

Embodied Contextual Agent in Information Delivering Application

Catherine Pelachaud
Dipartimento di Informatica e
Sistemistica
University of Rome "La
Sapienza"
cath@dis.uniroma1.it

Valeria Carofiglio,
Berardina De Carolis and
Fiorella de Rosis
Dipartimento di Informatica
University of Bari
{carofiglio, decarolis,
derosis}@di.uniba.it

Isabella Poggi
Dipartimento di Educazione
University of Rome Three
poggi@uniroma3.it

ABSTRACT

We aim at building a new human-computer interface for Information Delivering applications: the conversational agent that we have developed is a multimodal believable agent able to converse with the User by exhibiting a synchronized and coherent verbal and nonverbal behavior. The agent is provided with a personality and a social role, that allows her to show her emotion or to refrain from showing it, depending on the context in which the conversation takes place. The agent is provided with a face and a mind. The mind is designed according to a BDI structure that depends on the agent's personality; it evolves dynamically during the conversation, according to the User's dialog moves and to emotions triggered as a consequence of the Interlocutor's move; such cognitive features are then translated into facial behaviors. In this paper, we describe the overall architecture of our system and its various components; in particular, we present our dynamic model of emotions. We illustrate our results with an example of dialog all along the paper. We pay particular attention to the generation of verbal and nonverbal behaviors and to the way they are synchronized and combined with each other. We also discuss how these acts are translated into facial expressions.

Keywords

interface and conversational agents, believable agent architectures, cognitive models and models of mind, human-like and believable qualities, personality and emotion in agents

1. INTRODUCTION

We aim at building a new type of user interface: a conversational embodied agent, that is an agent able to converse naturally with a user. The agent we have developed is a 3D face which looks quite realistic [25]. Dialoging with a pretty

face is not enough to make the conversation worth and interesting. The agent needs to be endowed with human-like qualities; she needs to be able to communicate complex messages that tightly combine verbal and nonverbal signals. As humans, we converse not only through words but also with our hands, our face, our gaze... We show our emotions or we repress ourselves from showing them. We smile to greet a friend or we nod at her. We will stare at our enemy but we will avoid mutual gaze when we are ashamed of our acts or when we are lying. Facial expression, gaze direction, head movements and so on convey essential information that, as speakers, we make great use of and, as perceivers, we are able to decode. Nonverbal behaviors may be viewed as signals (e.g. raised eyebrow or smile), but also as meanings (emotion, certainty of what we say, deictic, and so on). We want our agent to be embodied, that is to have a large repertoire of facial expressions that enable it to express different communicative functions.

But our goal is also to develop a believable agent. The definitions of believability that have been proposed involve several dimensions: personality, affect, social intelligence and, in particular, consistent behavior [23, 1, 19]. Our Agent, named Greta, is embodied in a 3D talking head. It has a personality and a social role, and the capability of expressing emotions, consistently with the context in which the conversation takes place and with its own goals.

We developed the first prototype of this Agent for an information delivering application in the context of the EU project MagiCster¹. The type of conversations we simulate at present are information-giving dialogs, in which the main function of Greta is to provide some kind of information to the User, in a given domain. As MagiCster is a 'mixed-initiative' system, the User can ask questions to Greta; this opens a question-answering sub-dialog, after which Greta revises, if needed, her discourse plan according to the User's request. Greta combines appropriately verbal and nonverbal signals when delivering information, to establish a natural communication with the User. She has therefore to achieve a very rich expressiveness during the conversation, by show-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2002 ACM 0-89791-88-6/97/05 ...\$5.00.

¹IST project IST-1999-29078, partners: University of Edinburgh, Division of Informatics; DFKI, Intelligent User Interfaces Department; Swedish Institute of Computer Science; University of Bari, Dipartimento di Informatica; University of Rome, Dipartimento di Informatica e Sistemistica; AvatarME

ing the communicative functions that are typically used in human-human dialogs; for instance: affective, metacognitive, performative, deictic, adjectival, and belief relation functions [29]. She is able to show mixed expressions, where several expressions may blend with each other according to a belief network model (Section 6).

To get an agent to be believable, the agent should act consistently with her goals, state of mind and personality. This means that the agent’s behavior and appearance should be synchronised with her speech and should be made consistent with the meaning of the sentences she pronounces; it should therefore be a reflection of her mental state and should display it by avoiding, at the same time, an ‘over-expression effect’. Finally, Greta should know when to show and when to hide an expression, according to the situation in which interaction occurs. We claim that, to this aim, Greta should be designed as a ‘BDI Agent’ (Belief, Desire, Intention) [30] and her mental state should include a representation of the beliefs and goals that drive the feeling of emotions and the decision of whether to display or to hide them. Moreover, Greta’s affective state should be dynamic and evolve through time. The formalism we employ to represent the affective state of our Agent is a Dynamic Belief Network (DBN) that computes the triggering and the evolution of her emotions. An emotion may decay over time, may increase of intensity and may mix with another emotion: our models are able to encompass these different cases (Section 4).

