Interaction Design and Architecture(s) Journal - IxD&A, N. 5-6, 2009, pp. 99-106

Towards a Socially Intelligent ECA

Valeria Carofiglio, Berardina De Carolis, Irene Mazzotta, Nicole Novielli, Sebastiano Pizzutilo
Dipartimento di Informatica, Università degli Studi di Bari
Via Orabona, 4 – 70125 – Bari, Italy
{carofiglio,decarolis,mazzotta,novielli,pizzutilo}@di.uniba.it

ABSTRACT
Embedding the HCI technology with human preferences and behaviour justifies the attempt of implementing emotional and social intelligence aimed at exceeding the single ability to help the user. In this paper we present an Embodied Conversational Agent's (ECA's) architecture and methods useful to interpret the user affective attitude during her dialog with an ECA and behaving 'believably' in its turn. In particular, we present an agent architecture that is general enough to be applied in several application domains and that can employ several ECA's bodies according to the context requirements.

Author Keywords
Embodied Conversational Agent, Affective Computing.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

1. INTRODUCTION

Embodied Conversational Agents (ECA) can be seen as a new metaphor of human-computer 'intelligent' interaction which promises to be effective [1] if the hypothesis that 'characters contribute to more sociable and user-friendly interfaces' is taken for granted [2]. A well designed ECA should give the users the illusion of cooperating with a human partner rather than just 'using a tool'. The more the agent succeeds in this goal the more the users are expected to attach some anthropomorphic features to them and to show signs of affective (emotional, social) involvement in the interaction. Therefore, in developing a "computer conversationalist" that is embedded in a ECA and that is able to exhibit these capabilities it is important to conceive its architecture so as to:

(i) start from the interpretation of the spoken or written utterance;

(ii) reason on the various information the user intends to convey (emotion, social attitude, performative, content, etc.) and then to trigger communicative goals according to the current belief representation of the state of the world;

(iii) achieve these goals through a set of communicative plans ("what to say") that can then be rendered as a combination of voice and animations of the agent's body ("how to say").

This paper describes our experience in the design and implementation of a scalable architecture of a believable ECA that interacts with the user for providing advices in a domain where considering social and affective factors is crucial. To this aim, the architecture will be designed so as to dynamically manage different agents in an asynchronous mode (in different times and without any change in the architecture of others agents) allowing the modelling and building of different agent's functionalities, like showing some form of social and emotional intelligence which requires the ability to recognize those factors and utilize them to optimize the efficacy of the provided advice [3].

The paper is structured as follows: Section 2 describes the agent architecture; then in Section 3 we provide an overview of the Interpersonal Stances Modeling; in Section 4 we present a Model of Emotion Activation and finally in Section 5 we describe the Dialog Modeling component with an example of implemented dialog (Section 6). Conclusions and future work directions are shown in the last Section.

2. THE ARCHITECTURE

The ability to exhibit an emotional state and/or social signs is a shallow form of the intelligence an agent can show. The recognition of the social attitude and of the emotional state of the interlocutor should be utilized to drive reasoning behind the dialog between the user and the ECA. This implies studying how these factors may affect the ECA architecture. In our opinion, when developing an ECA, the following issues should be addressed:

- the user move should be interpreted so as to detect, beside the linguistic content: i) which is the social attitude of the user and ii) which emotions arise during the dialog;

- how these factors influence the dialog course by changing the priority of communicative goals, dialog plan and surface realization of communicative acts.

Figure 1 illustrates the architecture we propose to handle these issues. This architecture has been conceived as composed by two main functional modules: the "mind" and the "body" of the agent.

The user move is a rich information source that allows extracting knowledge about the user's intention, her social attitude, emotional state, and so on. In our approach, the "mind" of the agent uses two main different knowledge sources for reasoning on the user move and then formalizing beliefs that are useful for planning its dialog move: the user and the agent models.

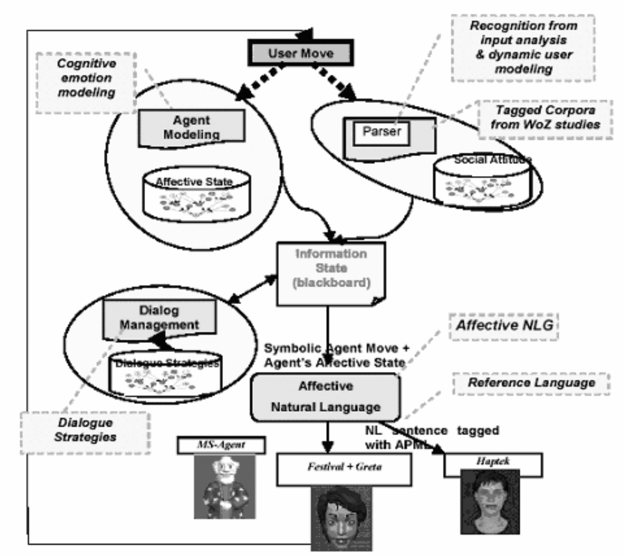

Figure 1. An overview of the ECA architecture.

