Influences on Embodied Conversational Agent's Expressivity: Toward an Individualization of the ECAs

Vincent Maya, Myriam Lamolle, Catherine Pelachaud

Influences on Embodied Conversational Agent's Expressivity: Toward an Individualization of the ECAs

Vincent Maya
LINC
IUT of Montreuil
University Paris VIII
maya@iut-univ.paris8.fr

Myriam Lamolle
LINC
IUT of Montreuil
University Paris VIII
m.lamolle@iut-univ.paris8.fr

Catherine Pelachaud
LINC
IUT of Montreuil
University Paris VIII
c.pelachaud@iut-univ.paris8.fr

Abstract

We aim at creating not a generic Embodied Conversational Agents (ECAs) but an agent with a specific individuality. Our approach is based on different expressivities: the agent's expressivity, the communicative of behavioral expressivity. Contextual factors as well as factors such as culture and personality shape the expressivity of an agent. We call such factors "influences". Expressivity is described in terms of signals (e.g. smile, hand gesture, look at) and their temporal course. In this paper, we are interesting in modelling the effects of influences may have in the determination of signals. We propose a computational model of these influences and of the agent's expressivity. We have developed a taxonomy of signals according to their modality (i.e. face, posture, gesture, or gaze), to their related meaning and to their correspondence to expressivity domains (the range of expressivity than they may express). This model takes also into account the signals dynamic instantiation, i.e. the modification of signals to alter their expressivity (without modify the corresponding meaning).

1 Introduction

We aim at creating an **E**mbodied **C**onversational **A**gent (ECA) that would exhibit not only a consistent behavior with her personality and contextual environment factors but also that would be defined as an individual and not as a generic agent. Most of the agents that have been created so far are very generic in their behavior type. We want to simulate that two different agents may behave differently in a same context and may express the same felt emotion differently, even if they belong to the same social-cultural sphere. Given a goal in mind, people differ in their manner of expressing themselves.

Several studies have shown the importance to consider complex information such as cultural factors, personality, environment setting when designing an agent (Brislin (1993), H. Morton and Jack (To appear), Isbister and Nass (Forthcoming) and Lee and Nass (1998)). These factors affect the interaction a user may have with the agent. Personality also is an important aspect that makes people look and act very differently. Gender, age (child vs. teenager), our social role (e.g., mother, doctor), our experience and memory intervene in the manner we interact with others, we talk about things. These differences arise at different levels: the formulation of our thought as well as their expression. They have influences on the surface generation and realization level during the dialog phase as well as on the selection of the non-verbal behaviors and of their expressivity (Ruttkay et al. (2003)).

At first we propose a taxonomy of the influences types. These influences act on what to say and when as well on how to say it and to express it. In this paper, we concentrate on the effects of the influences on the facial, gaze, gesture and body behaviors but we do not consider the modification of the speech (choice of word variation of paralinguistic values, of intonation tones). However we assume that the input text embeds these effect. Expressivity acts not only on the selection of a non-verbal behavior to convey a meaning but also on its *expressivity*, i.e. the *strength* of this non-verbal behavior. For example, in order to express the surprise, the agent may raise her eyelids, and the more the surprise is strong, the more the eyelids are up.

We model expressive effects within our framework based on **APML**, (Affective Presentation Markup Language, see DeCarolis et al. (2004)), based on a taxonomy of communicative functions proposed by I. Poggi (see Poggi et al. (2000)), and the ECA **GRETA**, showed in figure 1 and figure 7 (see Pelachaud and Poggi (2002)), we model their effects. From a tag that indicate only the meaning the agent has to express and its expressivity, we want to obtain a tag that indicate the signal to use, among all that are stored in libraries, and the technical value allowing the system to modulate this signal, that we call the dynamic instantiation coefficient.

In this paper, after presenting a state of art in section 2, we describe a taxonomy of influences in section 3. We then define what we mean by expressivity in section 4 and

Figure 1: Greta

the agent's model in section 5. In section 6, we describe our normalization tools allowing one to associate expressivity and dynamic instantiation information to signals, in order to use the model proposed in this paper. Finally we provide an overview of our APML translator.

2 State of the art

Agents exhibiting emotional behaviors have received quite some interest. In Ball and Breese (2000), the authors developed a model in which the emotion an agent is undergoing may affect her verbal and non-verbal behavior. They built a Belief Network that links emotion with verbal and non-verbal manifestation. Fiorella de Rosis and her colleagues have developped a computational model of emotion triggering using a dynamic Belief Network (Carofiglio et al. (in press)). Their model is able to determinate not only which emotion is triggered after a certain event for a given agent but also it is able to compute the variation of this emotion over time: this emotion may increase or decrease in intensity or it may also evolve in another emotion. The computational model uses a Belief Desire Intention (BDI) model of the agent's mental state. Fuzzy logic has been used to either model the triggering of emotions due to events (El-Nasr et al. (2003)) or to map facial expressions of an emotion with a given intensity (Bui et al. (2003)).

Emotions have also been considered during the interaction of a user with a system. Within the EU project Safira (Höök (To appear)), a new interaction device have been developed to interact with actors of a video game. This device is SenToy, a teddy bear with sensors attached to its joints (Paiva et al. (2003)). The user moves around the toy using a set of pre-defined moves with a given expressivity. These emotional behaviors are detected. When recognized by the system they are used to drive the behavior of one of the agents in the video game.