This paper gives an overview of the current state of the system development by describing its architecture and its various components. After referring some related works to our, we will continue by describing an example of dialog between the agent and the user, in the medical domain (Section 3). The dialog plan behind this example is very simple, and we will introduce it just in order to evidence how the verbal component is enriched with nonverbal behaviors. MagiCster will be linked, in the future, to a more sophisticated dialog planner, such as Longbow [10], which will enable us to make the discourse planning phase more dynamic and reactive. The example will illustrate the type of results we have achieved so far, as well as the requirements that should be provided in the future. We will refer to this example all along the paper, when describing the system architecture and its components. After giving an overview of the system, we pay particular attention to the integration between verbal and nonverbal behaviors in the discourse and to the way that conflicts in nonverbal signals integration are solved.

2. RELATED WORKS

In the construction of embodied agents capable of expressive and communicative behaviors, an important step is to reproduce affective and conversational facial expressions on synthetic faces [3, 7, 6, 18, 20, 29, 16]. For example, REA, the real estate agent [6], is an interactive agent able to converse with a user in real-time. REA exhibits refined interactional behaviors such as gestures for feedback or turn-taking functions. Cassell and Stone [8] designed a multi-modal manager whose role is to supervise the distribution of behaviors across the several channels (verbal, head, hand, face, body and gaze). BEAT [9] is a toolkit to synchronize verbal and nonverbal behaviors. Cosmo [18] is a pedagogical agent particularly keen on space deixis and on emotional behavior: a mapping between pedagogical speech acts and emo-

tional behavior is created by applying Elliott’s theory [14]. Ball and Breese [3] apply bayesian networks to link emotions and personality to (verbal and non-verbal) behaviors of their agents. André et al. [2] developed a rule-based system implementing dialogs between lifelike characters with different personality traits (extroversion and agreeableness). Marsella et al. [22] developed an interactive drama generator, in which the behaviors of the characters are consistent with their emotional state and individuality. In most of the mentioned agents, behaviors are mainly viewed as responses to events and actions: what is simulated is how the Agent responds with an emotion to event occurring in the domain and how emotions affects the Agent’s behavior [22, 15]. The system we propose in this paper aims at developing an Agent that is able to be affected by an emotion and to express it both verbally and nonverbally in its communicative acts. To model such an agent, we will consider, as we will see, several variables, such as the Agent’s goals and personality and information about the Interlocutor.

3. EXAMPLE OF INTERACTIONS

Let’s imagine that our embodied Agent, Greta, is a doctor able to give a patient information about a drug prescription. First of all, Greta has to inform the User about his health state. In the present example, the User is suffering from angina, which is reported by Greta. Then the User may ask any question related to his illness, such as how to cure it or how long it will probably last².

User0: *through a graphical interface, the user introduces himself and specifies the main topic of the conversation: to get information about his health state.*

Greta0: I’m sorry to tell you that you have been diagnosed as suffering from what we call angina pectoris, which appears to be mild.

User1: What is angina pectoris?

Greta1: This is a spasm of chest resulting from over-exertion when heart is diseased.

User2: Is it possible to cure it?

Greta2: Yes, certainly: a drug therapy does exist for this problem. To solve your problem, you should take two drugs. The first one is Aspirin and the second one is Atenolol.

In this example, we see 3 pairs of dialog turns. Every turn appears as a Question-Answer pair in which the user asks a question to the Agent that processes it and answers it. The Agent is provided with a knowledge base allowing her to understand the User and to formulate an answer, i.e. to provide the requested information. Since the User has been characterized as being a patient, the doctor-agent will explain in more detail the patient’s disease since, most probably, the latter does not have a deep medical knowledge. In addition, in this particular conversational topic, more emotions are involved than if the doctor-agent was talking to another doctor or even a nurse [13]. Indeed, in these latter cases the agent may speak in a harsher way (there is no direct involvement) and in a more concise way (both the other doctor and the nurse have a common knowledge and should be able to infer information from fewer medical references). On the contrary, a patient requires more information and

²We are currently leaving aside interesting aspects of the interaction between the User and Greta such as interruption of the User, misunderstanding by Greta of the User’s sentence and so on. This level of interaction will be addressed in future research.

a particular attention to the way information is exposed to him and expressed.

While conversing with the User, Greta will display various expressions that accompany her speech. In her first speaking turn (Greta0) she wants to express her empathy with the User. She will do it not only verbally (“I’m sorry to tell you”) but also nonverbally (by displaying the expression of being “sorry-for”). Expressing empathy will not be necessary if the conversational partner is a doctor or a nurse. To play down on the seriousness of the illness, Greta will emphasise both verbally and nonverbally the fact that it is still in a “mild” form. Another facial expression that will be seen in this dialog is a particular gaze direction, that plays a deictic function to indicate a given point in space. In turn Greta1, Greta indicates her chest while saying ‘a spasm of CHEST’ and looks at the User, in turn Greta2, while saying ‘YOUR problem’.