The user model component allows to reason on the user's beliefs (i.e. the user move "I love fruit!" will be transformed into the correspondent belief that can be used to adapt the dialog strategy) and on the user's social attitude during the dialog (i.e. the user move "It's nice to talk to you!" will be interpreted as a sign of friendly disclosure towards the ECA). While beliefs on knowledge, preferences and interests of the user are inferred according to an approach previously employed in another system [4], in this paper we will explain how the user social attitude is recognized and monitored with a dynamic model based on Belief Network (DBN) [5].

As far as agent modelling is concerned, it is also based on a DBN which mainly aims at triggering emotions that arise in the agent mind during the interaction, in a given situation, according to the agent's personality and to the social context in which the dialog occurs.

Starting from what has been inferred by the user model component and from the emotions triggered in the mind of the agent, the dialog management module computes the agent move using a strategy that will be explained later in the paper.

The information exchange among these modules is managed using a common blackboard called information state [6]. It represents the memory of the agent and stores beliefs about the current state of the dialog, the dialog history, the current dialog move and the move scheduled for execution. This approach allows employing different methods and techniques giving to the architecture a degree of openness and scalability.

The Body

While the move computed by the "mind" module contains the meaning to express ("what to say"), the "body" has to convey these meaning according to its communicative capabilities ("how to say"). In order to decouple meanings from signals we use a mark-up language: APML [7]. These meanings include the communicative functions that are typically used in human-human dialogs: for instance, syntactic, dialogic, meta-cognitive, performative, affective, deictic, adjectival and belief relation functions [8].
The use of a reference language gives the possibility to use different bodies and different platforms and devices without changing the mind of the agent. In fact, in order to express the same meaning using different signals according for instance to the context or to the capabilities of the body of the employed ECA, each ECA's body has a conditional meaning-signal table that allows to appropriately translate an APML tag into tags expressed in Signal Expression Markup Language (SEML). SEML tags define the expressions that can be performed on each channel of the Body as described in [9].
Let's see now in more details how these modules work.

3. INTERPERSONAL STANCES MODELING

After several forms of 'anthropomorphic behaviour' of users towards technologies were demonstrated [10], various terms and concepts have been employed to denote this behaviour and describe it. Paiva [11] talks about *empathy* Hoorn and Konijn [12] address the concept of *engagement, involvement*, *sympathy* and their contrary, *distance*. Cassell and Bickmore [1] adopt the Svennevig's theory of *interpersonal relations*.

We refer to Scherer's concept of *interpersonal stance*[1] as a category which is *"characteristic of an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, coloring the interpersonal exchange in this situation (e.g. being polite, distant cold, warm, supportive, contemptuous)":*
In particular, in referring to the social response of users to ECAs, we distinguish warm from cold *social attitude*, according to the Andersen and Guerrero's definition of interpersonal warmth [13] as "*the pleasant, contented, intimate feeling that occurs during positive interactions with friends, family, colleagues and romantic partners*".

We studied this attitude and the factors affecting it by observing the verbal and prosodic behaviour of 60 subjects conversating with an ECA in a Wizard of Oz simulation study [14]. More details about the WoZ study may be found in [15]. In particular, we defined a markup language (Table

---

[1] http://emotion-research.net/deliverables/D3e%20final.pdf.

1) for the user moves after carefully examining our corpus and considering suggestions from the studies about verbal expression of social attitude [13,16,17]. Dynamic recognition of these individual signs during the dialogue enables not only to estimate the overall social attitude value but it also allows the agent to adapt its dialogue plan accordingly: for example, if the user tends to talk about herself, in the following moves the ECA will use this information to provide more appropriate suggestions. The overall social attitude of the user will be inferred dynamically from the history of the signs recognized during the dialogue to adapt the ECA's language style, voice and facial expression.

