# GRETA : Towards an Interactive Conversational Virtual Companion

**Elisabetta Bevacqua, Ken Prepin, Radoslaw Niewiadomski, Etienne de Sevin, Catherine Pelachaud**
**CNRS – LTCI, Telecom-ParisTech**

## 1 INTRODUCTION

Researches have shown that people tend to interact with computers characterized by human-like attributes as if they were really humans (Nass et al., 1997, Reeves and Nass, 1996). For example, in their studies Nass and Reeves saw that, while interacting with computers, people apply rules of politeness and felt uneasy when large faces were displayed on a screen, like the talking head was invading their personal space (Reeves and Nass, 1996). Consequently, humane-machine interface designers should aim to implement interactive systems that simulate human-like interaction. The more this type of interfaces is consistent with human style of communication, the more their use will become easy and accessible (Ball and Breese, 2000). Such level of consistency could be reached using humanoid artefacts able to apply that rich style of communication that characterizes human conversation. The recent technological progress has made the creation of this type of humanoid interfaces, called Embodied Conversational Agents (ECAs), possible. An ECA is a computer-generated animated character that is able to carry on natural, human-like communication with users (Cassell et al., 2000b). For this purpose, all the researches engaged in ECAs development share common goals: on the one hand, they want to implement agents who can simulate humans' verbal and non-verbal behaviour, like speaking in natural language, performing gestures, displaying facial expressions, shifting their gaze and moving their head like humans do in everyday life. On the other hand, researchers aim to provide these virtual artefacts with the capability of understanding what humans say, interpreting their non-verbal signals and using all this information to decide how to react and respond.
Several ECAs have been developed so far. They exhibit human-like communicative capabilities: they can talk, listen, grab one's attention, look at another one, show emotion, and so on (Cassell et al., 2000a; Gustafson et al., 1999; Gratch and Marsella, 2004; Kopp and Wachsmuth, 2004; Pelachaud, 2005; Heylen, 2006; Gratch et al., 2007). They can play different roles, from a companion for old people or young kids, to a virtual trainee, a game character, a pedagogical agent or even a web agent (Johnson et al., 2005; Moreno, in press; Cassell et al., 1999; Hall et al., 2006; Bickmore et al, 2007).

In this chapter we present our work toward building a conversational companion. Conversing with partner(s) means to be able to express one's mental and emotional state, be a speaker or a listener. One needs also to adapt to our partner's reactions to what one is saying. We have developed an interactive ECA platform, Greta (Pelachaud, 2005). It is a 3D virtual agent capable of communicating expressive verbal and nonverbal behaviours as well as listening. It can use its gaze, facial expressions and gesture to convey a meaning, an attitude or an emotion. Multimodal

behaviours are tightly tied with each other. A synchronization scheme has been elaborated allowing the agent to display a raise eyebrow or a beat gesture on a given word. According to its emotional or mental state, the agent may vary the quality of its behaviours: it may use more or less extended gesture, the arms can move at different speeds and with different acceleration (Mancini & Pelachaud, 2008). The agent can also display listener behaviours (Bevacqua et al, 2008). It interacts actively with users and/or other agents providing appropriate timed backchannels. Interaction means also the interactants ought to adapt each other behaviours; dynamic coupling between them needs to be considered (Prepin & Revel, 2007).

In the remaining of this chapter we describe the implementation of our ECA system. We describe in more details the modules linked to listener model.
In the next section we introduce the SAIBA framework that Greta complies to. Representation languages used to control the agent are presented. Then each module of the ECA system is explained. The chapter ends with the description of several applications of our system.

# 2 GRETA

GRETA's architecture follows the design methodology proposed in (Thórisson et al., 2005) and is compatible with the SAIBA framework (Vilhjálmsson et al., 2007) (see next subsection 3.1). Its architecture is modular and distributed. Each module exchanges information and data through a central message system by the means of whiteboards as defined by Thorisson (Thórisson et al., 2005). It allows internal modules and external software to be integrated easily. The system is designed to be used in interactive applications working in real-time. Interactive applications of our system were developed within the eNTERFACE[1], SEMAINE[2] and CALLAS[3] EU-projects (see section Interactive Applications).