In order to produce such dialog exchanges, the agent should have a knowledge base expressed as a set of beliefs about the world, events, actions and so on. She should also be provided with some goal (what she wants to achieve) and with a plan establishing how to achieve her goal (decomposition of the goal into sub-goals). The “elementary” sub-goals have to be achieved through specific communicative functions, each of which is to be expressed by specific verbal and nonverbal signals; moreover, the signals have to be synchronized with each other in a realistic way. Making a deictic gesture (denoted by a gaze direction) at the wrong place might be interpreted wrongly. So a methodology to synchronize verbal and nonverbal signals is required. But we also want the agent not to be simply a cold and robot-looking agent but an agent that we can trust and feel sympathy for. The agent should show emotions and express them in the proper way at the appropriate time; therefore, a model of emotion and personality is required for the agent. Moreover, emotion might evolve through time and these models should be dynamic, that is they must evolve during the conversation; the same event may not provoke the same emotion depending on the different conversational contexts. In our previous example, we already pointed out that the same agent, a doctor, may behave (verbally and non-verbally) in a different way, depending on the characteristics of the interlocutor and on the relationship they have with each other. So, a model of the conversational context is required. We have listed so far a list of minimum requirements our agent should have to be able to dialog in a believable way with the user. Let us see how this is achieved in our system.

4. SYSTEM ARCHITECTURE

As mentioned previously, the type of conversation we simulate is *information-giving* dialogs in the form of *question / answer* sub-dialog. Figure 1 shows the different components of our system architecture. Three main components are included: a manager of the Agent’s *Mind*, a *Dialog Manager*, a *plan Enricher (Midas)* and a generator of the Agent’s *Body*.

When the dialog starts, a dialog goal in a particular domain is set and passed to the Dialog Manager (DM). From this goal, an overall discourse plan is produced for the Agent, by retrieving an appropriate ‘recipe’ from a plan library. This plan represents the way in which the Agent will try to achieve the specified communicative goal during the conversation. the way that a goal may be achieved depends, as well, on the cognitive model of the Agent (what we call

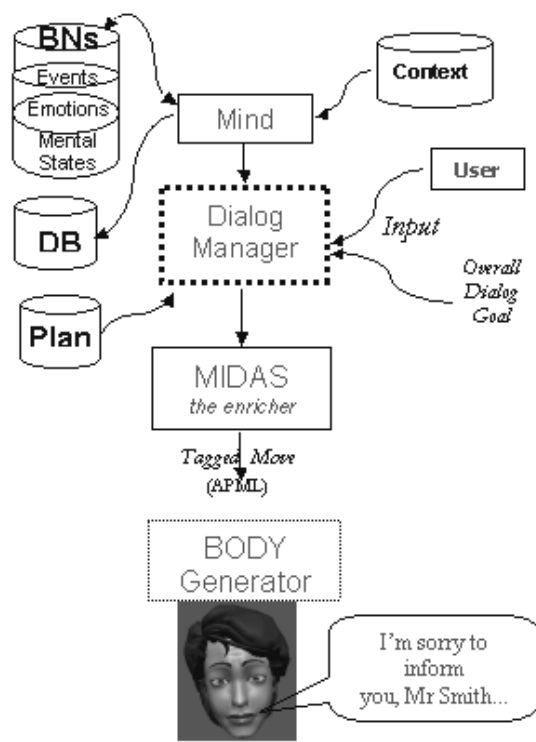


Figure 1: The architecture of our conversational agent (BN: Belief Network; DBN: Dynamic BN)

her ‘Mind’) that is her beliefs, desires, intentions, with the relations linking them and the levels of uncertainty attached to these links. This model is employed to simulate how the Agent reacts (both affectively and rationally) to events occurring during the dialog. The way the dialog goes on is a function not only of this plan but also of the following User Moves and of what we call the Social Context of the conversation [12]. Indeed, when talking with somebody we adapt our behavior and sayings to our discussion partner (what is our relation with her, what we think are her intellectual capacities and so on) and to the location of the conversation. We also behave differently depending on the topic of the conversation, that is depending on how we are related to a particular event or object that we may refer to or that was mentioned in the conversation. The social context describes the agent’s role and the relationship existing between the User and the agent. This context also describes the objects, the events and the actions that may occur in the domain and may influence the Agent’s mental state in a way that depends on her ‘personality’ (that is how the Agent reacts to an object/event/action, what is the relation of the Agent to them). Once the output dialog move has been selected by the DM, this one asks Mind whether a particular affective state of the Agent should be activated and with which intensity. In the next step the dialog move is enriched by the Midas module that adds tags indicating the communicative functions to be synchronized with the verbal stream. Then, this enriched move is passed to the Body Generator, that interprets and renders it by producing the corresponding expressive behavior.

Let us now describe these components in more details:

Mind : Mind is responsible for updating the Agent’s mental state by deciding whether a particular affective state should be activated and with which intensity. It decides, as well, whether the felt emotion should be displayed and how, according to the context variables, and revises the Agent’s goals after an emotion has been elicited. The mental state of the Agent is represented as a dynamic belief network (DBN) that is built automatically, at every dialog turn, from two main components: the network (BN) that represents the Agent’s mental state in the previous dialog turn and the networks that represent the event(s) occurred in the interval between the two turns, with their possible causes and effects. For example, let us consider the User2 move: “Is it possible to cure it?”. By representing this event in a BN, we may simulate a situation in which Greta interprets the sentence³ as a “state of anxiety”. This triggers her goal of reassuring the patient that a solution of his problem exists: “Yes, certainly: a drug therapy does exist for this problem.”