Tab. 1 - linguaggio di markup

Linguistic Signs of Social Attitude with definition
Friendly self-introduction: The subjects introduce themselves with a friendly attitude (e.g. by giving their name or by explaining the reasons why they are participating in the dialogue).
Colloquial style: The subject employs an informal language, dialect, proverbs
Talks about self: The subjects provide more personal information about themselves than requested by the agent.
Personal questions to the agent: The subject tries to know something about the agent's preferences, lifestyle etc., or to give it suggestions in the domain.
Humor and irony: The subjects make some kind of verbal joke in their move.
Positive or negative comments: The subjects comment the agent's behavior, experience, domain knowledge, etc.
Friendly farewell: This may consist in using a friendly farewell form or in asking to carry-on the dialogue.

Three PhD students labelled independently the corpus of WoZ dialogues with our markup language. According to the result of the annotation experiment we defined a set of linguistic cues that could be considered as *salient* [18] for every given of social attitude. These cues are organized into *semantic* categories. Every new user move is categorized as 'showing a particular sign of social attitude' if it includes some word sequences belonging to *semantic categories* which are defined as 'salient' for the considered sign. Recognition of linguistic signs of social attitude is performed by using Bayesian classification and can be enriched with acoustic analysis of user move, as described in [14].

Dynamic Modeling of the User Attitude. The user modeling procedure integrates (i) language analysis for linguistic cues extraction and (ii) a dynamic belief network (DBN) which considers the context in which the move was uttered. DBNs [5], also called time-stamped models, are local belief networks (called time slices) expanded over time; time slices are connected through temporal links to constitute a full model. The method allows us to deal with uncertainty in the relationships among the variables involved in the social attitude estimation (Table 2). The DBN formalism is particularly suitable for representing situations which gradually evolve from a dialog step to the next one. We applied results of the corpus analysis to learn from the annotated data a model of the user's mental state [19] which includes the dimensions of interest for dialog adaptation. In particular: in learning the temporal part of our DBNs, we took every single user move in the corpus as an independent observation and applied the K2 algorithm [20]; in learning the temporal link between the monitored variable Satt at two subsequent time instants, we took every dialog as an observation to measure the conditional probability that Satt takes a given value at time t, given its value at time t-1.

The DBN (Figure 2) is employed to infer how the social attitude of the user evolves during the dialog in relation to the dialog history. The social attitude is the *hidden* variable of our model, that is the variable we want to monitor, which depends on *observable* ones, such as the 'stable' characteristics of the users (their background and gender), the context in which the move was entered (previous agent move) and the linguistic features of the user move recognized by our Bayesian classifier (leaf nodes of our DBN). *Intermediate* variables are the signs of social attitude listed in Table 1

Links among variables describe the causal relationships among stable characteristics of the users and their behaviour, via intermediate nodes. DBNs, as employed in this paper, are said to be 'strictly repetitive models'. This means that structure and parameters of individual time slices is identical and temporal links between two consecutive time slices are always the same. We use a special kind of strictly repetitive model in which the Markov property holds: the past has no impact on the future given the present. In our simulations, every time slice corresponds to a user move, the stable user characteristics do not change from time to time (this is why we omitted the nodes Back and Gend from the figure) and temporal links are established only between dynamic subject characteristics in two consecutive time slices.

Tab. 2 – Variables of our model

Variable category	Variable Name	Label
Stable user characteristics	Background	Back
	Gender	Gend
Context	Last agent move type	Ctext
	User move type	Mtype
Monitored variable	User attitude towards the agent	Satt
Signs of social attitude	Familiar style	Fstyl
	Friendly self-introduction	Fsint
	Talks about self	Perin
	Question about the agent	Qagt
	Friendly farewell	F-Fw
	Comments (positive and negative)	Comm
Result of linguistic analysis	Cues of familiar style	Pfstyl
	Cues of friendly self-introduction	Pfsint
	Cues of talks about self	Pperin
	Cues of questions to the agent	Pqagt
	Cues of friendly farewell	Pffw
	Cues of comments	Pcomm

At the beginning of interaction, the model is initialized by assigning a value to the stable user characteristics (e.g. female user with background in Humanities). At every dialog step, knowledge about the context and evidence produced by linguistic analysis are entered and propagated in the network: the model revises the probabilities of the social attitude node. The new probabilities of the signs of social attitude and stage of change are used in formulating the next agent move, while the probability of the social attitude node supports revising high-level planning of the agent behavior.

