantness, novelty, etc) [29].

4.2 Expressive Behavior

Conversation is made of action (the act of speaking) and perception (the act of listening). Speaker and hearer adapt each other behaviors as the interaction evolves. Interaction involves not only speech but also non-verbal behaviors. A speaker does not behave in the same way depending on several contextual factors: she adapts her speech content and her behavior depending on the evolution of the interaction, on her relation with her conversation partner, on how this partner reacts to her speech. Quantity of gesturing, smiling, gaze between speaker and listener are highly related [30]. The externalization of nonverbal behaviors does play an important role in the communication process. Their perception interacts with the judgement one made of the speaker. To model different agent's behavior we have decided to take such a stand point: to model what is visible; that is to model the signals and how they are produced. We do not model the processes that was made to arrive to the display of such and such signals, we simply model the externalization part.

We do not aim at modelling the different factors (such as culture, personality, profession) that characterize an agent. Our work does not intend either to model how different agents would differ in their emotional reaction to an event, what culture or personality mean in their emotional reaction or to model where does a certain type of behavior come from. We are interested in understanding and modelling how a given communicative act would be expressed quantitatively and qualitatively. We are aware that our work fully rely on the modelling of complex factors such as culture, role in a society and the like. But, for our synthetic agent, we have elaborated a computational model of emotional behavior and its expressivity, leaving on the side the modelling of the what, why, how and where does expressivity come from.

Having decided to approach the problem from the visible aspect of behaviors, we turn our attention to define a set of parameters to describe them. In previous work we have defined a taxonomy of communicative behaviors based on their communicative meaning [5]. The behaviors were defined as a (meaning, signal) pair. The pairs were elaborated based on video corpus analysis. To a given meaning may be associated different set of signals. For example the meaning 'emphasis' (emphasis of a word) may co-occur with a raise eyebrow, or a head nod, or a combination of both signals. Vice versa, a same signal may be used to convey different meanings; e.g. a raise eyebrow may be sign of surprise, of emphasis, or even of suggestion. The second element of the pair, the signal, was defined in a quite static manner: no notion of dynamic variation of, for e.g., intensity, temporal duration, strength of movement was built in. Since, now, we aim at creating expressive agents we had to overcome such a limitation. We have decided to define a signal not only by its static definition (such as facial expression, gesture shape) but also by other parameters. To define them we looked in the literature of perception studies to see which parameters were investigated [14, 31]. Six dimensions representing behavior expressivity are defined. They are described in the next section.

4.3 Expressivity Dimensions

The expressivity dimensions have been designed for communicative behaviors only. Each dimension acts differently for each modality. For the face the dimensions act mainly on the intensity of the muscular contraction and its temporal course (how fast a muscle contracts). On the other hand, for an arm gesture, expressivity works at the level of the phases of the gesture: for example the preparation phase, the stroke, the hold as well as on the way 2 gestures are coarticulated one in another. We follow the taxonomy proposed by D. McNeill [2] to characterize gesture phases.

- *Overall activation*: corresponds to the quantity of movement across several modalities during a conversational turn (passive/static or animated/engaged). This parameter sets how many behaviors the agent displays while talking.
- *Spatial extent*: amplitude of movements. For the agent's face this parameter determines the quantity of physical displacement of the facial animation parameters involved in the expression. Then, spatial extent expressivity parameter will expand or condense the entire space in front of the agent that is used for gesturing.
- *Temporal*: duration of movements (e.g., quick versus sustained actions). The temporal parameter modifies starting and ending times of a facial expression. Gestures are synchronized with speech, but they may occur before the speech they accompany or after [2].
- *Fluidity*: smoothness and continuity of overall movement (e.g., smooth, graceful versus sudden, jerky). This parameter acts over several behaviors of a same modality. For two successive gestures, this dimension specifies how smoothly one gesture will map into the second one. While for the face it specifies the overall muscle contraction. Thus, as the movement gets more abrupt, there would be an increase of the muscles speed of contraction.
- *Power*: dynamic properties of the movement (e.g., weak versus strong). It corresponds to higher acceleration and deceleration magnitudes of the gesture. It also influences lip shape by controlling lip muscle tension.
- *Repetitivity*: tendency to rhythmic repeats of specific movements along specific modalities. This parameter aims to express how often a behavior is repeated. For gestures we refer to the technique of stroke expansion that we have previously introduced in [32] to capture coarticulation/superposition of beats onto other gestures. Stroke expansion repeats the meaning-carrying movement of a gesture so that successive stroke ends fall onto the stressed parts of speech following the original gesture affiliate.