Three kinds of nodes may be found in the Agent’s mental state: ‘belief’ nodes, ‘goal nodes’ and ‘goal-achievement’ nodes. We view personality in terms of levels of importance given to achieving ‘terminal’ goals (the biological goals of survival and reproduction) and ‘instrumental’ goals (the everyday goals that serve the terminal ones). In addition, we view emotions as caused by the belief that an important goal will fail or will be achieved. For this reason, we attach a ‘weight’ to goal-achievement nodes, that depends on the Agent’s personality. Every goal-achievement node is associated with an emotion that it might activate; for instance, ‘envy’ is associated with the goal of ‘dominating others’, while ‘happy-for’ is associated with ‘desiring the good of others’. Achieving the first goal is important for ‘ambitious’ people, while achieving the second one is important for ‘altruistic’ people. Variation in the intensity of an emotion is calculated as a function of the weight the Agent attaches to reaching the corresponding goal and the variation in the probability that the goal will be achieved. Time decay of emotions is represented by linking goal nodes at consecutive time instants with an arc whose associated conditional probability table is a function of the Agent’s “temperament”. This enables us to evaluate how time decay depends, on one hand on the type of emotion and, on the other hand, on the Agent’s tendency to stay more or less long time in a given (positive or negative) emotional state. Once an emotion has been triggered, the decision of whether to display it or to hide it may be modeled by a utility diagram (for more details on the emotion modeling component, see [5]).

Dialog manager : The dialog manager (DM) is built on the top of TRINDI toolkit that enables simulating a computational model of dialogs [17]. DM controls the

³At present, the interpretation algorithm is very rough, as we focus our attention on generation rather than interpretation issues.

dialog flow by iterating the following steps⁴. We denote Greta by G and the User by U:

1. After an ‘overall dialog goal’ has been specified, an appropriate discourse plan is selected from the library of plan recipes and the first move is generated according to the first step of the plan. The ‘overall dialog goal’ becomes the main topic of the conversation. In the example, the main topic is: Greta, the doctor, has to explain to the User, the patient, the therapy to cure angina. Thus, the first step Greta has to perform is to inform the User that he is affected by angina. This is represented in formal terms by:
Overall dialog goal: Explain(G,U,Therapy(angina))
First step: Inform(G, U, Has(U, angina))
2. At the end of the first move, the initiative is passed to the User, that can ask questions to the agent on any subject among the main topics under discussion. It could ask the therapy to follow, the duration of the disease and so on. In the example cited above, the user asks: “What is angina pectoris?”
3. the User move is translated into a symbolic communicative act (through a simplified interpretation process) and is passed to the DM. So the user’s first move is translated as: Ask(U, G, Definition(angina))
4. The DM decides “what to say next” by selecting the sub-plans to execute. The User has asked the definition of angina. As Greta has already told him he got the disease, she stored her belief that the User knows he has the angina: (Bel G Bel U Has(U, angina)). So, in this speaking turn, Greta can directly explain what is the disease. This move is represented as: Explain(G, U, Definition(angina)).

cycle : The DM goes on by cycling over steps 2 to 4 until the user leaves the conversation.

The output of the DM is a discourse plan which may include just one ‘primitive’ communicative act (for instance: a greet, a thanks, an inform, a request) or may be more complex (for instance: ‘Describe an object with its properties’). In both cases, it is represented as an XML-tree structure, according to a “Discourse Plan Markup Language” (DPML) [11]. A symbolic representation of the User move is also passed to the Mind module, that updates Greta’s mental state and possibly returns the name of one or more ‘emotions’ that should be associated with the communicative act. An appropriate tag is associated, accordingly, to the corresponding part of the discourse plan. For example, after User1 move, Greta will show empathy for him and display the emotion ‘I am sorry-for you’; after User2 move, Greta will show relief. So the tag ‘sorry-for’ will be attached to her answer node in the discourse plan.

⁴We illustrate each step using the example described in section 3.

```

<node name="n1" goal="Explain(Has(U,disease))" role =
"satellite" focus="disease" RR = "Elaboration-Object-Attri-
bute"> <node name="n2" goal="Inform(Has(U, disease))"
role="nucleus" focus="Has(U, disease)" RR="null"/> <node
name="n3" goal="Inform(Severity(disease))" role="satellite"
focus = "Severity(disease)" RR="null"/> </node>