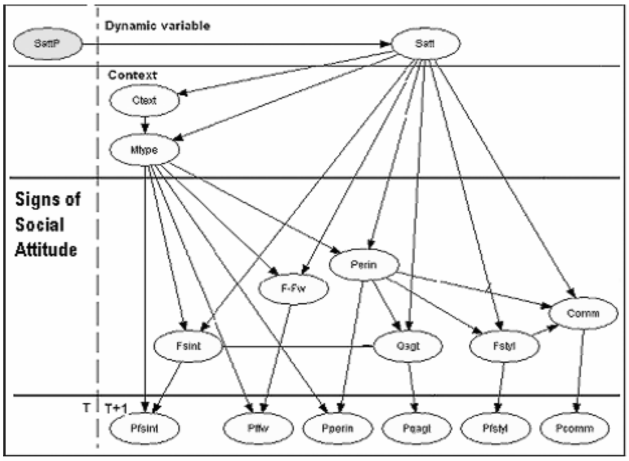

Figure 2 – User Model for the Social Attitude, a generic time-slice

We performed an evaluation of the model to examine how a variation in the threshold of the probability of the monitored variable (Satt) affects sensitivity and specificity of the model in recognizing this feature. For more details about model validation please refer to the study described in [15].

4. A MODEL OF EMOTION ACTIVATION

In our emotion modeling method [21] we pay particular attention to how emotions change of intensity with time, how they mix up and how each of them prevails, in a given situation, according to the agent's personality and to the social context in which the dialog occurs. So far, we focused our attention on event-driven emotions in Ortony, Clore and Collin's (*OCC*) theory [22]. In this theory, *positive* emotions (happy-for, hope, joy, etc.) are activated by *desirable* events while *negative* emotions (sorry-for, fear, distress, etc.) arise after *undesirable* events.

Events concerning the agent are in the *Well-being* category (joy, distress), events concerning other people are in the *FortuneOfOthers* category (happy-for, sorryfor, envy and gloating) while future events are in the *Prospective* category (fear, hope). In Oatley and Johnson-Laird's theory, positive and negative emotions are activated (respectively) by the belief that some goal will be achieved or will be threatened [23]. A cognitive model of emotions that is built on this theory should represent the system of beliefs and goal behind emotion activation and endows the agent with the ability to *guess the reason why she feels a particular emotion and to justify it*. It includes the ingredients that enable representing *how the Agent's system of goals is revised* when emotions are felt and how this revision influences planning of subsequent dialog moves.

Our model of emotion activation is represented with a *DBN* [5]. We use *DBN*s as a goal monitoring system that employs the observation data in the time interval (*Ti, Ti*+1) to generate a probabilistic model of the agent's mind at time *Ti*+1, from the model that was built at time *Ti*. We employ this model to reason about the consequences of the observed event on the monitored goals. We calculate the intensity of emotions as a function of the *uncertainty* of the agent's beliefs that its goals will be achieved (or threatened) and of the *utility* assigned to achieving these goals. According to the utility theory, the two variables are combined to measure *the variation in the intensity of an emotion* as a product of the change in the probability to achieve a given goal, times the utility that achieving this goal takes to the agent [24].

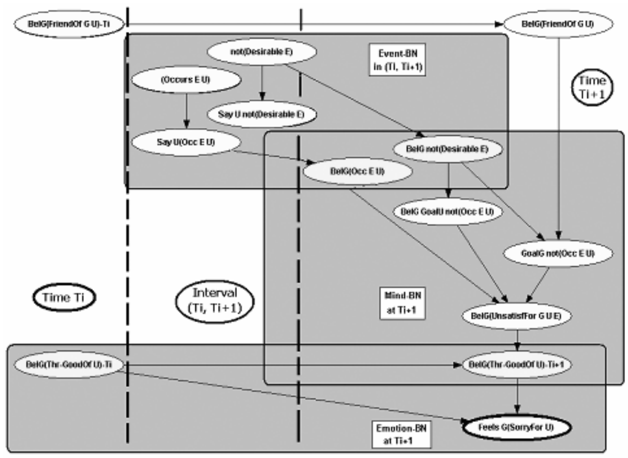

Figure 3. A portion of the DBN that represents the agent's mental state showing the triggering of Sorry-For