Several talking heads able to show emotions have been developed. In particular, in Kshirsagar et al. (2001), the authors have developed an agent that is able to react to the user's emotion detected through his facial expressions. This reaction is based on a computational model of emotion behavior that integrates a personality model. Carmen's Bright IDEAS (Marsella et al. (2000)) is an interactive drama where characters exhibit gestures based on their emotional states and personality traits. Through a feedback mechanism a gesture made by of a character may modulate her affective state. A model of coping behaviors have been developed by Marsella and Gratch (2003). The authors propose a model that embeds information such as the personality of the agent, his social role.

In an attempt to model cultural behavior for a talking head, King et al. (2003) have proposed a simple model using a table of correspondence between a given meaning and its associated behaviors. Scott King built such a table for each culture he considered (English and Maori). We are aware of very few other attempts.

The role of social context in an agent's behavior have been considered. DeCarolis et al. (2001) propose a model that decides whether an agent will display or not her emotion depending on several contextual and personality factors. Prendinger et al. (2002) integrate contextual variables, such as social distance, social power and thread, in their computation of the verbal and non-verbal behavior of an agent. They propose a statistical model to compute the intensity of each behavior. Rist and Schmitt (2003) modelled how social relationship and attitudes toward others affect the dynamism of an interaction between several agents.

To control the behavior of ECAs, several representation languages has been developed. Theses languages specify the agent's behaviors. They serve as interface between the different modules of the architecture of an agent system. The languages may embed various levels of abstraction: ranging from the description of the signals (a smile, a head nod) in VHML (VHML) to semantic information (rheme/theme, iconic) in Piwek et al. (2002), going through communicative function (performative, emotion) (DeCarolis et al. (2004)). Of particular interest to our work is the language SCRipting Emotion-

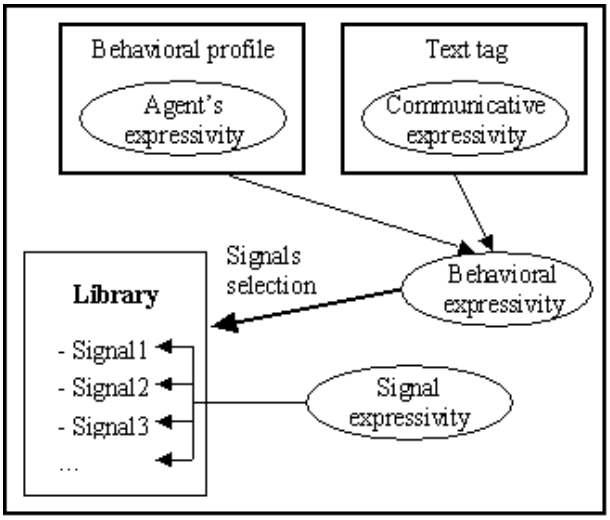

Figure 2: Expressivity Types

based Agent Minds (SCREAM). SCREAM has been designed to create emotionally and socially appropriate responses of animated agents placed in an interactive environment (Prendinger et al. (2002)).

The work that is more related to our is Ruttkay et al. (2003). The authors aim at creating agents with style. The authors aim at creating agents with style. They developed a very complex representation language based on several dictionaries. Each dictionary reflects an aspect of the style (e.g. cultural or professional characteristics or personality). They also defines the association meaning to signals. In this language the authors embed notions such as culture, personality, gender but also physical information such as gesturing manner or tiredness. To create an agent with style one needs to select a set of values (e.g. an Italian extrovert professor). The proper set of mappings between meanings and signals is then instantiated. The authors modelled explicitly how factors such as culture and personality affect behaviors. We distinguish our work from them in the sense that we do not modelled such factors, rather we modelled the different types of influences that may occur and how they may modulate an agent's behaviors.

3 Taxonomy of Influences

We call "influence" different factors: contextual factors as well as factors such as culture and personality. These factors shape the expressivity of an agent. Influences may act on the selection of a non-verbal behavior to convey a meaning (i.e. on the choice of the signals), on the expressivity of this behavior (e.g. on their intensity level), in order to lessen it or to accentuate it and on the communication strategies.

We differentiate three types of influences. The first type contains the *intrinsic* influences. We consider that each human being has a set of the conscious and unconscious habits that are reflected in the content of her discourse and that define her attitude and her behavior when she talks. These habits derive, amongst others, from her personality, her age, her sex, her nationality, her culture, her education and her experiences (Brown and Nichols-English (1999)). For example, some non-verbal behaviors are very culturally dependent, such as the emblems, gestures that may be directly translated into words. This might be the case with some iconic or metaphoric gestures that may take their origins in the culture (the action of eating will not be represented identically if one is used to eat with a fork or with sticks). However, not all gestures are culturally dependent. The type of gestures one makes while conversing might not differ over various cultures as much as one would have thought at first. The main discrepancy is more in the quantity of gestures made rather than on the type of gestures itself (Cassell (2000)).