```

Midas : The Midas module has the role of translating the symbolic representation of a dialog move into an Agent's behaviour specification. In order to overcome integration problems between the DM and Body components and to allow their independence and modularity, we based the specification of Midas on XML. Thus we have developed our own XML-language specification called Affective Presentation Markup Language (APML) [11]. The role of APML is to mark up the outcome of the DM with tags denoting the meaning value of communicative functions. These tags will then be instantiated by the body generator module as a given facial expression, head movement or even gaze behavior. The algorithm applied by Midas translates the DPML-based tree-structure into an APML-based structure, through a set of transformation rules that depend on the information attached to nodes in the discourse plan: rhetorical relation (RR) name and type, communicative goal, discourse focus and so on. That is, the transformation algorithm reads recursively a given DPML tree, down to its leaves. Initially, the root tag <APML> is introduced, followed by the <turn-allocation> tag whose type attribute is set to "take", to indicate that the agent takes the initiative. After this step, the appropriate recursive schema is activated, according to the value of the 'RR attribute' attached to the node. The general rule is to put the <belief-relation> tag emphasis on the RR marker and on the satellite for nucleus-satellite RRs, but only on the RR marker for multi-nuclear ones (i.e. Ordinal Sequence, Contrast, etc.). The type attribute of the <belief-relation> tag is set with the name of the RR. When the algorithm reaches a leaf node, the **generate-performative** function is called and the recursion ends. This function is responsible for the surface realisation, in which the <performative> element is generated. If Mind has established that an emotion is felt by the Agent in correspondence with that performative, the affect attribute of the performative tag is set to that emotion's name. Besides generating the <performative> tag, the generate-performative function produces the verbal part of the speech act. An <adjectival> tag is added when the argument of the communicative goal is a quantitative attribute of the discourse focus. Finally, the <deictic> tag is set when the argument of the communicative goal is described in the domain knowledge base as 'referrable through its coordinates'. In the example cited above, "severity" is a quantitative property of diseases, which is also the discourse focus: therefore, the <adjectival> tag is generated around the attribute-word (see example of turn S0, Section 5).

Body Generator : The body generator module interprets the APML-tagged dialog move and decides which signal to convey on which channel for each communicative

act⁵. In a previous work, we have provided a definition of communicative act [29] as being a (meaning, signal) pair where the 'meaning' item corresponds to the communicative value of the 'signal' item. Then, we have elaborated a taxonomy of nonverbal communicative acts [29]; these acts may be divided into 5 main groups that provide information on (1) deixis and information on physical or metaphorical properties of referents (acts as an adjective); (2) the degree of certainty with which the Agent believes what she is saying; (3) the expression of an Agent's goal: the performative of her sentence, the topic-comment distinction, the discourse rhetorical relations, the turn allocation in conversation; (4) the expression of emotions; and finally, (5) on the kind of thinking activity in which the Agent is currently engaged. For each of these acts and their meanings, we defined a list of corresponding facial signals; that is, we built a lexicon of facial expressions [28] and gaze [29], as a set of rules that map every signal to a given meaning. For example: within the 'certainty' class belonging to the Agent's belief group, raising eyebrow is used to mark that the Agent is 'uncertain' of what she is saying, while a 'small frown' will denote that she is certain about it. We based our work on findings from the literature as well as on the results of analysis of a corpus of videotapes, in which facial expression and head movements were described using the notation system developed by Poggi and Magno-Caldognetto [27]. (The reader may refer to [29] for a more detailed description of the communicative functions of Greta's facial expressions). Figure 2 shows some examples of these expressions. These signals are then translated into facial parameters as defined within the facial model.

Body Model : The Body we use is a combination of a 3D face model compliant with the MPEG-4 standard [25] and of Festival [4], a speech synthesizer. The facial model is capable of expressing the nonverbal communicative functions foreseen for our conversational agent. The text of each dialog move with its tags is given as input to Festival, which provides the duration of the phonemes and a wav file (an audio file). Phonemes are the smallest temporal unit we are considering. Knowing the phoneme duration enables us to retrieve the exact duration of any expressions as defined by the tags in the dialog move, thus ensuring the synchrony between speech and other visual activities. The facial expressions are translated into facial parameters. It may occur that during this process co-occurring expressions may require the same facial signals (e.g., eyebrow, head direction) getting rise at a conflict. Such a conflict is solved using a belief network as explained in Section 6). When no conflict occurs on overlapping expressions (e.g an expression is 'a smile' and another expression corresponds to a 'wide eye opening'), the final one is obtained by simply adding them (from the previous example, the final

⁵Until now we have been concentrating on the generation of facial expression, leaving aside, for the moment, any work of the voice. This choice is partly determined by our study purpose, but it is also partly due to the state of the art to define synthesized emotive voice that often lack of naturalness.



Figure 2: Expression of ‘suggest’, ‘surprise’ and ‘tiny’

expression would be a smile and a wide eye opening.

To connect the various components (Mind, Midas, Greta, and the DM itself), we use a Java class (jcontroller), which controls activation, termination and information exchange for the various processes involved in the dialog management, via socket.

5. EXAMPLE OF DIALOG OUTPUT

Taking again the example of section 3, we now describe how it is generated by the Midas component of our system. The tag names refer to a given communicative function whose values are specified by quotes. For example, in turn Greta0, the affective communicative act is specified with the value ‘sorry-for’.

U0: selection of the overall dialog goal.