## 2.1 The SAIBA framework

SAIBA[4] is an international research initiative whose main aim is to define a standard framework (i.e. a conceptual architecture and associated standard languages, see section Standard Languages) for the generation of virtual agent behaviour (Vilhjálmsson et al., 2007) (see Figure 1).

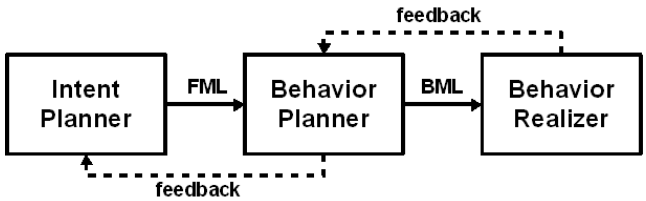


Figure 1: SAIBA architecture.

[1] Summer Workshop, eNTERFACE project, France 2008; http://enterface08.limsi.fr/
[2] FP7 STREP SEMAINE project IST-211486, http://www.semaine-project.eu
[3] FP6 IP CALLAS project IST-034800, http://www.callas-newmedia.eu
[4] http://wiki.mindmakers.org/projects:SAIBA:main

It defines a number of levels of abstraction, from the computation of the agent's communicative intention, to behaviour planning and realisation. The *Intent Planner* module decides the agent's current goals, emotional state and beliefs, and encodes them into the Function Markup Language (FML) (Heylen et al., 2008), a standard language still being defined (see section Languages for ECAs). To convey the agent's communicative intentions, the *Behavior Planner* module schedules a number of communicative signals (e.g., speech, facial expressions, gestures) which are encoded with the Behaviour Markup Language (BML). The BML specifies the verbal and nonverbal behaviours of ECAs (Vilhjálmsson et al., 2007) (see section languages for ECAs being defined). Finally the task of the third element of the SAIBA framework, *Behavior Realizer*, is to realize the behaviours scheduled by the *Behavior Planner*. It receives input in the BML format and it generates the animation.

## 2.2 SAIBA examples

There exists several implementations like SmartBody (Thiébaux et al., 2008) and BML Realizer (Árnason and Porsteinsson) that are SAIBA compatible.
SmartBody (Thiébaux et al., 2008) is a modular, distributed open-source framework for animating ECAs in real time. It corresponds to the *Behavior Realizer* module of the SAIBA architecture. It takes as input BML code (including speech timing data and the world status updates); it composes multiple behaviours and generates character animation synchronized with audio. The verbal content is generated by an external TTS system. BML used within SmartBody is a subset of the standard (Thiébaux et al., 2008); but it offers some extensions as well. SmartBody can be used with the Nonverbal Behaviour Generator (Lee and Marsella, 2006) that corresponds to the *Behavior Planner* in the SAIBA framework. It is a rule-based module that generates BML annotations for nonverbal behaviours from the communicative intent and speech text. On the other hand, SmartBody can be used with different characters, skeletons and even different rendering engines.

BMLRealizer (Árnason and Porsteinsson) created in the CADIA lab is another implementation of the *Behavior Realizer* layer of the SAIBA framework. It is an open source animation toolkit for visualizing virtual characters in 3D environment that is partially based on the SmartBody framework. As input it also uses BML; the output is generated with the use of the Panda3D rendering engine.

## 2.3 Greta's implementation of SAIBA

Greta's architecture is an almost full implementation of the SAIBA framework. It is composed of three main modules (see Figure 2), offering solution for the *Behavior Planner* and the *Behavior Realizer* and a partial implementation of the *Intent Planner.*
In the SAIBA standard the *Intent Planner* is dedicated to generate intentions of a Speaker virtual agent. To be able to control a Listener agent, we have introduced the *Listener Intent Planner*, which generates automatically the communicative intentions of the listener. In the current state of our system, when the agent is the Speaker (and not the Listener) its intentions are pre-defined manually in an FML-APML input file

(FML-APML is detailed in the Section 3.3.1). In future works, they should be generated by a *Speaker Intent Planner*.

