APML, a Mark-up Language for Believable Behavior Generation

Berardina De Carolis,
Valeria Carofiglio
Dipartimento di Informatica
University of Bari
decarolis@di.uniba.it
carofiglio@di.uniba.it

Massimo Bilvi
Dipartimento di Informatica e Sistemistica
University of Rome "La Sapienza"
bilvi@dis.uniroma1.it

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

ABSTRACT

Developing an embodied conversational agent able to exhibit a human-like behavior while communicating with other virtual or human agents requires to enrich the dialogue of the agent with nonverbal information. Our agent is defined as two components: a Mind and a Body. Her mind reflects her personality, her social intelligence as well as her emotional reaction to events occurring in the environment. Her body corresponds to her physical appearance able to display expressive behaviors. To specify the format of the dialogue move that should act as an interface between the agent's Mind and her Body, we designed a Mind-Body interface that takes as input a specification of a discourse plan in an XML language (DPML) and enriches this plan with the communicative meanings that have to be attached to it, by producing an input to the Body in a new XML language (APML). Moreover we have developed a language to generate facial expressions. This language combines facial basic expressions with operators to create complex facial expressions. The purpose of this paper is to describe these languages and to illustrate our approach to the generation of behavior of an agent able to act consistently with her goals and with the context in which the conversation takes place.

1. INTRODUCTION

Humans communicate combining signals of different nature. Signals like body posture (leaning forward), gestures (pointing at something, describing object dimensions), facial expressions (smile, nose wrinkling), gaze (making eye contact, looking down or up, to a particular object) may be used in combination with speech. Moreover, the way in which people communicate, and therefore the employed signals, is influenced by their personality, goals and affective state and by the context in which the conversation takes place [5]. Developing a "computer conversationalist" that is embedded, for instance, in a virtual human-like body (i.e. a smart avatar) and is able to exhibit these added dimensions of communication means moving from natural language generation to multimodal behavior generation.

The purpose of this paper is to describe how we enrich a dialogue generator with information on context-adapted behavior in an embodied conversational agent (ECA). Our agent shows a personality and a social intelligence and is able to react emotionally to events occurring in the environment, consistently with the context in which the conversation takes place and with her goals. To achieve such a context-adaptable multimodal behavior, our planner decides in the content planning step, which verbal and non-verbal signals to employ in every conversational move. This approach requires exploiting knowledge about the mental and physical capabilities of the agent during planning. The planner must decide which are the discourse steps that the agent has to carry out to achieve the given communicative goal; in addition, it has to indicate the combination of signals through which every step of the planned discourse has to be rendered. The main advantage of this alternative is that the dialogue move will be planned consistently with the agent's mental state, by establishing how to combine verbal and non verbal components.

In our system, the agent is seen as an entity made up of two main components, a 'Mind' and a 'Body', which are interfaced by a common I/O language. This language assures the independence of both components and it maintains the modularity of the system. During the conversation, the agent's Mind decides what to communicate, by considering the dialogue history, the conversational context and her own current cognitive state [2]. The Body "reads" what the Mind has decided to communicate. It interprets and renders it at the surface level, according to the communicative channels that it can employ: different bodies may have different expressive capabilities and therefore may use different channels. To achieve a rich expressivity, the output of the agent's Mind cannot be just a combination of symbolic descriptions of communicative acts. It should include, as well, a specification of the 'meanings' that the Body will have to attach to each of them. The Mind of our conversationalist has therefore to be able to perform the following functions: select an appropriate dialogue move, decide whether, in correspondence of that move, an emotion has to be displayed and, finally, specify the meanings that have to be conveyed through the selected move; these meanings include the communicative functions that are typically used in human-human dialogues: for instance (as far as facial expression and gaze are concerned), affective, meta-cognitive, performative, deictic, adjectival and belief relation functions [16].