5 Multi-level representation for naturalistic corpus : emotion, multimodal behaviors and context

In order to model realistic emotional behavior, literature should be completed by the collection and annotation of context-specific audio-visual data. The EmoTV corpus features 50 videos samples of TV interviews with emotional behaviors

[33]. A multilevel coding scheme has been defined after a first annotation phase. Emotion and multimodal annotations are annotated both at the global level of the video and at the level of individual emotional segments of the video. The contextual descriptors are also defined at the global level. The main difficult point of such a representation is to find the useful levels of description in term of granularity and temporality. The specificities of the multi-level coding scheme used for EmoTV are to enable annotation of both emotion labels and abstract dimensions, non-basic emotional patterns, two labels for labelling an emotion, the emotional context including some appraisal-based dimensions, a coarse temporal description of intensity variation in each segment and both a global description of perceived signs of emotion in the different modalities, and a more detailed description of multimodal behaviors in each segment [34].

Five sets of attributes represent the context namely *emotional context* including some appraisal dimensions (degree-of-implication, cause-event, person-event relation, time of event), *item interview context* (theme, place), *video-taped person* (age, gender, race), *overall communicative goal of the video-taped person* which combines consequence-event and communicative function, *recording context* (camera, character, acoustic quality, video quality).

Both verbal categories and abstract dimensions are used in order to study their redundancy and complementarity. In order to find an appropriate list of emotional labels, different strategies can be used [35] [28]. Two expert annotators labelled the emotion they perceived in each emotional segment, each time selecting one label of their choice (free choice). This resulted in 176 fine-grain labels (after a normalization phase) which were classified into the following set of 14 broader categories: anger, despair, disgust, doubt, exaltation, fear, irritation, joy, neutral, pain, sadness, serenity, surprise and worry. We have kept several levels of granularity. The coarse-grained level is composed of the 6 well-known Ekman classes [8] plus the "neutral" and "other" classes. The EmoTV coding scheme also features two classical abstract dimensions [36]: activation (passive, normal, active) and valence (negative, neutral, positive). The intensity (low, normal, high) and control dimensions (controlled, normal, uncontrolled) have also been added since they provide relevant information for the study of real-life emotion. Furthermore, for each segment coarse temporal descriptors for intensity variation are used.

The goal of the EmoTV corpus is to provide knowledge on the coordination between modalities during non-acted emotionally rich behaviors. It does not aim at providing detailed data on each individual modality.

The speech transliteration including non-verbal events markers was done using the Linguistic Data Consortium (LDC)[3] transliteration norm. Prosodic and spectral cues are automatically extracted.

In the videos only the upper body of people is visible. The coding scheme contains tracks for each visible modality: torso, head, shoulders, arms, facial expressions, gestures and global body. Torso, head and shoulders contain a description of the pose, and of the movement. Pose and movement annotations

[3] http://www.ldc.upenn.edu/

thus alternate. Head pose contains a primary position attribute (adapted from the FACS coding scheme): front, turned left / right, tilt left / right, upward / downward, forward / backward. A secondary position is available for representing combinations of positions (e.g. head to the right and down). Head primary movement observed between the start and the end pose is annotated with the same set of values as the primary position attribute. A secondary movement enables the combination of several movements. (e.g. head nod while turning the head). Tool-based annotation of gesture has already been studied [17]. We have kept some classical attributes and focused on repetitive and manipulator gestures which occur frequently in the EmoTV corpus.