The second type contains the *external* influences, that refer to the environment setting, such as the light conditions, the sound intensity, the spatial layout or the function of the conversation site. These factors may affect some speech and behavior characteristics. For instance, in a crowed room, some wide gestures are simply impossible. Likewise, in a noisy room, a person has to increase the volume of her voice due to pragmatic considerations. She will also be inclined to amplify her gestures. In the opposite, in a religious building or in a museum, a person ought to remain quiet and move silently due to social considerations. Another factor of influences is how the agent is placed in the environment; a person sitting in an armchair will not gesticulate as a person standing in a hallway.

The third type contains the *mental* and *emotional* influences. The mental state of the agent affects greatly the way the agent will behave: it modifies the prosody of speech, the amplitude of a facial expression, the movement tempo. A person does not talk and does not behave the same way whether she is angry or not. Her relationships with her interlocutor modulate also her behavior: she does not behave in the same way with a friend, and unknown person, an employee, a child or a doctor (DeCarolis et al. (2002)). The agent's mental state evolves all along the conversation. Her emotion varies through time, her goals and beliefs get modified as the conversation evolved.

In this paper, we oppose the *intrinsic* influences to the other ones, which we group in the *contextual* influences. The intrinsic factors are constant during a dialog session, whereas the contextual ones may vary. The contextual factors increase or decrease the effects of the intrinsic factors, or even cancel them.

```
 1. <librarySignal
 2.  format = "FAP">
 3.  <signal name = "smile"
 4.     fap4 = "-100" fap8 = "-100"
 5.     fap9 = "-100" fap51= "-100"
 6.     fap55= "-100" fap56= "-100"
 7.     fap12=  "150" fap13=  "150"
 8.     fap59=  "150" fap60=  "150"
 9.     fap6 =   "50" fap7 =   "50"
10.     fap53=   "50" fap54 =  "50"
11.  />
12.  <signal name = "joy_eyelids"
13.     fap19=  "128" fap20=  "128"
14.     fap21=   "64" fap22=   "64"
15.  />
16.  <signal name = "joy_1"
17.     combination = "smile"
18.     combination = "joy_eyelids"
19.  />
20.
... </librarySignal>
```

Figure 3: Example of signals library

4 Expressivity

We call expressivity the value that allows the system to relate *strength* to the communication act.

We do not aim at modelling what culture or personality mean, nor do we aim at simulating expressive animations. We limit our scope at representing influences that would modify the set of behaviors and the quality motions a particular agent will display to communicate a given meaning within a specific context.

According to the concepts they are applied to, we differentiate several expressivities, schematized in figure 2.

4.1 Communicative expressivity

The input text is marked with tags specifying the communicative function the agent aims at displaying. Each tag may have an attribute corresponding to the degree of expressivity attached to a given meaning for an agent. Is the agent puzzled or completely confused, slightly angry or madly angry? In figure 9, at line [10], the value of *communicative expressivity* related to the joy of the agent is *0.8*.

4.2 Agent's expressivity

Agent's expressivity is related to the qualitative property of behavior. Does an agent has the tendency to play down (acts so as not to be noticed) or on the opposite wants to catch all looks by acting wild? This value is given in the agent's definition. In figure 6, this expressivity is described by the lines [10] to [16].

```
 1. <library
 2.  modality = "face">
 3.  <expression meaning = "joy"
 4.    name = "smile"
 5.    ref = "0.4"
 6.    min = "0.3" max = "0.6"
 7.    minCoef = "0.9" maxCoef ="1.5"
 8.    dynInstType = "duration"/>
 9.  <expression meaning = "joy"
10.    name = "eyelids_joy"
11.    ref = "0.2"
12.    min = "0.1" max = "0.4"
13.    minCoef = "1.0" maxCoef ="1.0"
14.    dynInstType = "amplitude"/>
15.  <expression meaning = "joy"
16.    name = "joy_1"
17.    ref = "0.7"
18.    min = "0.7" max = "0.9"
19.    minCoef = "0.8" maxCoef ="1"
20.    dynInstType = "amplitude"/>
21.  <expression meaning="anger"
22.    name ="anger_mouth"
...    ...
... </library>
```

Figure 4: Example of expression library

4.3 Behavioral expressivity

The behavioral expressivity represents the way that the considered agent expresses the tag meaning, taking into account her characteristics and the contextual factors that may modify her expressivity. It is the result of the computation from the *communicative expressivity*, the *agent's expressivity* and the contextual factors that influence the way that she expresses the communication act. It intervenes during the selection of the signal and during the computation of its *dynamic instantiation* (see section 6): it modifies the quantity of movements related to these signals, their amplitude, their duration, their dynamism and/or their repetitiveness.

4.4 Signals expressivity

In order to choose the appropriate signals that best corresponds to a given meaning, the system has to know the expressivity related to each signal. For example, it has to know that mild smile is less expressive than a large smile. *Signal libraries* (see figure 3) contain *basic signals*(*smile*, from the line [3] to the line [11]), and *hight level* signal (i.e. defined as a combination of *basic* signals such as the signal *joy_1* define from the line [16] to the line [19]).