To specify the format of the dialogue move that should act as an interface between the agent's Mind and its Body, we designed a Mind-Body interface that takes as input a specification of a discourse plan in an XML language (Discourse Plan Markup Language, DPML) and enriches this plan with the meanings that have to be attached to it, by producing an input to the Body in a new XML language (Affective Presentation Markup Language, APML). After describing the system architecture, we will introduce APML, illustrating how it works as the Mind-Body interface.

2. MAGICSTER ARCHITECTURE

This work is part of a larger project, MagiCster, an EU-project[1] which aims at creating believable conversational agent. The architecture of MagiCster is shown in Figure 1. It is made up of two main components (a Mind and a Body), interfaced by a Plan Enricher. The agent's Mind includes a Content Planner, a Dialogue Manager and an Affective Agent Modeling module. This module is responsible for updating the agent's mental state, that is her goals and beliefs. The Body is a 3D face/avatar, with a speech synthesiser [1] for animated spoken delivery. We will briefly describe each module, to focus our description on the Mind-Body interface.

The Affective Agent Modeling module decides whether a particular affective state should be activated and with which intensity and whether the felt emotion should be displayed in a given context [5].

The Content Planner is responsible for the generation of the discourse plan appropriate to the context [4]. A discourse plan is a tree identified by its name; its main components are the nodes that are identified, as well, by a name; nodes include mandatory attributes describing the communicative goal, the discourse focus and the rhetorical elements (role in the Rhetorical Relation (RR) of the father-node and rhetorical relation). The DPML DTD is:

```
<!ELEMENT d-plan (node+)>
<!ATTLIST d-plan
        name CDATA #REQUIRED
>
<!ELEMENT node (node*, info*)>
<!ATTLIST node
        name CDATA #REQUIRED
        goal CDATA #REQUIRED
        role (root | nucleus | sat) #REQUIRED
        RR CDATA #REQUIRED
        focus CDATA #REQUIRED
        >
```

The XML-based annotation of discourse plans is justified by two reasons. The first one is that it enables building a library of standard explanation plans that can be instantiated when needed, to be used in any application context. The second one is that XML may be a standard interface between the generator modules, to favour resources distribution and reuse.

The Dialogue Manager is built on top of the TRINDI architecture [10], which provides an engine for computing dialogue moves and a space in which information relevant to the move selection and effect can be stored. Such information may be, for instance, the agent's mental state and the current plan. After a plan has been selected from the library of plan recipes, the first agent move is generated according to the first step of this plan. In the case the agent is dialoging with a User, the dialogue starts and the DM controls its flow by iterating the following steps, until the conversation ends [13]:

1. the initiative is passed to the User, that can make questions on any of the topics under discussion;
2. the User move is translated into a symbolic communicative act (through a simplified interpretation process) and is passed to the DM;
3. the DM decides "what to say next" by selecting the sub-plan to execute.

The Plan Enricher translates the symbolic representation of a dialogue move into an agent's behaviour specification. A dialogue move may be a 'primitive' communicative act (for instance: a 'greet', a 'thanks', an 'inform', a 'request') or a more complex plan (for instance: 'Describe an object with its properties'), annotated according to DPML. The choice of the appropriate communicative act (an implore rather than an order; showing or not showing an emotion) is based on the conversational context [2]. An algorithm translates this DPML-based tree-structure into another XML-based language (APML), through a set of transformation rules that depend on the information attached to nodes in the discourse plan: rhetorical relation name and type, communicative goal, discourse focus and so on.

The Face and Body Animation interprets the APML-tagged dialogue move and decides how to convey every meaning (by which combination of signals). As mentioned previously, the Body we use at present is a combination of a 3D face model compliant with the MPEG-4 standard [12] and a speech synthesiser[2] [1].

3. A MARKUP LANGUAGE FOR BEHAVIOR SPECIFICATION: APML

We have developed XML-based languages, that include high-level primitives for specifying behaviour acts similar to those performed by humans, in order to control easily the behavior of ECAs.

An effort in building a standard in this direction is represented by the Human Markup Language [8]. This language allows one to specify human communicative behaviors at a very high-level. The aim of HML is to "develop Internet tools and repository systems which will enhance the fidelity of human communications" [8]. Its specification modules include tags allowing the representation of physical, cultural, social, kinetic, psychological, and intentional features used by humans in communicating information.