The coding scheme enables the annotation of structural phases of gestures [2]: preparation (bringing arm and hand into stroke position), stroke (the most energetic part of the gesture), sequenceOfStroke (a number of successive strokes), hold (a phase of stillness just before or just after the stroke), retract (movement back to rest position). We have selected the following set of values for the gesture function (the gestures that are more frequent are listed first; representational gestures and emblems revealed to be very few after the annotation phase):

- *manipulator*: contact with body or object, movement which serve functions of drive reduction or other non-communicative functions, like scratching oneself; manipulator target (chest, hairs, eyebrows, nose, mouth); object that the video taped person is holding,
- *beat*: synchronized with the emphasis of the speech,
- *deictic*: arm or hand is used to point at an existing or imaginary object; deictic target (self, camera, other),
- *representational*: represents attributes, actions, relationships about objects and characters,
- *emblem*: movement with a precise, culturally defined meaning.

Movement quality is annotated for torso, head, shoulders, gestures, global pose and movement. The attributes of movement quality that we selected as relevant in our corpus are: the number of repetitions, the fluidity (smooth, normal, jerky), the strength (soft, normal, hard), the speed (slow, normal, fast), the spatial expansion (contracted, normal, expanded).

6 Description of the GRETA ECA system

We have developed a system that generates the behaviors of a talking ECA. To determine speech-accompanying non-verbal behaviors the system relies on a taxonomy of communicative functions proposed by Isabella Poggi [5]. A communicative function is defined as a pair (meaning, signal) where meaning corresponds to the communicative value the agent wants to communicate and signal to the behavior used to convey this meaning. The former ones are represented as a set of goals and beliefs the speaker has the goal to communicate. In the taxonomy communicative functions are differentiated in information about speaker's

beliefs, speaker's intentions, speaker's affective state and metacognitive information about speaker's mental state.

Our system, called Greta, takes as input the text the agent has to say and outputs the animation of the agent. The input text is augmented with information related to the ways the agent wants to say her text. Depending on the type of communicative acts that are specified in the input file, the agent will display different behaviors.

To control the agent we are using a representation language, called 'Affective Presentation Markup Language' (APML) where the tags of this language are the communicative functions [37].

Our system takes as input the text (tagged with APML) the agent has to say [38]. The system instantiates the communicative functions into the appropriate signals. The output of the system is the audio and the animation files that drive the facial model. The APML tags, corresponding to the meaning of a given communicative function, is converted into their corresponding facial signals. The conversion is done by looking up the definition of each tag into the library that contained the lexicon of the type (meaning, signals). Finally, we proceed with the animation generation for the agent. The animation is obtained by conversing each facial signal in their corresponding facial and body parameters.

7 A Representation Scheme for an Expressive Agent

We want to simulate that different agents may behave differently in a same situation and express their felt emotion differently. This representation allows us to define that an agent has a very expressive face or that she rarely uses wide arm movements, etc. For example, the simulation of one's nonverbal behavior, by two different agents to express anger produces two different perceivable animations. We do not aim at modelling what culture or personality mean, nor do we aim at simulating expressive animations. In this section, we detail the representation of the different levels of agent's expressivity [39] [40] in relation to modalities (face, gesture, gaze, posture, head) and we explain the computation of contextual factors effects.

7.1 Global Expressivity

In the input text, the tags are defined for the *default agent*. To allow for the generation of an expressive ECA, we associate to each agent a *behavioral profile* which specifies, on the one hand, the agent's expressivity, i.e. the agent's predispositions (which modalities the agent prefers to use) and the global expressivity (how the modalities are used), and on the other hand, the effects of the contextual factors.

The agent's predispositions represent the expressivity level of each modality. For example, an agent *Agent1* may be more expressive with the face and gestures than the default agent (i.e. her facial moves and her gestures are more visible

than the *default agent's* one) but less than the posture. The predispositions, given as input, is constant during a dialog turn.

The own agent's expressivity is represented by her predispositions to display a communicative act in the different modalities. But, the agent can use a modality through different dimensions: spatial, temporal, fluidity, power, repetitivity and overallActivity (see section 4.3). These values lessen or accentuate the intensity, the velocity, the duration, the delay of the chosen signals for the corresponding modality in the animation engine to express the communicative acts specified in the input text. The *spatial* and *temporal* parameters are local to an communicative act and can be modulated by the *fluidity*, *power* and *overallActivity* parameters. For example, according to the agent's description factors, this agent gets a large fluidity in her movements but her gestures are close to her body (the spatial dimension is set to small). We also specify which modalities are more expressive than the others; i.e. which modalities display the most expressive behavior.