To be able to instantiate the *behavioral expressivity* into a set of expressive signals, the animation engine has to know the signals that are potentially available and to compute the appropriate *signal expressivity*. In order to integrate this *expressivity* type, we define *expression libraries*

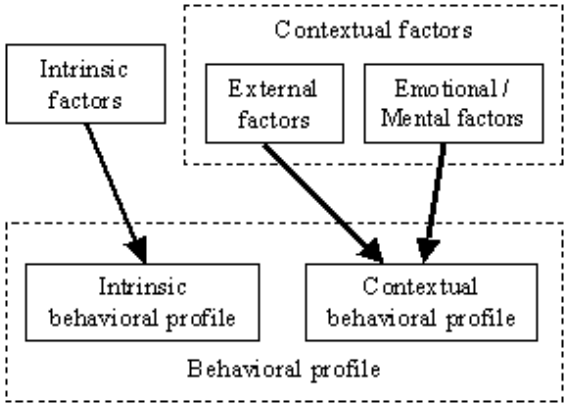

Figure 5: Agent's behavioral profile

```
 1. < agentDefinition
 2.    nameAgent        = "Agent1"
 3.    redundancyStrategy = "mono"
 4. >
 5.    <modHierarchy
 6.        face     = "1.0"
 7.        gesture  = "0.5"
 8.        position = "0.3"
 9.        gaze     = "0.4" />
10.    <intrinsicfactors
11.        face     = "0.4"
12.        gesture  = "0.4"
13.        posture  = "-0.2"
14.        <!--gaze     = "0"-->
15.    />
16. </agentDefinition>
```

Figure 6: Agent's definition

(see figure 4). These libraries contain for each communicative act a list of (meaning, signal) pairs. They also contain information about *expressivity domain* of these signals and about the way to modulate their expressivity. These *expressions libraries* have also the advantage to be independent of the format of the *signals libraries*.

To indicate the expressivity associated to signals, we use fuzzy set that define domains where the signals are appropriately usable. The attribute *ref* represents the expressivity associated to the related signal. The attributes *min* and *max* represent respectively the minimal and the maximal expressivities that can be associated to the signal. We assume that the attributes *min* and *max* take into account the signal distortion possibilities; that is the modulation of the signal does not change the meaning associated to it. The *expression libraries* indicate also the extreme distortion coefficients *coefMin* and *coefMax* to compute the distortion coefficient related to a given expressivity (see section 6.3).

Behavior expressivity may be expressed not only through the signals, and their expressivity, but also by combination of signals dispatched over modalities. We differentiate the *modal signal expressivity*, which concerns the qualitative parameters that determine the choice between several signals of a same modality with a same meaning, and the *inter-modal signal expressivity*, which is modelled by defining the functions that relates the behaviors across the modality, such as redundancy (i.e. expression of the same meaning with several signals of different modalities), complementarity (e.g. saying "*he goes to the stadium*", and complementing it with an iconic gesture that means "*he drives to the stadium*"), substitution (e.g. straight index over the mouth to mean silence), and masking (e.g. masking sadness by a smile).

5 Agent's Model

In previous section, we have shown that even two agents may have in mind a same communicative act, they may behave differently, using various expressivity or using different modalities. In this section, we feather refine our model by specifying the agent's preferences to use a given modality (e.g. an agent may have a very expressive face). We also model how the agent dispatches her behavior over different modalities.

In our model, the information that allows the system to obtain these results is described within the tag $< agentDefinition >$ (see figure 2). In this section, we describe the different elements of this tag.

5.1 Intrinsic behavioral profile

In figure 2, we associate to the agent a behavioral profile, which specifies, on the one hand, the agent's expressivity, i.e. the intrinsic factors, and, on the other hand, the effects of the contextual factors. This profile specifies the agent's expressivity depending on the modalities. It allows one to define that an agent has a very expressive face or that she rarely uses wide arm movements.

These intrinsic factors is described in the element $< intrinsicFactors >$. It associates a numeric value to the attributes *face*, *posture*, *gaze* and *gesture*. These values lessen or accentuate the expressivity of the tag meaning for the related modality. According to the description of *Agent1* in figure 6, this agent is more expressive for the face and for the gestures than the *default agent* (i.e. facial moves and her gestures are more accentuated than the *default agent's* ones but less for the posture. At line [14] the default value of the gaze attribute is 0 meaning the agent does not use this modality to communicate this specific meaning. This intrinsic profile, given aa input, is constant during a dialog session.

Figure 7: Face variation of Greta

5.2 Modality Hierarchy

In order to choose the modality that the agent will use for a given tag, we define a priority scheme on the behavior. We associate to each modality (face, gaze, gesture and posture) a numeric value that represents their preferential level in the hierarchy.

In case several modalities have the same hierarchical level the system considers the expressivity of all the signals of the concerned modalities to choose a signal for this level. In the agent's definition (see figure 6), the lines [6] to [10] describe this hierarchy. According to this description, *Agent1* uses mainly facial expressions.

5.3 Inter-modal functions

From the *communicative expressivity*, the system obtains a *behavioral expressivity*. The system computes how to express this expressivity (see algorithm described in section 6) according to the signals expressivity. This latter expressivity is defined in the expression libraries (see. figure 2). For a same meaning, several signals of different modalities may be associated (e.g. anger can be express with a frown thin lips, looking straight and tense movement). The *behavioral expressivity* is then related to how the signals are dispatched over all the modalities.

Substitution and complementarity modify the text content and are represented by tags (e.g. saying "*he goes to the stadium*", and complementing it with an iconic gesture that means "*he drives to the stadium*" for complementarity, or raising straight index over the mouth to mean silence, for substitution). Therefore, these tags must be defined in the input text.