HML is a language at a very abstract level: using it for controlling specific agent bodies may be difficult and may require developing complex interpreters to translate a very abstract specification into low-level body actions. For this reason, researchers tend to develop their own languages, more suited to the type of embodied agent they wish to control. For instance, Virtual Human Markup Language, VHML [18] encompasses a large number of tags related to

[1]IST Project IST-1999-29078, partners: University of Edinburgh, Division of Informatics; DFKI, Intelligent User Interfaces Department; SICS; University of Bari, Dipartimento di Informatica; University of Rome, Dipartimento di Informatica e Sistemistica; AvatarMe.

[2]We are currently working toward the establishment of XML tags for intonation specification

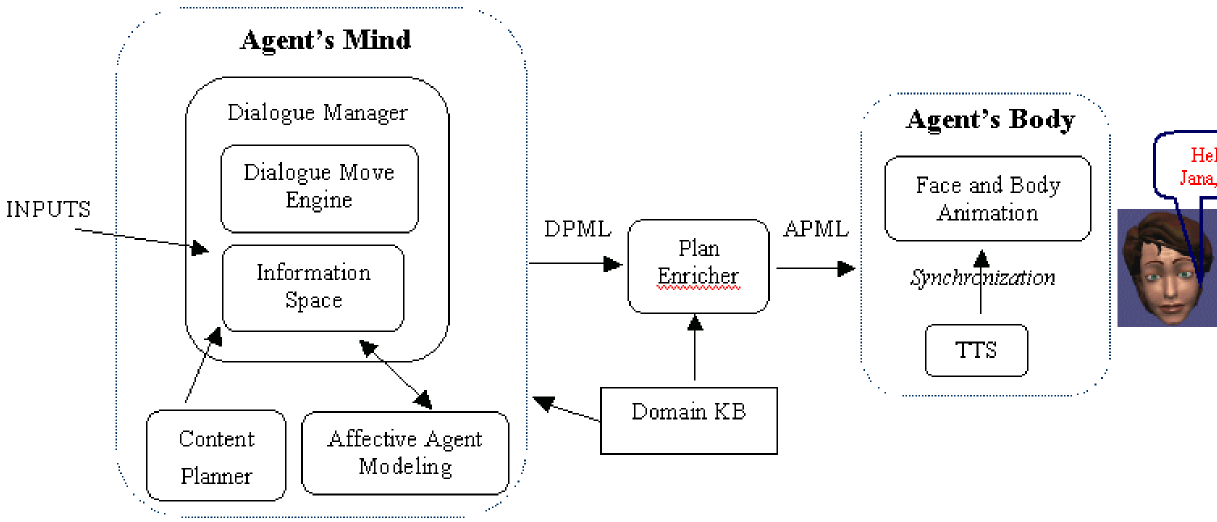

Figure 1: MagiCster system architecture

synthetic agents: some tags are related to facial expressions, to gesture, to emotion but also to dialogue management, synthesis speech and so on. The language offers tags on several levels: from tags representing the right raised eyebrow to the tags representing the emotion 'happiness'.

MPML (Multimodal Presentation Mark-up Language) has been developed with the aim of enabling authors of Web pages to add agents for improving human-computer interaction [9]. The design of this language has been driven by the choice of Microsoft Agent as a body. For instance, the tag for specifying a predefined animation sequence (<act>) takes, as a possible value, one of the MS-agent's animations.

Another XML language that was designed for generating embodied agent's behavior has been defined in BEAT [3]. In this system, the XML language is used for tagging both the agent's input and its output. The input is an utterance that is parsed into a tree structure; this tree is manipulated to include information about non-verbal signals and then serialised again in XML. The output language, specifying the agent's behavior, contains tags describing the type of animation to be performed and its duration.