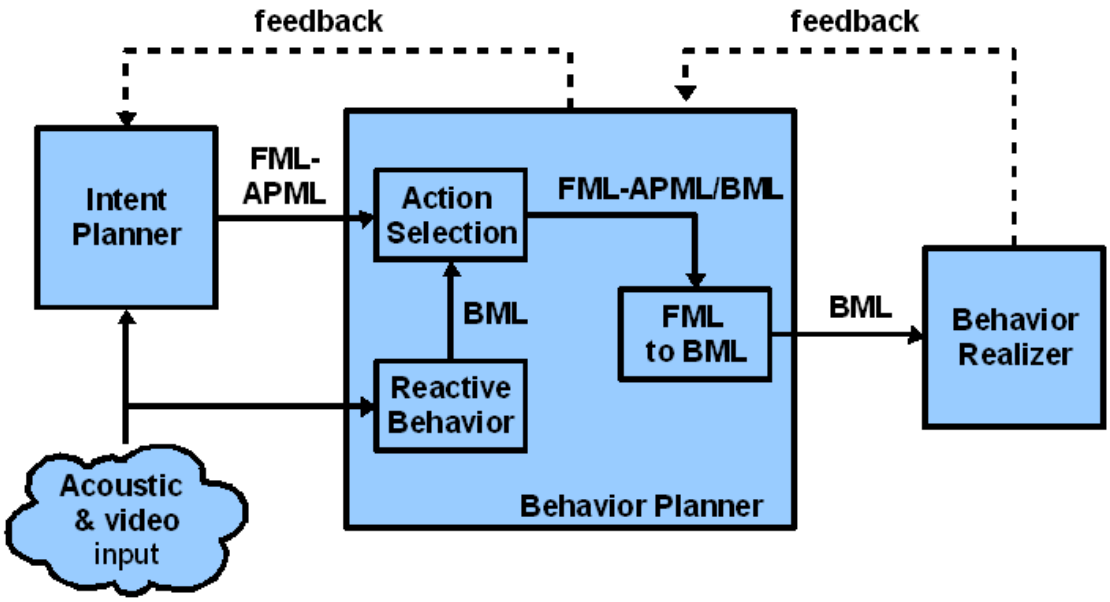



Figure 2: GRETA's architecture.

Each communicative intention generated by the *Intent Planner* (both when the agent is Speaker or Listener) is transmitted to the *Behavior Planner*, in FML-APML language. The *Behavior Planner* proposes a list of corresponding possible nonverbal behaviours, written in BML language (see section Languages).
These behaviour signals are sent to the *Behavior Realizer* that generates MPEG-4 Facial Animation Parameter (FAP) and Body Animation Parameters (BAP) frames. Finally, the animation is played in the *FAP-BAP Player*.
All modules are synchronized by the Central Clock and communicate with each other through the Psyclone whiteboard (Thórisson et al., 2005).

### 3.3.1 The Agent Languages

Two representation languages are used in the Greta's architecture to formalise the exchange of information between modules: the Function Markup Language (FML-APML) and the Behaviour Markup Language (BML). The FML, is still at a very early age of specification. We propose a temporary solution that we called FML-APML. It encompasses the language APML (DeCarolis et al., 2001) we have been using as well as other functionalities.
The second language, BML, is derived from the SAIBA framework and so far this language has become an almost standardised language, commonly used by ECAs community. Greta implements the current version of BML with several extensions that allows one to exploit better the capabilities of our agent.