Strategies for redundancy and masking, from a cognitive point of view as well as from a computational point of view, may vary according to the context or to the considered agent. For the moment, we consider that the masking is mainly contextual and we define its strategies in the section 6.1. Conversely, an agent may mainly employ a specific redundancy strategy. This information is given in the agent's definition (see line [3] in figure 6). We define several strategies :

- "*mono*" : only signals of one modality is used.

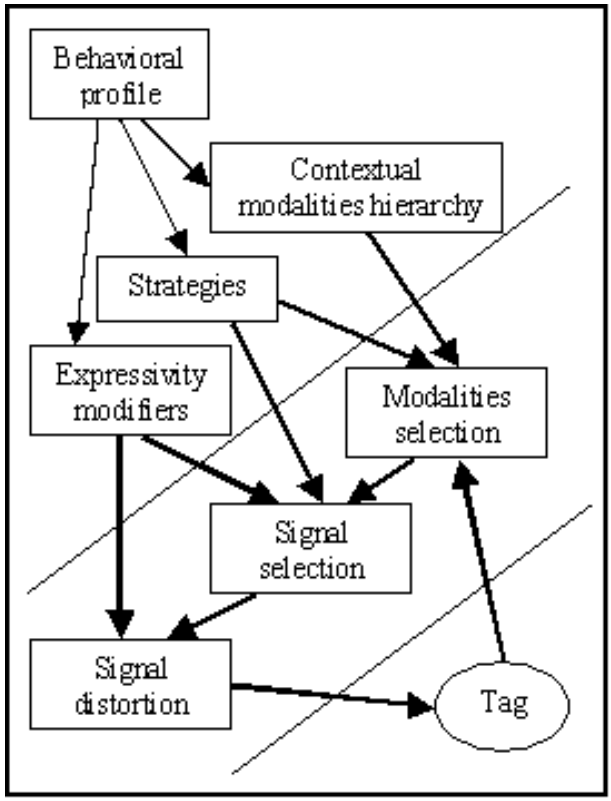

Figure 8: Tags transformation Steps

- "*maximal*" : signals of all possible modalities are selected.

- "*additional*" : redundancy is used only for the expressivity superior to a given threshold.

These strategies may be restricted a specific communicative act, such as "*certainty*" or "*performative*".

The redundancy, more than any other signal association, raises the problem of the coherence of the signals choice. Let us imagine that the system choose to express an emotion with facial and gestural signals. By the process described in the section 6, we obtain a facial signal inconsistent with the gestural signal. For example, in order to express redundantly the agent's anger, the system decides to use, on the one hand, signals related to the face crispation, and on the other hand, wide arms movements.By considering, for the sake of the example, that these signals are inconsistent, our system has to be able to determine this inconsistence (thanks to a *coherence library*) and to propose a substitution solution.

6 System overview

Our system takes as input a *intrinsic behavioral profile* that represents the agent's communication characteristics

```
 1. <agent
 2.    nameAgent       = "Agent1"
 3.    contextCoeff    = "0.5"
 4.   <!--MaskingStrategy =
 5.                  "multimodal"-->
 6. >
 7.   <performative type = "inform">
 8.      <rheme>
 9.        <affective type = "joy"
10.           expressivity = "0.8">
11.            I'm happy.
12.        </affective>
13.      </rheme>
14.   </performative>
15. </agent>
```

Figure 9: Example of input

(see figure 2 and figure 5) and a text with tags specifying communicative functions (Poggi (2003)).

The representation language used for the tags is the Affective Presentation Markup Language (APML) (DeCarolis et al. (2004)). A tag represents the meaning associated to a given communicative function. Most tags contain an expressivity attribute.In the *expression libraries*, each meaning is associated to a list of possible signals that may describe it.

From a given intrinsic behavioral profile, the system instantiates the tags into a set of signals that are then translated into animation parameters. During this instantiation phase, the process of the agent individualization is done in three steps: the modality selection, the signals choice and the signal dynamic instantiation (see figure 8).

6.1 Modalities selection

For each tag, the system has to decide the modality (*face*, *gesture*, *gaze* or *posture* one) to use to express the given meaning.

In most cases, the decision is based on the modality hierarchy: among the modalities that have at least one expression which allows the system to represent the meaning, it choose the one with the highest priority and that is not used yet, in order to prevent conflicts. Conflicts may occur for embedded tags acting on the same text span. These tags may use the same modalities for their corresponding signals. Conflicts may arise if the tags require to use the same modality. In case conflicts may not be solved using the behavior hierarchy scheme. We select one tag to prevail over the others. For gaze and gesture we choose the most embedded tag while for face we choose the outer tag (Poggi (2003)).

Some contextual factors may however modify this hierarchy. For example, for an agent that expresses her communication acts mainly by facial expression, the anger or the nervousness may incite her to use gestures more intensively.

In the tag $< agent... >$ in the input text (see figure 9), the value *mono* of the attribute ***redundancyStrategy*** indicates that *Agent1* does not use redundancy. Therefore the system selects the face modality to express *"joy"*, as this modality is not yet employed to express the performative *"inform"* (expressed by the gaze signal *"look_at"*.