As for BEAT, APML has been designed to represent the communicative functions. Poggi et al. [16] defined a communicative function as a (meaning, signal) pair, where the meaning item corresponds to the communicative value of the signal item. For instance, a smile can be the signal of a "joy" emotion or of a back-channel. This distinction between the meaning and the signal, i.e. the way in which the meaning can be communicated, has driven the design of APML. Due to the architectural choice of Mind-Body separation, tags should not specify the signal to be conveyed but only the meaning associated to a given communicative act. This dichotomy meaning/signal allows us to create different types of agents. Agents with different personality, gesture style, or even culture may be represented [6, 17] by associating to each meaning of the communicative functions different signals. An agent is therefore defined by a set of pairs (meaning, signal) that may be stored in a library. We are currently working on associating movement characteristics to specify how signals get instantiated [17]. A nervous person will have expressions associated to communicative acts that appear and disappear really fast (i.e. the expressions will be defined by short onset and offset) while a gracious person won't. In the next section 4 we explain the language we are using to describe facial expressions.

APML - Affective Presentation Markup Language - DTD is:

```
<!ELEMENT APML (turn-allocation+, perfomative*,
                turn-allocation*)>
<!ENTITY %TA-TYPE "(take|give)">
<!ENTITY %P-TYPE "(inform|ask|greet|request|)">
<!ENTITY %BR-TYPE "(adj|ElabObjAttr|ElabGenSpec|
                    justification|motivation|)">
<!ENTITY %A-TYPE "(joy|sorry-for|distress|)">
...
<!ELEMENT turn-allocation (performative*)>
<!ATTLIST turn-allocation type %TA-TYPE #REQUIRED>
<!ELEMENT belief-relation (#PCDATA|performative)>
<!ATTLIST belief-relation type %BR-TYPE #REQUIRED>
<!ELEMENT performative ((adjectival|deictic)*,
                         belief-relation*)>
<!ATTLIST performative type %P-TYPE #REQUIRED
                     affect %A-TYPE #IMPLIED
                     certainty %C-TYPE #IMPLIED>
<!ELEMENT adjectival (#PCDATA)>
<!ATTLIST adjectival type %ADJ-TYPE #REQUIRED>
<!ELEMENT deictic (#PCDATA)>
<!ATTLIST deictic obj CDATA #REQUIRED>
...
```

we are showing here the DTD instead of the XML-Schema for space reasons, since Schemas have a less compact representation than DTDs. The first part of the DTD defines the enabled values for the tag attributes while the second part specifies tags and their nesting in the definition of a valid APML structure. Every dialogue turn specified with this language starts with the root tag <APML>. To indicate that the agent is taking or giving the initiative, the turn-allocation tag is used: its type attribute can take the value "take" or "give".

The tag Performative may have up to two attributes on top of its own value-type:

- type: the performative type, that can take any of the values specified in P-TYPE domain (e.g. suggest, criticise),
- affect: an emotion in the A-TYPE set,
- certainty: of what is being communicating.

These attributes represent the communicative function of 'emotion' and 'certainty'. In our view [15], some communicative functions modulate the meaning of the performative per se, and they usually span the whole communicative act as the performative does. Other communicative functions modulate a single word or semantic element of the utterance and usually last only the time of that word or semantic element. The affective and the certainty functions act on the whole phrase of a discourse as does the performative and, therefore, may be represented as attributes to the performative tags; while the other communicative functions (adjectival, belief relation, emotional emblem [7] and so on) have a more local property as they act on the word(s) they refer to and would correspond to separate tags. The <belief-relation> tag takes as type attribute the name of the RR. The <adjectival> tag specifies that a specific word plays an iconic function. Eye aperture (greater or smaller) may mimic the size of an object, the importance of a person, etc. The <deictic> tag indicates a reference to objects, persons having a specific position in the domain space: the agent may refer to them by using pointing gestures, by gazing at them etc.