7.2 Modality Hierarchy

The predisposition behavioral profile, just explained, indicates the effects of the agent's expressivity for each modality. Another factor for distinguishing agents among each other, is the modality hierarchy [41]. This hierarchy represents the modalities over which the agent is the most expressive. She may mainly use her hands to communicate or her face will be very lively, almost grimacing. To each modality (face, gaze, gesture, posture, head), we associate a value which represents the preferential level in this hierarchy. In case several modalities have the same preferential level, we consider that agent's nonverbal behavior to express a communicative act is visible through several modalities [42].

8 System Overview

Given a tagged-input file, the system instantiates the tags into a set of signals. To do so, it looks in a library the signals that correspond to the given meanings. Then it selects the signals that express the tags meaning, according to its attributes values and the agent's preferences. In the next sections, we describe the agent's contextual behavioral profile. We also detail the various selection stages of our system: the modality selection and the signals pre-selection. The first selection corresponds to determining which modality the agent uses; the second selection consists in ordering the set of possible behaviors having an equivalent meaning, from the most adequate solution to the least. This ordering takes into account the expressivity of the agent.

8.1 From Global to Local Non Verbal Behavior Specification

For each tag of the input text, the system has to decide the modality (face, gesture, gaze, posture, head one) to express the given meaning taking into account

the weight of a communicative act and the global expressivity of the agent. This decision is based on the global agent's expressivity. Among the modalities that have at least one expression which allows the system to represent the meaning, the system chooses the one with the highest priority and that is not used yet, in order to prevent conflicts.

8.2 Pre-Selection of Non-Verbal Behavior

To obtain the local expressivity of each modality, the system selects a set of expressions from a library. The expression is selected if its range of values contains the wanted expressivity value [43]. Each local expression contains the signals (representing the non verbal behavior) to play by the animation engine.

Currently, if no expression is selected, the system chooses the nearest expression. So, the animation engine can display at least one non-verbal behavior for local expressivity.

Then, the system has to order the set of expressions based on the agent's definition. This ordering allows us to obtain a list of non-verbal behaviors in the order of the agent's preferential use. This pre-selection is sent to the animation engine of the Greta system that chooses the "most adequate" non-verbal behavior.

9 From corpus analysis to ECA specification

In this section we briefly describe an example of generating the animation of an ECA from the annotation of a video. The image in figure 2(a) is from the EmoTV corpus. In section we have provided an example of the Anvil annotation for this video sequence.

In Greta we do not consider the complete annotation of the given video clip. As mentioned in section , we are concerned with the visible part of behaviors. So currently we leave aside all annotations regarding the description of the context. On the other hand we use the emotion labels as well as the description of the movement quality as input to our Greta system. We follow an analysis-synthesis loop approach to refine the animation of the ECA. The annotation of the video segment is re-written to follow the APML specification. In the example of figure 1 the annotated emotion is anger for the first half part of the segment and then it fades into despair for the rest of the segment. We have also used the annotation of the gesture strokes from the video segment to define emphasis tags in the corresponding APML text. This ensures that the gesture stroke of the ECA will happen with the emphasized words. Finally from the annotation of emotion and of multimodal behavior at the global level, we define the agent's behavioral profile. At this point, given the APML text and the agent's behavioral profile, the system automatically computes the expressivity parameters values (see 4.3) for each of the signals the agent has to produce. The animation engine considers both the signals and their expressivity to generate the agent's animation.

Fig. 2. (a) A real scene annotated by ANVIL displaying a blended emotional behavior combining sadness and anger. (b) A first simulation with the Greta system.