#### *FML-APML*

FML encodes communicative and emotional functions the agent aims to transmit. Our version of this language, FML-APML, is an XML-based markup language for representing the agent's communicative intention and the text to be uttered by the agent. The communicative intentions of the agent correspond to what the agent aims to communicate to the user: e.g., its emotional states, beliefs and goals. FML-APML

uses a similar syntax as BML one. It has a flat structure and allows defining explicit duration for each communicative intention. Each tag represents one communicative intention; different communicative intentions can overlap in time. We consider the following tags (see (Mancini & Pelachaud, 2008) for more details):

- certainty: is used to specify the degree of certainty the agent intends to express;
- performative: represents the agent's performative e.g. suggest, approve, or disagree;
- theme/rheme: represents the topic/comment of conversation; that is, respectively, the part of the discourse which is already known or new in the participants' conversation;
- belief-relation: corresponds to the metadiscoursive goal, i.e. the goal of stating the relationship between different parts of the discourse;
- turntaking: models the exchange of speaker turns;
- emotion: describes the emotional state of the agent;
- emphasis: is used to emphasize the agent's verbal or nonverbal message;
- backchannel: describes the listener's communicative intentions, i.e. its will and ability to continue, perceive, understand the interaction and its attitude towards the speaker's speech (if it believes or not, likes or not, accepts or refuses what is being said) (Allwood et al., 1992)
- world: refers to objects of the world.

We remark that this language allows us to describe the agent's communicative functions when it is either speaker or listener.

### *BML: Behavior Markup Language*

BML language is not yet a standard, however researchers agreed on a "common" BML syntax specification to allow one to exchange BML files and engines between different systems, as described in (Vilhjálmsson et al., 2007). The BML language allows us to specify the nonverbal signals that can be expressed through the agent communication modalities. Each BML top-level tag corresponds to a behaviour the agent is to produce on a given modality: head, torso, face, gaze, body, legs, gesture, speech, lips. In the current version for each modality one signal can be chosen from a short fixed list. Each signal has a start time and duration defined. This temporal information can be absolute (in seconds) or relative, in relation to the other verbal or nonverbal signals.

The BML language version we have implemented in our agent contains some extensions which allow us to define labels to use a larger set of signals which can be produced by the agent and to specify the expressivity of each signal (Mancini & Pelachaud, 2008).

***Signal label.*** In the common BML syntax it is possible to specify just a small set of signals for the agent. For example one can specify only 4 mouth shapes: flat, smile, laugh and pucker. In our version of BML we have two types of information about a signal: the *type* attribute, which is mandatory and refers to the small set of signals defined in the BML common version, and the tag *reference* which is used by our agent to perform a nonverbal behaviour from a larger set of signals our agent can perform.

***Expressivity parameters.*** Our agent can dynamically modulate multimodal signals using a small set of high level parameters that we call *expressivity parameters*

(Hartmann et al., 2006). They influence the quality of movement: for example, the gesture of raising a hand can be performed quickly or slowly and so on. Expressivity parameters are not included in the common BML syntax but can be interpreted by our *Behavior Realizer*. Thus, in the implementation of BML in our system, we can specify not only *which* signals the agent has to perform but also *how* it will execute them.

### 3.3.2 Intent Planner

The *Intent Planner* module has the task of computing the agent's intentions when being a listener or a speaker. This module consists of two modules: the *Listener Intent Planner* and the *Speaker Intent Planner* that calculate the agent's behaviour respectively while it is listening and while it is speaking. So far just the module to compute the listener's intentions has been implemented and the Speaker Intent Planner is still under construction. Together, these two modules will correspond to the Intent Planner in the SAIBA framework.

#### Listener Intent Planner

The *Listener Intent Planner* computes the backchannel signals that the agent provides while listening. This module implements three types of backchannels: reactive, response as well as mimicry. Reactive backchannels derive from a first process of perception of the speaker's speech and they show contact and perception; response backchannels are generated by a more aware evaluation that comprehends memory and cognitive process (Kopp et al., 2008). Finally, the mimicry backchannels derive from the imitation of the speaker's behaviour. For mimicry we mean the behaviour displayed by an individual who does what another person does (Van baaren, 2003). This type of behaviour has been proven to play quite an important and positive role during conversations (Warner et al., 1987; Chartrand and Bargh, 1999).
Several researches (Allwood et al., 1992; Poggi, 2005) have noticed that listeners can emit signals with different levels of intentionality: they react instinctively to the speakers behaviour, generating signals at a very low level of control; or they can decide to consciously emit a signal in order to show their reaction to the speakers speech and even act with the intent to influence the speaker's behaviour. The Listener Intent Planner generates listener's signals with a low level of awareness. We do not take into account the agent's intention to consciously provide a backchannel signal; backchannels are emitted unintentionally.
To generate backchannel signals, the Listener Intent Planner uses two modules called *response/reactive backchannel* and *mimicry*. In order to decide *when* a backchannel should be emitted and to select *which* communicative functions the agent should transmit, the Listener Intent Planner component needs three data as input:

- the user's verbal and nonverbal behaviour that is tracked through a video camera and a microphone;
- the user's estimated interest level. Such a level is calculated evaluating the user's gaze, head and torso direction within a temporal window (Peters et al., 2005);
- the agent's mental state towards the interaction.

Research has shown that there is a strong correlation between backchannel signals and the verbal and nonverbal behaviours performed by the speaker (Maatman et al., 2005; Ward and Tsukahara, 2000). Models have been elaborated that predict when a backchannel signal can be triggered based on a statistical analysis of the speaker's behaviours (Maatman et al., 2005; Morency et al., 2008; Ward and Tsukahara, 2000).

From the literature (Maatman et al., 2005; Ward and Tsukahara, 2000) we have fixed some probabilistic rules to prompt a backchannel. Our system analyses speaker's behaviors looking for those that could prompt an agent's signal; for example, a head nod or a variation in the pitch of the user's voice will trigger a backchannel with a certain probability. When a rule is found, the probability that certain behaviour provokes a backchannel depends also on the user's estimated level of interest. This value is used by the system to vary the backchannel emission frequency: when the interest level decreases it may be a sign that the user might want to stop the conversation (Schegloff and Sacks, 1973), consequently the agent shows disengagement as well and provides less and less backchannels.
When a backchannel is triggered, the response/reactive backchannel, and mimicry modules compute which type of backchannel should be displayed. The response/reactive backchannel module uses information about the agent's beliefs towards the speaker's speech to calculate the backchannel signal. This information is stored in the agent's mental state that describes how the agent is reacting to the user's speech. In our system the agent's mental state is represented as a list of the listener's communicative functions. We use Allwood's and Poggi's taxonomies of communicative functions of backchannels (Allwood et al., 1992; Poggi, 2005): understanding and attitudinal reactions (liking, accepting, agreeing, believing, being interested). The response/reactive module takes into account the agent's mental state to decide which communicative functions the agent should convey. Then, the appropriate signals to display will be selected from a backchannel lexicon that we have elaborated in previous studies (Bevacqua et al., 2007; Heylen et al., 2007). When no information about the agent's beliefs towards the speaker's speech is given, the response/reactive module selects a pre-defined backchannel among those signals that have been proven to show contact and perception, like head nod and raise eyebrows.
When fully engaged in an interaction, mimicry of behaviours between interactants may happen (Lakin et al., 2003). The mimicry module determines which signals would mimic the agent. So far we are considering solely speaker's head movement in the signals to mimic. A selection algorithm (explained below) determines which backchannel to display among all the potential signals that are outputted by the two modules.

### 2.3.3 Behavior Planner

The *Behavior Planner* takes as input both the agent's communicative intentions specified by the FML-APML language and some agent's characteristics (i.e. *baseline*). The main task of this component is to select, for each communicative intention, the adequate set of behaviours to display. The output of the *Behavior Planner* is described in the BML language. It contains the sequence of behaviours with their timing information to be displayed by our virtual agent.