4. FACIAL DESCRIPTION LANGUAGE

Humans are very good at showing a large spectrum of facial expressions; but at the same time, humans may display facial expressions varying by very subtle differences, but whose differences are still perceivable. We have developed a language to describe facial expressions as (meaning, signal) pairs. These expressions are stored in a library. When the planner enriches the discourse move with a communicative meaning, the program looks in the library to which signals it corresponds and the APML tag gets instantiated by the corresponding signals values. Defining facial expressions using keyword such as 'happiness, raised eyebrow, surprise' does not capture these slight variations. In our language, an expression may be defined at a high level (a facial expression is a combination of other facial expressions already pre-defined) or at a low level (a facial expression is a combination of facial parameters). The low level facial parameters correspond to the MPEG-4 Facial Animation Parameters (FAPs) [12]. The language allows one to create a large variety of facial expressions for any communicative functions as well as the subtleties that distinguish facial expressions. Paradiso et al [11] have established an algebra to create facial expressions. The authors have elaborated operators that combine and manipulate facial expressions. Our language has the only purpose to create facial expressions that are associated to a given communicative functions. We have worked out a method to combine facial expressions due to

Figure 2: The combination of "left raised eyebrow" (left) and "right raised eyebrow" (centre) produces a raised "eyebrow" movement (right)

distinct co-occurring communicative acts using a Bayesian Network [15].

We consider two items: "facial basis" (FB) and "facial display" (FD). An FB is a basic facial movement such as right raised eyebrow, upper lip raise, jaw opening, left upper eyelid lowered and so on. FBs include also eye and head movements such as nodding, shaking, turning the head and the eyes. An FB may be represented as a set of MPEG-4 compliant FAPs or recursively, as a combination of other FBs using the '+' operators:

$$(1) FB = \{fap3 = v_1, \ldots\ldots\ldots, fap69 = v_k\};$$

where v_1, ... ,v_k specify the intensity value for the FAPs 3-68[3].

$$(2) FB' = FB_1 + FB_2;$$

where FB_1 and FB_2 may be defined in turn as a set of FAPs or FBs.

Head and eye directions and movements are defined separately. The head direction is specified as:

$$FB = \{head_h = d_h, head_v = d_v, head_t = d_t\};$$

where $head_h$, $head_v$ and $head_t$ are, respectively, the pitch, yaw and roll angles of the head. The values d_h, d_v and d_t specify the head direction in degrees along the three axis.

The eyes direction may be defined similarly.

FBs corresponding to head movements such as nodding and shaking may be represented by:

$$FB = \{head_m = (nod|shake), amp = d, period = p\};$$

where $head_m$ can be "nod" or "shake", amp is the amplitude of the movement (in degrees) and "period" represents the duration of a single nod/shake cycle.

For example *raised eyebrow* may be defined as a combination of the *left* and *right* raised eyebrow:

$$raised\_eyebrow = left\_raise\_eyebrow + right\_raise\_eyebrow;$$

where *left_raise_eyebrow* and *right_raise_eyebrow* are defined, respectively, as:

[3]we do not consider the FAP 1 and FAP 2 associated, respectively, to the six basic expressions and visemes as we are generating our own expression of emotion as well as visemes [14]

Figure 3: The "raised eyebrow" expression (left) and its more intense equivalent (right)

$$left\_raise\_eyebrow = \{fap31 = 50, fap33 = 100, fap35 = 50\};$$

$$right\_raise\_eyebrow = \{fap32 = 50, fap34 = 100, fap36 = 50\};$$

Figure 2 illustrates the resulting *raise_eyebrow* FB. The operator '*' increases or decreases the intensity of a single facial basis:

$$FB' = FB * c;$$

Where FB is a "facial basis" and 'c' is a constant. For example if we want a more intense raised eyebrow (figure 3):

$$large\_raise\_eyebrow = raise\_eyebrow * 2;$$

4.1 Facial Displays

Every facial display (FD) is made up of one or more FBs:

$$FD = FB_1 + FB_2 + FB_3 + \ldots\ldots + FB_n;$$

For example we can define the 'surprise' facial display as:

$$surprise = raised\_eyebrow + raised\_lid + open\_mouth;$$

We can also define a FD as a combination of two or more (already) defined facial displays using the '+' and '*' operators. For instance the "worried" facial display is a non-uniform combination of "surprise" (slightly decreased) and "sadness" facial displays (figure 4):

$$worried = (surprise * 0.7) + sadness;$$

5. AN EXAMPLE

In this section, we derive an example taken from a medical domain application. The User may dialogue with the agent and ask her about his physical state. The Dialogue Manager (DM) elaborates a discourse plan by consulting the domain model. This plan is then enriched by the plan enricher that translates a DPML-based tree-structure into an APML-based structure.

Let us suppose, the User is asking about the severity of his disease. The DM selects the following dialogue move which DPML recipe is:

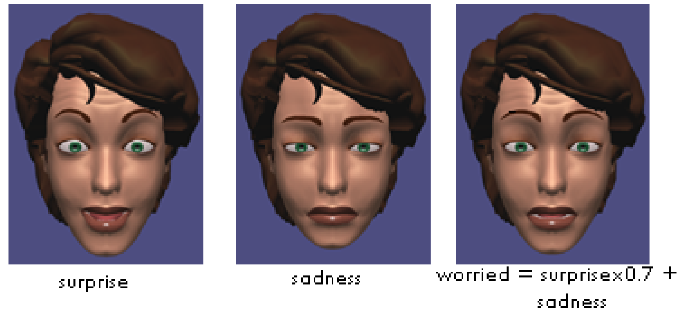

Figure 4: The combination of "surprise" (left) and "sadness" (centre) produces a "worried" facial display (right)

```
<node name="n1" goal="Explain(Has(U,disease))"
 role="sat"focus="disease"RR="ElabObjAttr">
<node name="n2" goal="Inform(Has(U, disease))"
 role="nucleus" focus="Has(U, disease)" RR="null"/>
<node name="n3"goal="Inform(Severity(disease))"
 role="sat" focus="Severity(disease)" RR="null"/>
</node>
```

Every dialogue turn of the agent starts with the turn-allocation function that indicates that she takes the initiative. Therefore, after the root tag <APML>, the <turn-allocation> tag is generated by setting up its type attribute to "take". Every RR attribute is transformed into a <belief-relation> whose type attribute is set with the name of the RR; every leaf node is transformed into a <performative> element. If Mind establishes that an emotion is felt by the agent in correspondence with the current node and that this emotion has to be displayed [2], the affect attribute of the performative tag is set to that emotion's name. The surface realization of the leaf node, corresponding to the text within the <performative> tags, is made by a generation function that produces the verbal part of the speech act and includes, if needed, two more tags: the <adjectival> one (when the argument of the communicative goal is a quantitative attribute of the discourse focus) and the <deictic> one (when the argument of the communicative goal is described in the domain knowledge base as 'reference-able through its coordinates'). In the example, the "severity" of the disease is a quantitative property of angina, which is the discourse focus: therefore, the <adjectival> tag is generated around the attribute-word "mild" (instantiation of the severity of the angina). The output of the plan enricher is the APML annotated text:

```
<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="ElabObjAttr"> which
<performative type="inform" certainty="uncertain">
appears to be <adjectival type="small">mild.
</adjectival> </performative> </belief-relation>
</turn-allocation></APML>
```

The APML tags get instantiated by their facial signals by

Figure 5: Expression of ‘sorry-for’, ‘certain’ and combination of both expressions with conflict resolution

looking up in the library associated to the agent. In this example “certain” corresponds to a frown and the tag “sorry-for” to the signals: inner eyebrow up, head aside, mouth corner down. There is a conflict in the eyebrow region. Our system resolves the conflict that may occur when more than one communicative functions span over the same text [15]. The conflict resolution uses a Bayesian Network that takes one or many communicative functions as input and output the final combined expressions. The final expression may contain the meanings of all the communicative functions, creating an expression of complex meaning. Figure 5 illustrates how the frown of “certainty” gets integrated within the facial expression of “sorry-for”.