10 Conclusion and Perspectives

In this paper we have presented a methodology based on corpus analysis to create expressive ECAs. We have also proposed a representation scheme and a computational model for such an agent. We have explained how the annotation of expressivity in TV interviews is compatible with the specifications of our ECA. We will apply this protocol on a selection of video displaying basic and non basic emotional patterns. We will try to use the hybrid scheme used in the corpus for annotating each segment with two labels in order to consider non basic emotional patterns. The procedure will be validated via perceptual tests for evaluating how much the contextual cues, the emotion and the multimodal behaviors are perceptually equivalent in the original video and the simulation of the corresponding behavior by the ECA.

Acknowledgments

This work has been partially supported by the Network of Excellence Humaine (Human-Machine Interaction Network on Emotion) IST-2002-2.3.1.6 / Contract no. 507422 (http://emotion-research.net/). We are very grateful to Bjoern Hartmann for implementing the expressive behavior module and to Vincent Maya for his help in this project.

References

1. Kendon, A.: Human gesture. In Ingold, T., Gibson, K., eds.: Tools, Language and Intelligence. Cambridge University Press, Cambridge (1993)
2. McNeill, D.: Hand and mind - what gestures reveal about thoughts. University of Chicago Press (1992)
3. Calbris, G.: The semiotics of French gestures. University Press, Bloomington: Indiana (1990)

4. Ekman, P., Friesen, W.: Facial Action Coding System. Consulting Psychologists Press, Inc., Palo Alto, CA (1978)
5. Poggi, I., Pelachaud, C., de Rosis, F.: Eye communication in a conversational 3D synthetic agent. AI Communications **13** (2000) 169–181
6. Scherer, K., Schorr, A., Johnstone, T.E.: Appraisal processes in emotion: Theory, Methods, Research. New York and Oxford: Oxford University Press (2001)
7. Ekman, P., Rosenberg, E., eds.: What the Face Reveals : Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (Facs) (Series in Affective Science). Oxford Univ Press (1998)
8. Ekman, P.: Basic emotions. In Dalgleish, T., Power, M.J., eds.: Handbook of Cognition & Emotion. John Wiley, New York (1999) 301–320
9. deMeijer, M.: The contribution of general features of body movements to the attibution of emotions. Journal of Nonverbal behavior **13** (1989) 247–268
10. Newlove, J.: Laban for actors and dancers. Routledge, New York (1993)
11. Boone, R.T., Cunningham, J.G.: Children's understanding of emotional meaning in expressive body movement. In: Poster presented at Biennial Meeting of the Society for Research in Child Development, Washington, DC (1996)
12. Boone, R., Cunningham, J.G.: Children's decoding of emotion in expressive body movement: The development of cue attunement. Developmental Psychology **34** (1998) 1007–1016
13. deMeijer, M.: The attribution of agression and grief to body movements : the effect of sex-stereotypes. European Journal of Social Psychology **21** (1991) 249–259
14. Wallbott, H.: Bodily expression of emotion. European Journal of Social Psychology **28** (1998) 879–896
15. Montepare, J., Koff, E., Zaitchik, D., Albert, M.: The use of body movements and gestures as cues to emotions in younger and older adults. Journal of Nonverbal Behavior **23** (1999) 133–152
16. Allwood, J., Cerrato, L., Dybkær, L., Paggio, P.: The mumin multimodal coding scheme. In: Workshop on Multimodal Corpora and Annotation „ Stockholm (2004)
17. Kipp, M.: Gesture Generation by Imitation. From Human Behavior to Computer Character Animation. PhD thesis, Universität des Saarlandes (2004)
18. Atifi, H., Marcoccia, M.: L'expression et la mise en scène des émotions à la télévision : l'articulation des émotions vécues, racontées et attribuées. In: Oralité et gestualité (ORAGE 2001) : Interactions et comportements multimodaux dans la communication, Aix-en-Provence, L'Harmattan (2001) 179–182
19. Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P.: Emotional speech; towards a new generation of databases. Speech Communication (2003)
20. Rousseau, D., Hayes-Roth, B.: Personality in synthetic agents. Technical report, Stanford Knowledge Systems Laboratory Report KSL-96-21 (1996)
21. DeCarolis, B., Pelachaud, C., Poggi, I., de Rosis, F.: Behavior planning for a reflexive agent. In: IJCAI'01, Seattle, USA (2001)
22. Prendinger, H., Descamps, S., Ishizuka, M.: Scripting affective communication with life-like characters in web-based interaction systems. Applied Artificial Intelligence **16** (2002) 519–553
23. Rist, T., Schmitt, M.: Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters. In: Proc. of AISB'02 Symposium on Animated Expressive Characters for Social Interactions. (2003) 79–84
24. Ruttkay, Z., van Moppes, V., Noot, H.: The jovial, the reserved and the robot. In: proceedings of the AAMAS03 Ws on Embodied Conversational Characters as Individuals, Melbourne, Australia (2003)