#### *Reactive Behavior*

The mutual adaptation necessary to enable verbal interaction between an ECA and a human is, in some way, highly cognitive: the speaker can have to re-plan its speech depending on listener’s reactions; the emotions of the agents (speaker or listener) can change throughout the dialogue and influence its behaviour.
However this mutual adaptation is also, in some other way, mostly reactive: it is not linked to the meaning of words or to changes within the mental state of the agent, but

more to psychophysical and dynamical properties of the exchanged signals (Murray and Trevarthen, 1985, Adamson and Frick, 2003, Nadel et al., 2005, Striano et al., 2005, Soussignan et al., 2006, Prepin and Revel, 2007, Auvray et al., 2009). Certain facial expressions such as joy or fear can have direct impact on the partner and induce immediate reaction such as respectively smile or fear, certain breaks or modulation within the speech appeal immediate answer. To enable partners of an interaction to imitate, to synchronise with each other or even to slow down or speed up together their rhythms of production, social agent ought to be endowed with the capability to produce such reactive behaviours.

This dynamical aspect of the interaction is much closer to the low-level of the agent system than to the high-level of the communicative intentions described by FML: to implement this dynamical part of the social agent, the ECA needs reactivity (realtime perception) and sensitivity (realtime adapted actions). An architecture enabling dynamical coupling (Prepin and Revel, 2007) has been adapted to Greta independently from SAIBA framework. This architecture implements two capacities giving to the agent dynamical properties and reactivity: the self-generation of dynamics (driven by an oscillator) and the sensibility to partner's behaviour (detected by movement analysis). When two agents interact with each other, coupling behaviours such as synchronisation and turn-taking emerge.

We are presently working on integrating this architecture to the *Behavior Planner*. The *Reactive Behavior* module will have certain autonomy from the rest of the architecture. It will short-cut the *Intent Planner*, getting input signals directly, i.e. the BML stream coming from the analysis of human's behaviour (see Figure 1), as well as the currently planned actions, i.e. the BML outputted by the *Behavior Planner*.

With these two sources of information, the *Reactive Behavior* module will propose to the Action Selection module (see next paragraph, Action Selection) two different types of data. On one hand, it can propose adaptation of the current behaviour by comparing its own actions to the actions of the speaker at a very low level. Tempo and rhythm of the partner's signal production are computed as well as the current synchrony between partners. Then corresponding adaptation of the current behaviour will be proposed; for example it can propose to slow down or speed up behaviours. This type of propositions may enable synchronisation or similarity of tempo with the user. On the other hand, the second type of data proposed by the *Reactive Behavior* is: By extracting from the user's behaviour salient events such as facial expressions, speech prosody or breaks, it will propose actions such as performing a backchannel, imitating the user, smiling etc.

Finally the *Reactive Behavior* will be able to propose real-time reactions or adaptations to the user's behaviour thanks to its partial autonomy. It will act more as an adaptator of the ongoing interaction than as a planner. It is a complementary part of the *Intent Planner*, much more reactive and also working at a much lower level. The ECA must be able to select or to merge the information coming from both this *Reactive Behavior* and the *Intent Planner*, using an Action Selection module.

### *Action Selection*

Tyrrell (Tyrrell, 1992) defines the task of action selection mechanism for an agent as "determining, from a set of available conflicting actions, the most appropriate ones". The goal of the Action Selection consists in adapting the actions according to the user's interest level (as it is perceived by the agent) and selecting which action is displayed among the possible (conflicting) ones.

Inspired from the free flow hierarchy approach (de Sevin and Thalmann, 2005), no choice is made before the Action Selection module. It receives all propositions of actions (see Figure 1) from the *Intention Planner* module such as response backchannels specified with FML (see section 3.2.2) and adaptation to behaviour tempo (written with BML) from the *Reactive Behavior* module. Thus, it has to choose between a more cognitive-driven and a more reactive-driven behaviour. To enable the Action Selection module to make a choice, actions are associated to priorities generated in the production modules. The first step is to compute the probability of proposed actions and normalize their priorities according to the user's interest level. The latter is considered as a good indicator of the successfulness of the interaction (Peters et al., 2005).
Based on this probability uniformity, the action selection module is now able to compare action priorities for actions that are conflicting at the signal level. For example, the ECA cannot generate a head shake to mimic the user and a head nod determined by the communicative function "agree" (Poggi, 2005). Only one of these two actions can be displayed at the same time and a selection is necessary. The selection is event-based; the algorithm is real-time. Finally, the action selection module chooses the most appropriate actions based on the priority values according to the user's interest level. The selected action is sent to the FML-to-BML module to be displayed by the ECA.