Let us consider, for instance, the triggering of *sorryfor* that is represented in Figure 3. This is a negative emotion and the goal that is involved, in this case, is *preserving others from bad*. The agent's belief about the probability that this goal will be threatened (Bel G (Thr-GoodOf U)) is influenced by her belief that some undesirable event E occurred to the user (BelG(Occ E U)). According to Elliott and Siegle [25], the main variables influencing this probability are the desirability of the event (Bel G not(Desirable E)) and the probability that the agent attaches to the occurrence of this event (Bel G (Occ E U)). Other factors, such as the social context (Bel G FriendOf G U)), affect the emotion intensity. The model of the agent state at time *Ti*+1 is built by automatically combining several *BN*s: the main one (Mind-BN) and one or more Event-BNs and Emotion-BNs. In the Event-BNs, the user moves are interpreted as *observable* consequences of occurred events, that activate emotions through a model of the impact of this event on the agent's beliefs and goals. The strength of the link between what the user said (Say U (Occ E U)) and the *hidden* event (Occ E U) is a function of the user sincerity; the link between this observation and the agent's belief (Bel G (Occ E U)) is a function of how believable the agent considers the user to be. Therefore, the more sincere the user is and the more likely the event is a priori, the higher will be the probability that G believes in the occurrence of the event E. Similar considerations hold for the evaluation of how *desirable* the event is (Bel G (Desirable E)); these nodes are leaves of the Event-BN. They are, as well, roots of Mind-BN: they influence G's belief that U would not desire the event E to occur (Bel G Goal U ¬(Occ E U)) and (if G is in a *empathy* relationship with U and therefore adopts U's goals), its own desire that E does not occur (Goal G ¬(Occ E)). This way, they concur to increase the probability that the agent's goal of *preserving others from bad* will be threatened.

Variation in the probability of this goal activates the emotion of *sorry-for* in G through the Emotion-BN. The intensity of this emotion is the product of this variation times the *weight* the agent gives to the mentioned goal. According to Carbonell, we define a personality as a cognitively plausible combination of weights the agent gives to the goals represented in the model [26].

The strength of the link between the goal-achievement (or threatening) nodes at two contiguous time instants defines the way the emotion, associated with that goal, decays, in absence of any event influencing it. By varying appropriately this strength, we simulate a more or less fast decay of emotion intensity. Different decays are attached to different emotion categories (positive vs. negative, FortuneOfOthers vs. Wellbeing and so on) and different temperaments are simulated, in which the *persistence* of emotions varies. The agents' affective state usually includes multiple emotions. Different emotions may coexist because an event produced several of them at the same time or because a new emotion is triggered while the previous ones did not yet decay completely. We describe in [24] how we modelled the two mentioned mixing metaphors (*microwave oven* and *tub of water*, in Picard's terminology).

5. DIALOG MODELING

The dialog manager includes three main layers:

- a *Deliberative* layer that selects the goal with the highest priority and the correspondent plan and stores in the agenda the actions of the plan.
- a *I/O Communicative layer* that executes the next action in the agenda.
- a *Reactive layer* that decides whether the goal priority should be revised, by applying reaction rules.

The dialog manager, and in particular the deliberative module, decides which goals to trigger and to pursue during the dialog, starting from the interpretation of the user move in terms of content and social attitude and according to the emotion triggered in the agent mind.

As the dialog evolves these factors may change what has been planned at two levels: i) by manipulating the inner aspects of the emotional response of our agent with an algorithm of activation/deactivation of its goals and of dynamic revision of their priorities; ii) by deciding whether the agent should manifest its emotion and how.

Handling these issues is the main task of the Reactive Layer. The idea is that the agent has an initial list of goals, each with its priority, some of which are inactive: every goal is linked, by an application condition, to a plan that the agent can perform to achieve it. The communicative actions correspondent to active plans are put in the agenda maintained by the information state. The agent starts the dialog by executing these actions but, as we said in the Introduction, the agent applies some form reasoning on the user move. The recognized social attitude and the emotion triggered in the agent mind are used to implement social and emotion-based dynamic revision of goals and consequently of the dialog.

To achieve this aim, the following knowledge sources are employed by the dialog management modules:

1. *The agent and the user model* , stored in the information state, with the interaction history. These models include two categories of factors:

– *long-term settings* that are stable during the dialog and influence the initial priority of the agent goals and therefore its initial plan, initiative handling and behaviour: agent's personality, its role, its relationship with the user;

– *short-term settings*, that evolve during the dialog and influence goal priority change and plan revision: in particular, the emotional state of agent and the social attitude of the user.

The agent's goals can be in one of the following relations among themselves:

– *Priority*: $gi < gj$: gi is more important, to the agent, than gj . If this relation holds and no constraints or preconditions are violated by satisfying it, gi will be achieved before gj .

– *Hierarchy*: $H(gi, (gi1, gi2, , gin))$: the complex goal gi may be decomposed into simpler subgoals $gi1, gi2, , gin$, which contribute to achieving it.

– *Causal Relation*: $Cause(gi, gj)$: the plan achieving the source goal gi is a precondition for the plan achieving the destination goal gj .

2. *Plans* are represented as context-adapted recipes; a recipe may be applied when some preconditions hold; its application affects the dialog state (agent's and user's mental state and interaction settings). In the healthy eating domain, our agent adopts the typical plan of intelligent advice systems:

– *situation-assessment*, to acquire information about the user,

– *describe-eating-problems*, to describe eating problems and their possible origin,

– *suggest-solution*, to describe how to eat better and to overcome problems,

– *persuade-to change*, to convince the users to change their eating habits.