25. Cassell, J., Stone, M., Hao, Y.: Coordination and context-dependence in the generation of embodied conversation. In: INLG. (2000) 171–178
26. Plutchik, R.: The psychology and Biology of Emotion. Harper Collins College, New York (1994)
27. Batliner, A., Fisher, K., Huber, R., Spilker, J., Noth, E.: Desperately seeking emotions or: Actors, wizards, and human beings. In: SpeechEmotion-2000. (2000) 195–200
28. Craggs, R., Wood, M.M.: A categorical annotation scheme for emotion in the linguistic content. In: Affective Dialogue Systems (ADS'2004). (2004)
29. Scherer, K.R.: Emotion. In Stroebe, M.H..W., ed.: Introduction to Social Psychology: A European perspective. Oxford: Blackwell. (2000) 151–191
30. Feyereisen, P., de Lannoy, J.: Gestures and speech: Psychological investigations. Cambridge University Press (1991)
31. Gallaher, P.: Individual differences in nonverbal behavior: Dimensions of style. Journal of Personality and Social Psychology **63** (1992) 133–145
32. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In: Computer Animation'02, Geneva, Switzerland, IEEE Computer Society Press (2002)
33. Abrilian, S., Devillers, L., Buisine, S., Martin, J.C.: Emotv1: Annotation of real-life emotions for the specification of multimodal affective interfaces. In: HCI International 2005, Las Vegas, USA (2005)
34. Abrilian, S., Martin, J.C., Devillers, L.: A corpus-based approach for the modeling of multimodal emotional behaviors for the specification of embodied agents. In: HCI International 2005, Las Vegas, USA (2005)
35. Cowie, R.: Emotion recognition in human-computer interaction. IEEE Signal processing Magazine (2001)
36. Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schroeder, M.: 'feeltrace': An instrument for recording perceived emotion in real time. In: ISCA Workshop on Speech & Emotion, Northern Ireland (2000) 19–24
37. DeCarolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a mark-up language for believable behavior generation. In Prendinger, H., Ishizuka, M., eds.: Life-like Characters. Tools, Affective Functions and Applications. Springer (2004) 65–85
38. Pelachaud, C., Carofiglio, V., Carolis, B.D., de Rosis, F.: Embodied contextual agent in information delivering application. In: First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS), Bologna, Italy (2002)
39. Badam, V.R.K., Gharpure, C.P.: A stochastic and multi-layered model for personality in computational agents. In: Personality in Computational Agents, Logan, UT, USA, Department of Computer Science of Utah State University (2002)
40. Kshirsagar, S.: A multilayer personality model. In: SMARTGRAPH '02: Proceedings of the 2nd international symposium on Smart graphics, New York, NY, USA, ACM Press (2002) 107–115
41. Theune, M., Heylen, D., Nijholt, A.: Generating embodied information presentations. In Stock, O., Zancanaro, M., eds.: Multimodal Intelligent Information Presentation. Kluwer Academic Publishers (2004) 47–69 ISBN=1-4020-3049-5.
42. Allwood, J.: Cooperation and flexibility in multimodal communication. In: CMC '98: Revised Papers from the Second International Conference on Cooperative Multimodal Communication, London, UK, Springer-Verlag (2001) 113–124
43. Maya, V., Lamolle, M., C, P.: Influences on embodied conversational agent's expressivity: Toward an individualization of the ecas. In: Proceedings of AISB 2004. (2004)