S0: <APML><turn-allocation type = “take”> <performative type = “inform” affect=“sorry-for” certainty = “certain”> I’m sorry to tell you that you have been diagnosed as suffering from what we call angina pectoris, </performative> <belief-relation type = “elaboration-object-attribute” > which <performative type=“inform” certainty=“certain”> appears to be<adjectival type = “small”>mild.</adjectival> </performative> </belief-relation> </turn-allocation> </APML>

U1: What is angina pectoris?

S1: <APML><turn-allocation type = “take”> <performative type = “inform” certainty=“certain”> <belief-relation type = “gen-spec”> This is </belief-relation> a spasm of <deictic = “chest”> chest </deictic></performative><belief-relation type=“cause-effect”> resulting from <performative type = “inform” certain = “certain” > over-exertion when heart is diseased. </performative> </belief-relation> </turn-allocation> </APML>

U2: Is it possible to cure it?

S2: <APML> <performative type=“suggest” affect=“relief”> Yes, certainly: a drug therapy does exist for <deictic = “user”>this </deictic> problem. you should take two drugs. </performative> <performative type = “inform”> <topic-comment type=“comment”>The first one</topic-comment> is Aspirin<belief-relation type=“sequence”>and </belief-relation> <topic-comment type=“comment”>the second one </topic-comment> is Atenolol. </performative> </APML>

In every answer of the Agent, we may notice that the dialog is annotated with tags. These tags represent the communicative functions associated with given words. Every dialog turn by the agent starts with a notification that the agent has now the speaking turn. If we start examining the first dialog pair, the doctor shows empathy to the user: while informing the user on his sickness, she will show she is sorry for his illness. This is a typical expression that might not

be shown if Greta dialogs with another doctor. She will not have to show empathy but just deliver a precise information; no expression of emotion will be shown in this case. So we may notice that the expression displayed by the agent depends on the person she is talking to. Looking back at the first dialog pair, we may also notice that the agent delivers an information, but at the same time the doctor is certain of her diagnosis and informs the user in an assertive manner, verbally and nonverbally. So the text is marked not only by an affective tag but also by a certainty tag. The last tag we may notice is related to the word ‘mild’. The agent shows her evaluation of the illness (the illness is mild) and displays an iconic movement, that is a movement that imitates the characteristic of ‘mild’. There is a parallel between the meaning of ‘mild’ in this sentence and its concurrent expression: small aperture of the eyes. So in the first turn the annotated communicative acts arise from different views and are made for different purposes: Greta wants to display a close relationship with the user by showing empathy; but she is also a doctor and should play her role by showing she is certain of what she is providing as information (the performative of the sentence). Finally, she gives an evaluation of the illness of the user. These communicative acts are then translated into facial expressions by our system. In the other dialog pairs, we may also find tags denoting a deictic function that indicates a point in space. Other tags that may be found are related to the rhetorical relations [21]. They provide information on how successive sentences are related to each other.

6. CONFLICT RESOLUTION

So far, we have characterized communicative acts by their functional meaning. But if we turn our attention to the signal part of a communicative act, we notice that a communicative act may be shown on different channels of the face such as eyebrows, mouth shape, gaze direction, head direction and head movement. The signal part of each communicative act is described by specifying (when necessary) a value for these channels. Each channel has a set of possible values. The channel ‘head direction’ may have the values: head turn right, head turn left, head turn up, head turn down, or head aside. Eyebrows may have the following shape: raised eyebrows, frown, or oblique eyebrows. An expression is defined as a combination of values for a set of given facial channels. For example: the expression of ‘sorry-for’ is shown by a specific head direction and, at the same time, by a specific eyebrow shape. An expression of ‘certainly-not’ is shown by a unique signal, a specific eyebrow shape. The first step done by the body generator is to translate every single tag in its corresponding facial expression. A tag may involve more than just one facial channel. In the annotated dialog shown in section 5, we may notice that several tags occur on a same text span. In the first dialog turn of Greta, Greta0, the first phrase *I’m sorry to tell you that you that you have been diagnosed as suffering from what we call angina pectoris*, is tagged four times (namely: turn-allocation, performative, affective and certainty). Each of the tags corresponds to a given facial expression. Some of these expressions may involve the same channel. Moreover, different values may be assigned for the same channel. ‘Sorry-for’ is expressed by oblique eyebrows while ‘certainly-not’ is done by an intense frown. A conflict arises in this case: what should be the shape of the eyebrow: raised or