6. CONCLUSIONS

In this paper, we have described the architecture of the behaviour generator of a believable conversational agent. In particular, we focused our discussion on the importance of Mind-Body separation and therefore on the need of an interface between the two modules. These interfaces should be able to represent the communicative functions that can be potentially realized by different bodies with different expressive capabilities. We have defined two XML-like mark up languages to represent the Mind’s output, i.e. the discourse plans, (called DPML) and the Body’s input (named APML): We have also described how a plan enricher transforms DPML trees into APML trees. Finally we have presented our language to generate subtle facial expressions.

7. ACKNOWLEDGEMENTS

We wish to acknowledge Fiorella de Rosis for her useful suggestions.

8. REFERENCES

[1] A.W. Black, P. Taylor, R. Caley, and R. Clark. Festival. http://www.cstr.ed.ac.uk/projects/festival/.

[2] N. De Carolis, C. Pelachaud, I. Poggi, and F. de Rosis. Behavior planning for a reflexive agent. In *IJCAI’01*, Seattle, USA, August 2001.

[3] J. Cassell, H. Vilhjálmsson, and T. Bickmore. BEAT : the Behavior Expression Animation Toolkit. In *Computer Graphics Proceedings, Annual Conference Series*. ACM SIGGRAPH, 2001.

[4] B. DeCarolis. Generating mixed-initiative hypertexts: a reactive approach. In *Proceedings of the International Conference on Intelligent User Interfaces*, Redondo Beach, CA, 5-8 January 1999. ACM Press.

[5] F. deRosis, C. pelachaud, N. De Carolis, V. Carofiglio, and I. Poggi. Modeling the dynamics of affective states in a conversational embodied agent. *Special Issue on “Applications of Affective Computing in Human-Computer Interaction”, The International Journal of Human-Computer Studies.*, to appear.

[6] F. deRosis, I. Poggi, and C. pelachaud. Tanscultural believability in embodied agents: a matter of consistent adaptation. In R. Trappl and P. Petta, editors, *Agent Culture: Designing Virtual Characters for a Multi-Cultural World*. Kluwer, to appear.

[7] P. Ekman. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, editors, *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*, pages 169–248. Cambridge University Press, Cambridge, England; New-York, 1979.

[8] HumanML. Human markup language. http://www.humanmarkup.org.

[9] M. Ishizuka, T. Tsutsui, S. Saeyor, H. Dohi, Y. Zong, and Helmut Prendinger. MPML: A multimodal presentation markup language with character agent control functions. In *Achieving Human-like Behavior in Interactive Animated Agents, Proceedings of the AA’00 Workshop*, pages 50–54, Barcelona, 2000.

[10] S. Larsson, P. Bohlin, J. Bos, and D. Traum. *TRINDIKIT 1.0 manual for D2.2*. http://www.ling.gu.se/projekt/trindi.

[11] A. Paradiso and M. L’Abbate. A model for the generation and combination of emotional expressions. In *Multimodal Communication and Context in Embodied Agents, Proceedings of the AA’01 workshop*, Montreal, Canada, May 2001.

[12] C. Pelachaud. Visual text-to-speech. In Igor S. Pandzic and Robert Forchheimer, editors, *MPEG4 Facial Animation - The standard, implementations and applications*. John Wiley & Sons, to appear.

[13] C. Pelachaud, V. Carofiglio, B. De Carolis, and F. de Rosis. Embodied contextual agent in information delivering application. In *First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS)*, Bologna, Italy, July 2002.

[14] 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.

[15] C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, To appear.

[16] I. Poggi, C. Pelachaud, and F. de Rosis. Eye communication in a conversational 3D synthetic agent. *AI Communications*, 13(3):169–181, 2000.

[17] Z. Ruttkay and C. Pelachaud. Exercises of style for virtual humans. In *Symposium of the AISB’02 Convention*, volume Animating Expressive Characters for Social Interactions, London, Avril 2002.

[18] VHML. Virtual human markup language. http://www.vhml.org.