The attribute *maskingStrategy* specifies the strategy to mask an expression by another one (see section 5.3). Agents may not mask their emotions in the same way and with the same efficiency than others. Masking strategy depends on the context and, in particular, on the relationships between the agent and its interlocutor. We define three strategies:

- "*attenuation*": the system attenuates the signals of the expressions, as to aim to show a neutral expression
- "*addition*": the system dissimulates the signal with a signal of an another modality, such as a hand in front of the mouth.
- "*replace*": the system replaces a signal by another one, such as a sad expression replaced by a polite smile.

6.2 Signals selection

6.2.1 Computation of the *behavioral expressivity*

First, in order to select the appropriate signals according to the influences, for the modality selected at the previous step, the system computes the *behavioral expressivity* for the desired modality. It sums up the *communicative expressivity* (given in the input text) to the value of the *intrinsic behavioral profile* related to the selected modality. This operation allows it to take the behavior of the agent for the modality into account. For example, the intrinsic behavioral may express that the agent is inclined to use wide gestures. The result of this operation is modified according to *contextCoeff*. This coefficient only aims to lessen or to accentuate the *behavioral expressivity*. This value is defined for all the text included between $< agent... >$ and $< /agent >$. The attribute *contextCoeff* expresses the effects of the contextual factors only at the level of the expressivity.

In the input text presented in figure 9, at the line [3], the attribute *contextCoef* indicates that the contextual coefficient *contextCoeff* is equal to *0.5*. This value may model for example a signal attenuation resulted on masking an expression by another one (e.g. in some situations, anger may not be shown and a polite smile may have to be displayed). Consequently, since its value is inferior to *1*, the expressivity of the tag "*inform*" and of the tag "*joy*" is lessened. In our example, the tag "*joy*" is associated to an expressivity of *0.8*. Considering that during the previous steps, the system has chosen to express the tag "*inform*" by the signal "*look_at*" of the gaze modality, it choose,

according to the modality hierarchy and to the presence of related signals in the different modalities, to use the facial modality to express the tag "*joy*". It adds to the *communicative expressivity* of the tag "*joy*", the *Agent1*'s facial expressivity, expressed by the attribute *face* of the element *intrinsicFactors* (see line [12] in figure 6). The value of this attribute is *0.4*. It obtains thus an intermediary expressivity of *1.2* (i.e. *0.8 + 0.4*). It applies then the contextual coefficient *contextCoeff* whose a value is *0.5*. The system outputs a *behavioral expressivity* e_{joy} of *0.6* (i.e. (*0.8*+*0.4*) x *0.5*).

6.2.2 Selection in the libraries

In the expression libraries, each signals description contains an *expressivity domain* (defined by a minimal and a maximal values) and a reference value. In the *expression library* described in figure 3, three signals set can represent the emotion related to the tag *"joy"*: the *basic* signals *"smile"* and *"joy_eyelids"*, defined independently of each other (Bui et al. (2003)), and the *hight-level* signal *joy_1*. The name of these signals allow the system to retrieve them in the related signals library (see figure 3).

The *behavioral expressivity* is compared to the expressivity domain of these signals, described in the *expression library* (see figure 4). We compare the value of the *behavioral expressivity* e_{joy} related to the tag *"joy"* with the boundary values *min* and *max* for each of these signals. We select the signal whose *domain* of expressivity contains the value e_{joy}.

If several signals can be selected, the system chooses according to the distance between e_{joy} and its reference expressivity or between e_{joy} and the nearest bound. There are several possible strategies. In our system, they are configurable in order to test their efficiency.

If e_{joy} does not belong to any domain (i.e. no signal with such an expressivity exists for this given meaning and this particular agent), the system chooses the expression with the nearest domain, and redefines the value of e_{joy} according to value of the nearest bound of this domain. In our example, the system selects the signal denoted "*smile*".

6.3 Signal dynamic instantiation

6.3.1 Computation of the dynamic instantiation coefficient

As seen in the previous section, the system obtains the name of the selected signal from the expression library, for a given modality and for a given expressivity. Now, it has to compute the dynamic instantiation to apply to this signal. This dynamic instantiation allows us to obtain the widest range of expressions and to modulate the expressivity. In our example, for an influence coefficient *contextCoeff* with a value of *0.4* instead of *0.5*, the system obtains a *behavioral expressivity* with a value of *0.48*. In this case, the system also uses the signal "*smile*", but *the dynamic instantiation* applied to the signals would have been able to make perceivable the difference.

To compute the dynamic instantiation *l'*, for a given *behavior expressivity e*, we consider:

- *l*: the distance between *ref* and *e*;
- *L*: the distance between *ref* and the appropriate boundary: *min* if *e* is inferior or equal to *ref*, *max* otherwise;
- *L'* the distance between *1* (i.e. the default coefficient, which indicates that the system has not to modify the signals stored in the library) and the coefficient related to the considered boundary (i.e. *coeffMin* or *coeffMax*).

The distance *l'* between the dynamic instantiation coefficient *dynCoeff* and *1* is such as that the ratio *l/L* is equal to the ratio *l'/L'*. Thus, $l' = L' * l/L$. The coefficient *distortCoeff* is superior to 1 iff *e* is superior to *ref*. In our example, for the signal "*smile*", as e_{joy} = *max*, we have *dynCoeff = coeffMax=1.5*.

For *contextCoeff* equal to *0.4*, we have said that the behavioral expressivity have a value of *0.48*. As the expressivity reference ref is *0.4* the *dynCoeff* is equal to *1.2* (i.e. *1+(1.5-1)*x(*0.48-0.4*)*/(0.6-0.4)*).