3. *Reaction rules* implement goal-revision strategies. They may produce, in general, the following effects on the dynamics of plan activation:

– *add details* when the user asks for more information;

– *reduce* details in case of urgency;

– *abandon temporarily a plan to activate a new subplan* to reassure, motivate or provide more details;

– *abandon a subplan* when its goal has been achieved: for example, when the user seems to know the information the agent is providing;

– *substitute a generic subplan with a more specific and situation-adapted one*;

– *revise the sequencing of plans, to respond to the User request of* taking the initiative. This is the most delicate situation: to be cooperative, the agent should leave aside its dialog plan and follow the user request; however, as we said, communicative goals may be linked by causal relations. Therefore, when the users show the intention to take the initiative in the dialog, the agent checks whether their goal may be activated immediately or whether some preconditions have first to be satisfied. It then satisfies these preconditions with the shortest subplan before satisfying the user request [27].

As far as emotions and social factors are taken into account, according to Oatley and Johnson- Laird that claimed that *human plans are much more flexible than those so far explored in AI* [23], our reactive planning method takes these factors into account from two points of view:

a) rules regulating the **goal priority** revision by formalizing the following strategies:

– in case of *urgent events*, reduce the detail of information provided by upgrading the priority of ”most relevant” subgoals and downgrading the priority of *details*;

– in case of *desirable or undesirable events* occurred to the user, display *altruistic* social emotions (sorry-for and happy-for) by means of ”full expression” goals, that is by verbal and nonverbal means, and give them the highest priority; revise the priority of other goals; hide *egoistic* social emotions as envy and gloating;

– in case of *desirable events* occurred to the agent, activate *surface expression* goals: use verbal and nonverbal means to express them but leave the priority of other goals unvaried;

– in case of *undesirable events* (again occurred to the agent), activate *behaviour control* goals: avoid displaying any emotional reaction by activating, at the same time, repair goals.

With these rules, we formalize a situation of *empathic reaction* in which the agent temporarily substitutes the presumed goals of the user for its own, when these goals are due to an emotional state of the user [28]. If an undesirable event occurs to the users, what they are presumed to need is to be convinced that the agent understands the situation and does its best to solve the problem. If something desirable occurs to them, they need to know that the agent shares their positive experience. If, on the contrary, the undesirable event does not concern the users, they probably want to be sure that this will not interfere negatively with the dialog.

b) **rules regulating emotion display**

Once felt, an emotion E may be hidden or displayed. Again, this decision (although not always taken consciously) may be influenced by personality factors and by the interaction context. In [29] we describe how this aspect of the emotional behaviour of our agent can be modelled by means of rules that regulate activation of display goals. For example, the following rule:

```
If (Feel G E) • Is-a (E WellBeing) • (Valence E
Negative) • (Adoptive G U) then Goal G  ¬ (Display
G E)
```

activates the goal of hiding a socially negative emotion felt, because the agent has an adoptive relation with the user (i.e. fear, anger, etc.).

On the contrary, the following one:

```
If (Feel G E) • Is-a (E WellBeing) • (Valence E
Positive) (Adoptive G U) then Goal G (Display G E)
```

activates the goal of showing *a positive emotion*.

6. AN EXAMPLE OF DIALOG

To test how incorporating affective and social factors may contribute to the believability of the dialogue, we built a domain-independent simulator that uses the architecture described so far. The following is an example dialogue in the application domain of healthy eating [30]. In this dialogue the agent and user moves are denoted respectively with V_i and U_j. V_i moves are generated by our dialogue simulator, while U_j moves are introduced in free text.

V0: Hi, my name is Valentina! I'm here to suggest you how to eat well. What's your name?
U0: My name is Jana.
V1: Hi Jana, nice to meet you. How can I help you?
U1: I've been eating too much and badly in the past year and I don't feel comfortable with myself!
V2: Oh! That's a bad situation indeed, but I understand! Let me try to see how can help you. Which is your age?
U2: I'm 17.
V3: Well: teens are notorious for being hungry frequently, because they need to eat more to support the major growth of their bodies. Maybe you only need an appropriate diet and change your eating habits.
U3: Yes, you are right! I definitely need a more regular regimen.
V4: Good, let's try to do something then. Did you ever think to change your diet?
U4: Yes, sometimes I did it and got very good results. Now, either because of my laziness or my stressing rhythms I can't lay myself into more regular habits. But now I want to try again!
V5: Good, that's the way I want to see you! You seem to be in a more positive attitude now!