oblique? Rather than using priority rules [9] that decide which communicative act should prevail or simple additive rules [7, 24] that add all signals in all channels, we have defined a methodology to resolve conflicts at the channel level. We have built a belief network (BN) that links communicative acts to channels (see Figure 3) [26]. The root nodes correspond to the communicative functions (the meaning element of communicative acts), while the leaf nodes correspond to the signals. The communicative functions we are considering have been described above (Section 4): performative, emotion, belief-relation, certainty, metacognitive, intonational structure, topic/comment, turn-taking. Two intermediate nodes linking performatives to leaf nodes are introduced: dominance and orientation. They correspond to the variables used in the definition of performatives [28] and serve to lower down the number of variables in the construction of the belief network. Dominance refers to the ‘power relationship’ underlying the choice of the performative and Orientation to the Interlocutor ‘in whose interest the requested action is addressed’. All performatives are characterized by these variables. For example, the performative ‘order’ is characterized as being ‘dominant’ and in ‘oneself-interest’. In the same way and for the same purpose as for performatives, two intermediate nodes are linked to the node ‘emotion’: ‘valence’ (positive or negative) and ‘time’ (past, current and future) [23]. Valence is commonly used to differentiate emotions (positive emotions such as ‘joy’ and ‘satisfaction’ versus negative emotions such as ‘anger’ and ‘sadness’). The intermediate node ‘time’ refers to the time at which the event that triggers the emotion is believed to happen [23]. ‘Fear’ or ‘distress’ refer to an event that might happen in the future, while ‘sadness’ or ‘resentment’ are due to events happened in the ‘past’. The leaf nodes are: eyebrow shape (raised eyebrow), head and gaze direction (head up, gaze down), mouth shape (smile) and head movement (nod). When co-occurring communicative acts happen on a same text span, the probability of the corresponding root nodes of the BN is set to 1. The output of this belief-network is a value for every facial channel involved in these co-occurring communicative acts. For example: if the node ‘emotion’ with the value ‘sorry-for’ and the node ‘certainty’ with the value ‘certainly-not’ are set to 1, the output value for the signals are the following: the eyebrows are given the value ‘frown’ while the head direction is set to ‘head-aside’. Had no conflict-resolution strategy been introduced, the ‘sorry-for’ expression would have been represented by a ‘oblique eyebrow’ and ‘head aside’, while ‘certainly-not’ would have been marked by a frown’. After the conflict resolution, the final expression obtained by combining both expressions is ‘frown’ and ‘head aside’. We can notice that the BN has cut off the ‘sorry-for’ signal (oblique eyebrow) at the eyebrow level. The resulting expression conveys both meanings at once, thus showing a complex meaning. It is a combination at the channel level but also at the meaning level. If the expressions ‘sorry-for’ and ‘certain’ have to be displayed simultaneously, since ‘certain’ is marked by a frown which has a lesser appearance weight than the frown denoting ‘certainly-not’, the BN will keep the eyebrow shape of the expression ‘sorry-for’. Let us see an other example. If the Agent is certain of what she is saying but wants to emphasise the focus of her sentence, two communicative functions will mark the emphasised part of the text: ‘certain’ and ‘comment’. Both are expressed through eyebrows, but the former

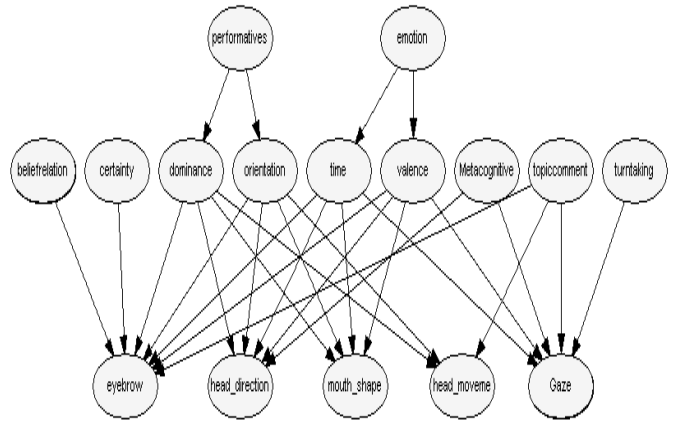


Figure 3: Belief-Network of facial channels



Figure 4: Expression of ‘sorry-for’, ‘certainly-not’ and combination of both expressions with conflict resolution

is through a ‘small frown’, the latter through an ‘eyebrow raising’. The result of the BN is that the ‘certain’ meaning is suppressed, while the ‘comment’ is displayed. This method allows us to combine expressions at a finer level, as the conflict is resolved at the channel level.

7. CONCLUSION

In this paper we have presented the minimum requirements that a conversational agent needs to be believable and expressive and how these requirements have been embedded in our system architecture. The type of conversation our agent is able to sustain with a User is of the type query/answer for an information-giving application. Our Agent respects the ‘belief, desire, intention’ structure. Her beliefs and goals are dynamically defined in the Mind component of the system. They evolve with time and according to what is happening during the conversation with the User. The agent displays facial expressions and gaze behaviors in synchrony with her verbal output. The meaning of the expressions is in correspondence with the state of mind the agent is in and with what she is saying. Expressions may convey complex meanings. Such an Agent offers a new type of user interface that may have valuable qualities in future applications.