We consider that the evolution of the dynamic instantiation coefficient is linear between 1 (i.e. the default coefficient) and the extreme values, but not necessarily between the extreme values.

Given the behavioral profile and a specific meaning, the system computes the appropriate value of the signal dynamic instantiation using a fuzzy logic approach. We point out that we are dependent on the signals libraries content: for two different agents for which the system use the same modality and the same signals to express a given meaning or for a same agent in two different contexts but that use in the both cases the same modality and the same signals to express a given meaning, the difference between the *behavioral expressivities* may induce at the level of dynamic instantiation coefficient, and consequently at the level of the animation a difference imperceivable for a human being.

6.3.2 Dynamic instantiation types

Expressivity ought to be modelled differently depending on the modalities (face, gesture, gaze and posture) it applies to. We consider several types of dynamic instantiation: *temporal* (e.g. mutual gaze duration, duration of a raised eyebrow), *spatial* (e.g. facial muscular contractions, width of the arms aperture) or *repetition*. For facial expression, variation of expressivity can be expressed through variation of muscular contractions as well as variation of its temporal course; while when talking about gaze, expressivity variations may be related to factors such as length of mutual gaze or length of looking at the conversation partners; while when talking about gesture,

```
 1. <agent
 2.    nameAgent       = "Agent1"
 3. >
 4.   <signal name = "look_at"
 5.     coeffDistort = "1"
 6.     distortion = "temporal">
 7.       <rheme>
 8.         <signal name = "smile"
 9.           coeffDistort = "1.2"
10.           distortion = "spatial"
11.         >
12.            I'm happy.
13.         </signal>
14.       </rheme>
15.   </signal>
16. </agent>
```

Figure 10: Example of output

it may be related to parameters such as the strength of a movement, its tempo, its dynamism or its spatial amplitude. Variation of expressivity may also be expressed by the rapid repetition of the same gesture (rapid head nods, fast beat gestures). In our example, the signal "*smile*" is associated to an "*amplitude*" dynamic instantiation, that is the system accentuates by *20%* the facial movements. Conversely, the system modulates the expressivity of the signal "*look_at*" by varying its duration (see figure 10).

6.4 Output

From the input text, the system applies several modifications until to obtain a text where tags $< signal >$ replace the communication act tags. Each of these modifications corresponds to a level of influences integration. The tags $< signal >$ are not associated to *expressivity* any more, but to the attribute *dynCoeff*. The figure 10 presents the output that the system processes.

At the computation level, these modifications of XML texts are applied according to XSLT stylesheets (see figure 11). XSLT (e**X**tensible **S**tylesheet **L**anguage **T**ransformation) is a language for transforming XML documents into other XML documents (see XSLT). The first transformation computes the *behavior expressivities* from the *communicative expressivities* in order to individualize the behavior according to the agent. The various libraries (*expression libraries* and *signal libraries*) are specified for a given input text, as well as the agents' definition. The second transformation creates, from the communication act tags, signals tags that are directly exploitable by the animation engine. The algorithms described in the previous sections are implemented in the stylesheets.

The output text of the figure 10 is simplified. A non simplified output contains temporal information related the signals. For example, for the facial signals, several other attributes are defined. They may specify the time

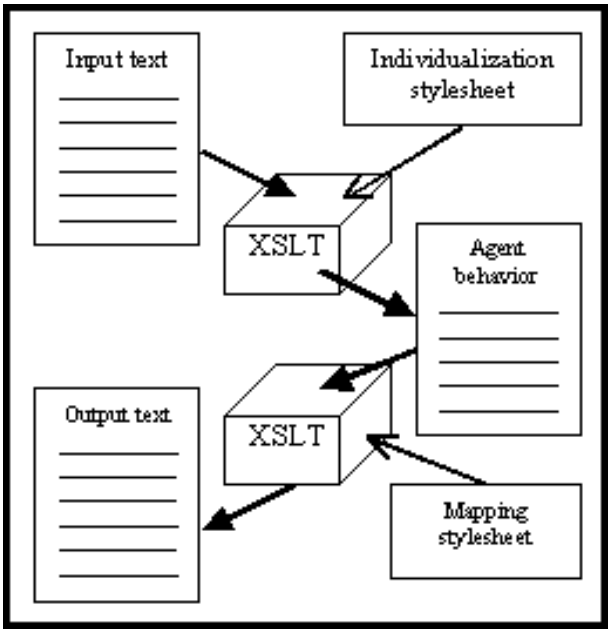

Figure 11: XSL Transformations

that the expression takes to reach its maximal intensity, the time during which the expression maintains its maximal intensity, the time that, starting from the maximal intensity, the expression changes into another expression, the time an expression waits until it raises, the time considered from the beginning of the tag, or the time an expression finishes to be shown before the end of the tag.

At the end, the system outputs a list of facial and body parameters that are used to drive the animation engine. To this end, we are using an MPEG-4 compliant animation engine, described in Pelachaud (2002).