In the present prototype, translation of the user moves into a symbolic communicative act is quite primitive (just based on keyword analysis and on the dialogue context). For instance the U_1 move '*I've been eating too much and badly in the past year and I don't feel comfortable with myself!*' is interpreted as Say U (Occ EatingTooMuch U), Say U not(Desirable EatingTooMuch U). Symbolic communicative acts are inputs of the cognitive emotion model which, in this example, activates the Sorry-For. At the same time, linguistic cues of Friendly Style and Talks about Self are detected and evidences about these signs contribute to increase the overall likelihood of observing a warm social attitude of the user. Hence, in the subsequent move (V_2) the agent reacts by expressing her Sorry-For ('*That's a bad situation indeed!*') and by reciprocating the warm social attitude through the use of some small talk ('*But I understand!*').

The next move U_2 does not show any particular sign of social attitude and does not provide any evidence which could potentially cause emotion triggering. Here the sorry-for decays due to the absence of any more stimulus. The dialogue goes on quite neutrally until the user claims her intention to change her diet, in U_3. This event causes the triggering of a *light* Happy-For, whose intensity depends on the belief of the agent about the user sincerity, that is how true the agent beliefs the user wants to change her diet given that the user claimed it.

Then, the user reacts to the agent question by friendly talking about self. As a consequence, an higher level of the user social attitude is estimated, causing the agent to reply with a friendly style in her next move ('*Good, that's the way I want to see you!*'). Moreover, the user states again her intention to change her diet causing an increase of the intensity of the Happy-For felt by the agent.

7. CONCLUSIONS

This research builds on prior work on affect modeling and dialog simulation. In this paper we combine social attitude and emotion modeling methods to build a scalable and open architecture for an emotionally and socially intelligent ECA.

In fact, every user move is rich of information (such as linguistic cues of social attitude) which goes beyond the pure content and meaning of sentences ('what user says'). These extra-rational information about the user state of mind can be exploited to enrich the user model and can be used by a socially and emotionally intelligent ECA, in order to tailor the dialogue strategy accordingly.

The two approaches to emotion and social attitude modeling have been validate in our previous research [31,14], with satisfying results.

We are aware of the limitations of our approach. In particular, translation of the user move meaning into symbolic form is currently implemented using a keyword-spotting based approach. In our future work, we plan to refine such analysis including contextual and acoustic information [32].

The main strength of the proposed ECA architecture is its openness and flexibility. In particular, we are able to simulate interactions in different conditions, by simply changing a few parameters describing the agent's personality. In this paper we show an example of adaptation by simulating the behavior of an empathic agent which reciprocates the social attitude of the user. In our future research we plan to perform evaluation studies in order to test which combination of personality traits of the agent best increase the user satisfaction.

Moreover, thanks to the independence of our architecture from the interaction mode, we plan to perform further investigation about spoken interaction. In particular, we will enrich the model for the analysis and interpretation of the user move using prosodic and acoustic parameters for improving the recognition of both (i) the actual communicative intention attached to the user move [33] and (ii) the recognition of the user level of social attitude [14].

ACKNOWLEDGMENTS

This research would not be without the encouragement, the suggestions and the teaching of Fiorella de Rosis. Her disappearance has been for us a sad loss, but her teaching incites us to continue the work.

REFERENCES

1. Cassell, J. and Bickmore, T., 2003. Negotiated collusion: modelling social language and its relationship effects in intelligent agents. User Modelling and User-Adapted Interaction, 13, 1-2, 2003.
2. Lee, E-J., and Nass, C., 1999. Effects of the form of representation and number of computer agents on conformity. In Proceedings of CHI, 238-239.
3. Neumann B., 1994. Scale in Distributed Systems, in Casavant, T. Singhal, M. (ed) pp.463-489, Los Alamitos, CA:IEEE Comp.Society Press.
4. de Rosis, F., Pizzutilo, S., Russo, A., Berry, D. C., Molina, F.J.N., 1992. Modeling the User Knowledge by Belief Networks. User Model. User-Adapt. Interact. 2(4): 367-388.
5. Jensen, F.V., 2001. Bayesian Networks and Decision Graphs. Springer
6. Larsson, S. and Traum, D. R., 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Nat. Lang. Eng.* 6, 3-4, 323-340.
7. De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M., 2004. APML, a Mark-up Language for Believable Behavior Generation. In H. Prendinger (Ed), Life-like Characters. Tools, Affective Functions and Applications, Springer.
8. Poggi, I., Pelachaud, C., de Rosis, F., 2000. Eye Communication in a Conversational 3D Synthetic Agent. AI Commun. 13(3): 169-182.
9. De Carolis, B., 2005. MyTutor: A Personal Tutoring Agent. In Proceedings of IVA 2005, 478-488.