8. REFERENCES

- [1] J. Allbeck and N.I. Badler. Consistent communication with control. In *Proceedings of the workshop on "Multimodal communication and context in embodied*

- agents", Montreal, Canada, May 2001. The Fifth International Conference on Autonomous Agents.
- [2] E. Andre, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogues for animated presentation teams. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.
 - [3] G. Ball and J. Breese. Emotion and personality in a conversational agent. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.
 - [4] A.W. Black, P. Taylor, R. Caley, and R. Clark. Festival. <http://www.cstr.ed.ac.uk/projects/festival/>.
 - [5] V. Carofiglio and F. de Rosis. Mixed emotion modeling. In R. Aylett and D. Canamero, editors, *Symposium of the AISB'02 Convention*, volume Animating Expressive Characters for Social Interactions, London, Avril 2002.
 - [6] J. Cassell, J. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjálmsón, and H. Yan. Embodiment in conversational interfaces: Rea. In *CHI'99*, pages 520–527, Pittsburgh, PA, 1999.
 - [7] J. Cassell, C. Pelachaud, N.I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Computer Graphics Proceedings, Annual Conference Series*, pages 413–420. ACM SIGGRAPH, 1994.
 - [8] J. Cassell and M. Stone. Living hand and mouth. Psychological theories about speech and gestures in interactive dialogue systems. In *AAAI99 Fall Symposium on Psychological Models of Communication in Collaborative Systems*, 1999.
 - [9] J. Cassell, H. Vilhjálmsón, and T. Bickmore. BEAT : the Behavior Expression Animation Toolkit. In *Computer Graphics Proceedings, Annual Conference Series*. ACM SIGGRAPH, 2001.
 - [10] M.G. Core, J.D. Moore, and C. Zinn. Supporting constructive learning with a feedback planner. In *AAAI Fall Symposium on "Building Dialogue Systems for Tutorial Applications"*, Cape Cod, MA, Nov. 2000.
 - [11] N. De Carolis, V. Carofiglio, and C. Pelachaud. From discourse plans to believable behavior generation. In *International Natural Language Generation Conference*, New-York, 1-3 July 2002.
 - [12] N. De Carolis, C. Pelachaud, I. Poggi, and F. de Rosis. Behavior planning for a reflexive agent. In *IJCAI'01*, Seattle, USA, August 2001.
 - [13] F. deRosis and F. Grasso. Affective natural language generation. In Ana Paiva, editor, *Affect in interactions*. Springer-Verlag, Berlin, 2000.
 - [14] C. Elliott. *An Affective Reasoner: A process model of emotions in a multiagent system*. PhD thesis, Northwestern University, The Institute for the Learning Sciences, 1992. Technical Report No. 32.
 - [15] J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the 5th International Conference on Autonomous Agents*, Montreal, Canada, May 2001.
 - [16] W.L. Johnson, J.W. Rickel, and J.C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *To appear in International Journal of Artificial Intelligence in Education*, 2000.
 - [17] S. Larsson, P. Bohlin, J. Bos, and D. Traum. *TRINDIKIT 1.0 manual for D2.2*. <http://www.ling.gu.se/projekt/trindi>.
 - [18] J.C. Lester, S.G. Stuart, C.B. Callaway, J.L. Voerman, and P.J. Fitzgerald. Deictic and emotive communication in animated pedagogical agents. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.
 - [19] A. Bryan Loyall and Joseph Bates. Personality-rich believable agents that use language. In W. Lewis Johnson and Barbara Hayes-Roth, editors, *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 106–113, Marina del Rey, CA, USA, 1997. ACM Press.
 - [20] M. Lundeberg and J. Beskow. Developing a 3D-agent for the August dialogue system. In *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing*, Santa Cruz, USA, 1999.
 - [21] W.C. Mann, C.M.I.M. Matthiessen, and S. Thompson. Rhetorical structure theory and text analysis. Technical Report 89-242, ISI Research, 1989.
 - [22] S. Marsella, W.L. Johnson, and K. LaBore. Interactive pedagogical drama. In *Proceedings of the 4th International Conference on Autonomous Agents*, Barcelona, Spain, June 2000.
 - [23] A. Ortony. On making believable emotional agents believable. In R. Trappl and P. Petta, editors, *Emotions in humans and artifacts*. MIT Press, Cambridge, MA, in press.
 - [24] C. Pelachaud, N.I. Badler, and M. Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1):1–46, January-March 1996.
 - [25] C. Pelachaud, E. Magno-Caldognetto, C. Zmarich, and P. Cosi. An approach to an italian talking head. In *Eurospeech'01*, Aalborg, Denmark, September 3-7 2001.
 - [26] C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, To appear.
 - [27] I. Poggi and E. Magno Caldognetto. A score for the analysis of gestures in multimodal communication. In L. Messing, editor, *Proceedings of the Workshop on the Integration of Gesture and Language in Speech*, Applied Science and Engineering Laboratories, pages 235–244, Newark and Wilmington, Del., October 7-8 1996.
 - [28] I. Poggi and C. Pelachaud. Performative faces. *Speech Communication*, 26:5–21, 1998.
 - [29] I. Poggi, C. Pelachaud, and F. de Rosis. Eye communication in a conversational 3D synthetic agent. *AI Communications*, 13(3):169–181, 2000.
 - [30] A.S. Rao and M.P. Georgeff. Modeling rational agents with a BDI-architecture. In J. Allen, R. Fikes, and R. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufman, 1991.