7 Conclusion

We aim at creating an individual agent. Individuality is forged by several factors such as personality, social role, culture. Modelling such factors is extremely complex. To overcome this difficulty we propose to model influences by their impact they have on the behaviors expressivity. At first, we have described a taxonomy of influences as well as a set of parameters that characterize expressivity. The system is still being developed. We then foresee to do evaluation tests to validate the strategies for the intermodal expressivity. We also aim at testing the validity of the dynamic instantiation coefficient for each signal: we have to verify if adding expressivity does not create other meaning perceivable in the agent's behavior.

Acknowledgements

References

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.

R Brislin. *Understanding culture's influence on behavior*. Hartcout Brace College Publishers, New-York, 1993.

C.M. Brown and G. Nichols-English. Dealing with patient diversity in pharmacy practice. *Drug Topics*, pages 61–69, September 1999.

The Duy Bui, D. Heylen, M. Poel, and A. Nijholt. Generation of facial expressions from emotion using a fuzzy rule based system. In D. Corbett & M. Brooks M. Stumptner, editor, *Proceedings of 14th Australian Joint Conference on Artificial Intelligence (AI 2001)*, pages 83 – 94, Adelaide, Australia, 2003. Springer.

V. Carofiglio, F. de Rosis, and R. Grassano. Dynamic models of mixed emotion activation. In L Canamero and R.Aylett, editors, *Animating Expressive Characters for Social Interactions*. John Benjamins, Amsterdam, in press.

J. Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In S. Prevost J. Cassell, J. Sullivan and E. Churchill, editors, *Embodied Conversational Characters*. MITpress, Cambridge, MA, 2000.

B. DeCarolis, V. Carofiglio, M. Bilvi, and C. Pelachaud. APML, a mark-up language for believable behavior generation. In *Embodied conversational agents - let's specify and evaluate them!, Proceedings of the AAMAS'02 workshop*, Bologna, Italy, July 2002.

B. DeCarolis, C. Pelachaud, I. Poggi, and F. de Rosis. Behavior planning for a reflexive agent. In *IJCAI'01*, Seattle, USA, August 2001.

B. DeCarolis, C. Pelachaud, I. Poggi, and M. Steedman. APML, a mark-up language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, *Life-like Characters. Tools, Affective Functions and Applications*, pages 65–85. Springer, 2004.

M.S. El-Nasr, J. Yen, and T. Loerger. FLAME - fuzzy logic adaptive model of emotions. *International Journal of Autonomous Agents and Multi-Agent Systems*, 3 (3):1–39, 2003.

H. McBreen H. Morton and M. Jack. Experimental evaluation of the use of ECAs in eCommerce applications. In Z. Ruttkay and C. Pelachaud, editors, *From Brows till Trust: Evaluating Embodied Conversational Agents*. Kluwer, To appear.

K. Höök. User-centred design and evaluation of affective interfaces. In Z. Ruttkay and C. Pelachaud, editors, *From Brows till Trust: Evaluating Embodied Conversational Agents*. Kluwer, To appear.

K. Isbister and C. Nass. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, Forthcoming.

Scott A. King, Alistair Knott, and Brendan McCane. Language-driven nonverbal communication in a bilingual conversational agent. In *Proceedings of CASA 2003*, pages 17 – 22, 2003.

S. Kshirsagar, C. Joslin, W. Lee, and N. Magnant-Thalmann. Personalized face and speech communication over the internet. In *Proc. of IEEE Virtual Reality*, pages 37–44, Tokyo, Japan, 2001. IEEE Computer Society.

E.-J. Lee and C. Nass. Does the ethnicity of a computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. In *WECC'98, The First Workshop on Embodied Conversational Characters*, October 1998.

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.

Stacy Marsella and Jonathan Gratch. Modeling coping behavior in virtual humans: Don't worry, be happy. In *proceedings of the /2nd International Conference on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, 2003.

A. Paiva, G. Andersson, K. Höök, D. Mourao, M. Costa, and C. Martinho. SenToy in FantasyA: Designing an affective sympathetic interface to a computer game. *Journal of Personal and Ubiquitous Computing*, 6(5-6):378–389, 2003.

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, 2002.

C. Pelachaud and I. Poggi. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation*, 13:301–312, 2002.

P. Piwek, B. Krenn, M. Schröder, M. Grice, S. Baumann, and H. Pirker. RRL: a rich representation language for the description of agents behaviour in NECA. In *Embodied conversational agents - let's specify and evaluate them!, Proceedings of the AAMAS'02 workshop*, Bologna, Italy, July 2002.

I. Poggi. Mind markers. In N. Trigo M. Rector, I. Poggi, editor, *Gestures. Meaning and use.* University Fernando Pessoa Press, Oporto, Portugal, 2003.

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

H. Prendinger, S. Descamps, and M. Ishizuka. Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence*, 16(7-8):519–553, 2002.

T. Rist and M. Schmitt. Applying socio-psychological concepts of cognitive consistency to negotiation dialog scenarios with embodied conversational characters. In *Proc. of AISB'02 Symposium on Animated Expressive Characters for Social Interactions*, pages 79–84, 2003.

Zs. Ruttkay, V. van Moppes, and H. Noot. The jovial, the reserved and the robot. In *proceedings of the AAMAS03 Ws on Embodied Conversational Characters as Individuals*, Melbourne, Australia, 2003.

VHML. **Virtual Human Markup Language**. http://www.vhml.org.

XSLT. **eXtensible Stylesheet Language Transformation**. http://www.w3.org/TR/xslt.