10. Reeves, B. and Nass, C., 1992. The Media Equation: How people treat computers, television and new media like real people and places. Cambridge University Press, New York, 1999. Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review 10*, 2, 453-469.

11. Paiva, A. (Ed), 2004. Empathic Agents. Workshop in conjunction with AAMAS'04.

12. Hoorn, J.F, and Konijn, E.A., 2003. Perceiving and Experiencing Fictitional Characters: An integrative account. Japanese Psychological Research, 45, 4.

13. Andersen, P.A. and Guerrero, L.K., 1998. Handbook of Communication and Emotions. Research, theory, applications and contexts. Academic Press.

14. de Rosis, F., Batliner, A., Novielli, N. and Steidl, S., 2007. 'You are Sooo Cool, Valentina!' Recognizing Social Attitude in Speech-Based Dialogues with an ECA. In A. Paiva, R. Picard and R. Prada (Eds): Affective Computing and Intelligent Interaction, Springer LNCS 4738, ISBN 978-3-540-74888-5, pages 179-190.

15. Clarizio, G., Mazzotta, I., Novielli, N. and de Rosis, F., 2006. Social Attitude Towards a Conversational Character. In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication. RO-MAN 2006, 2-7.

16. Polhemus, L., Shih, L-F and Swan, K., 2001. Virtual interactivity: the representation of social presence in an on line discussion. Annual Meeting of the American Educational Research Association.

17. Swan, K., 2002. Immediacy, social presence and asynchronous discussion, in: J. Bourne and J.C.Moore (Eds.): Elements of quality online education. Vol3, Nedham, MA. Sloan Center For Online Education.

18. Lee, C. M., Narayanan, S., Pieraccini, R., Combining Acoustic and Language Information for Emotion Recognition, Proc of ICSLP 2002, Denver (CO), September 2002.

19. Carofiglio, V., de Rosis F., and Novielli, N., 2005. Dynamic User Modeling in Health Promotion Dialogs. In J. Tao, T. Tan and R. W. Picard (Eds): Affective Computing and Intelligent Interaction, Springer LNCS 3784, 723-730.

20. Cooper, G.F. and Herskovitz, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Machine Learning , 9.

21. de Rosis, F., Castelfranchi, C., Goldie, P. and Carofiglio, V., in press. Cognitive Evaluations And Intuitive Appraisals: Can Emotion Models Handle Them Both?. Draft Chapter of the HUMAINE Handbook, Springer.

22. Ortony, A., Clore, G.L. and Collins, A., 1988. The cognitive structure of emotions. Cambridge University Press.

23. Oatley, K. and Johnson-Laird, P.N., 1987. Towards a Cognitive Theory of Emotions. Cognition and Emotion, 29-50,.

24. Carofiglio,V., F. de Rosis, and R. Grassano. 2008. 'Dynamic models of mixed emotion activation'. In: L. Canamero and R.Aylett (eds): Animating expressive characters for social interactions. John Benjamins Publ Co.pp.123-141

25. Elliott, C. and Siegle, G., 1993. Variables influencing the intensity of simulated affective states. In Proceedings of the AAAI Spring Symposium on Mental States.'93, 58-67.

26. Carbonell, J.C. 1980. Towards a process model of human personality traits. Artificial Intelligence, Vol. 15, pp. 49-74.

27. De Carolis, N., 1999. Generating Mixed-Initiative Hypertexts: A Reactive Approach. IUI,1999, ACM, 71-78.

28. Picard, R.W., 1997. Affective Computing. The MIT Press.

29. De Carolis, B:, Pelachaud, C., Poggi, I. and de Rosis, F., 2001. Behavior planning for a reflexive agent. Proceedings of IJCAI'01.

30. de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A. and De Carolis, B., 2006. User Modeling And Adaptation In Health Promotion Dialogs With An Animated Character. International Journal of Biomedical Informatics, pages 514-531.

31. Berry, D. C., Butler, L. T., de Rosis, F., 2005. Evaluating a realistic agent in an advice-giving task. International Journal of Man-Machine Studies 63(3): 304-327.

32. Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., Shriberg, E., Jurafsky, D., Martin, R. and Meteer, M., 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics, 26(3):339–373.

33. De Carolis B., Cozzolongo G., 2009. Interpretation of User's Feedback in Human-Robot Interaction. Journal of Physical Agents, 2, 2.