#### *FML-to-BML*

This module receives as input sequences of communicative acts specified by FML-APML tags. It interprets these tags and decides which signal to convey on which modality for each communicative act.

Behaviour is defined by a given facial expression or a particular hand configuration and arm position. But another element characterizes behaviour: the manner of execution of the behaviour; we call this parameter the expressivity of the behaviour. Until now the behaviour has been described statically: a facial expression is defined at its apex (Ekman, 1979) and the shape of a gesture is specified by the shapes it has over the various phases that composed it (e.g., preparation phase, stroke) (McNeill, 1992). The expressivity parameter refers to the dynamic variation of the behaviour along this static description, for e.g., the temporal duration and strength of the behaviour (Hartmann et al, 2006).

Since not every human exhibit similar behaviours quality, we have introduced the notion of baseline (Mancini and Pelachaud, 2008). An agent is described by a specific baseline; it tells the general tendency an agent has to use such and such modalities with such and such expressivity. Thus an agent doing large and fast gesture in general will be defined with a different baseline than an agent using rarely arm movement and facial expression. The baseline for each agent affects how the agent communicates a given intention or emotion. These models allow us to derive an agent able to display expressive nonverbal behaviours.

### 3.3.4 Behavior Realizer

The Behavior Realizer module generates the animation of our agent following the MPEG-4 format (Ostermann, 2002). The input of the module is specified by the BML language. It contains the text to be spoken and/or a set of nonverbal signals to be displayed. Facial expressions, gaze, gestures, torso movements are described symbolically in repository files. Each BML tag is instantiated as a set of key-frames that are then smoothly interpolated. The *Behavior Realizer* synchronizes the behaviours across modalities. It solves also eventual conflicts between the signals that use the same modality. The speech is generated by an external TTS[5] and the lips movements are added to the animation. We are currently using Mary (ref), Festival (ref) and Euler (ref).
When the *Behavior Realizer* receives no input, the agent does not remain still. It generates some idle movements. Periodically a piece of animation is computed and is sent to the FAP-BAP Player. This avoids unnatural "freezing" of the agent.

### 3.3.5 FAP-BAP Player

The FAP-BAP Player receives the animation generated by the *Behavior Realizer* and plays it in a graphic window. The player is MPEG-4 compliant. Facial and body configurations are described through respectively FAP and BAP frames.

### 3.3.6 Synchronization

The synchronization of all modules in the distributed environment is ensured by the Central Clock which broadcasts regularly timestamps through the whiteboard. All other components are registered in the whiteboard to receive timestamps.

# 4 Interactive applications

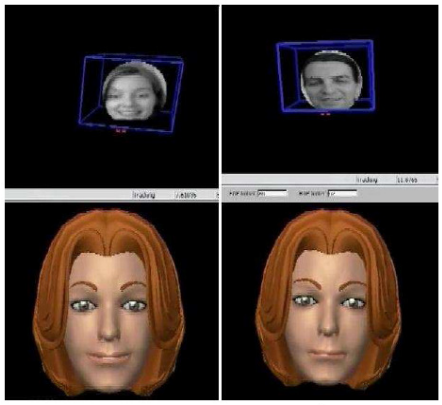
Our agent is designed to be able to manage a natural interaction with users and virtual agents. Our first step along this line has been to develop some applications that allow users to interact with Greta. We want to see if our agent can sustain an interaction with a user and how users judged the experience. Firstly, we have made the agent perceive the external world. Analysis components detecting the user's voice characteristics (like pauses and pitch variations) and nonverbal behaviours (such as head movements and facial signals like smile) can be connected to our system (Baklouti et al., 2008; Morency et al., 2005; Pure Data). The information, sent by these components through the whiteboard, is then used to plan the agent's response. The agent is able to process this data and to react to it nearly instantly. It allows the agent to be interactive with the user.

## *4.1 An interactive listening agent*

Our ECA system has been used to build an interactive listening agent (Bevacqua et al., 2008). In the SEMAINE project[6], we are developing a Sensitive Artificial

[5] MARY text-to-speech system developed by March Schröder, DFKI. http://mary.dfki.de/

Listener, SAL, (Douglas-Cowie et al., 2008). Our aim is to endow the agent with the capability to sustain emotionally coloured interaction, in particular by showing appropriate backchannels.



a b

Figure 3: A real-time listening agent: interaction between a user and the agent in the role of: a) Poppy and b) Spike.

In this project we are considering 4 SAL agents each defined with a specific baseline (Bevacqua et al., 2008). Depending on its baseline and its mental state, a SAL agent shows very different backchannels. Figure 3 shows a single frame of an interaction between a user and 2 SAL agents. The user was asked to tell a story to the agent that shows its participation through the production of backchannel signals.
In Figure 3a, the agent is Poppy an outgoing agent. The agent is showing positive backchannels using smile and head nod. On the other hand, Figure 3b illustrates Spike interacting with a user. Spike is very argumentative and uses mainly frown to signal negative backchannels such as disagreement. In this application, we have interfaced our system with Watson (Morency et al., 2005), a real-time head tracking system.

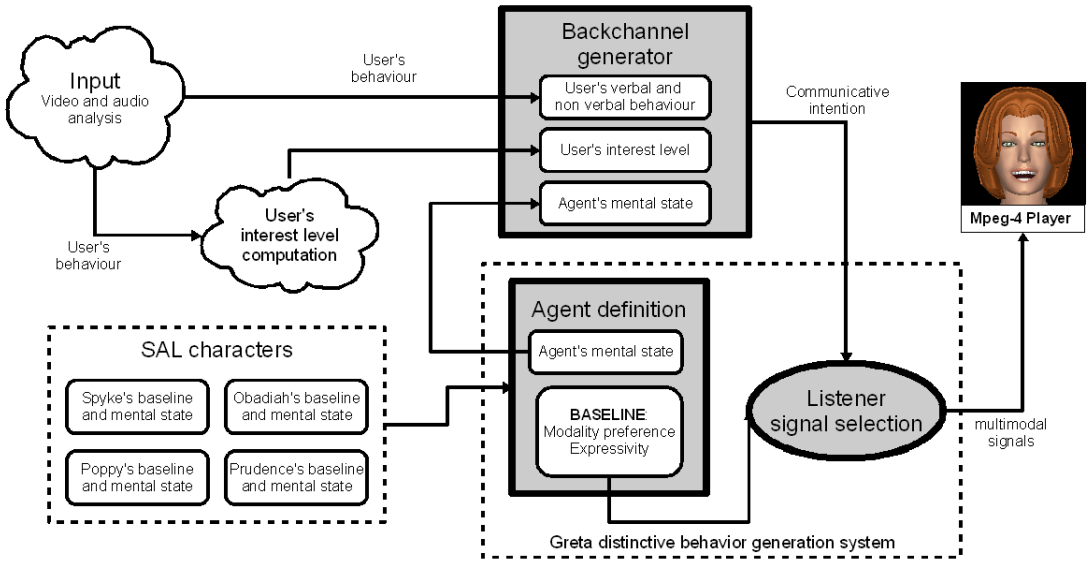



Figure4: Architecture of SAL system

## *4.2 Interactive listening virtual and robotic agents.*

Within the eNTERFACE summer workshop (Moubayed et al., 2008) which was devoted to research on multimodal interfaces, the *Listener Intent Planner* component has been connected with Pure Data (Pure Data), a graphical programming environment for real-time audio processing, and with a face detection module that can detect user's smiles, head nods and shakes (Baklouti et al., 2008). The backchannel signals generated by this system are displayed both by our 3D agent and by a SONY Aibo robot (see Figure 4). Since Aibo is a dog-like robot, an ad hoc backchannel lexicon had to be elaborated manually. For example, when the backchannel signal requested is a smile the dog wags its tail and bright lights on its head and on its back are turned on. The control of the behaviours of the agent and of the robot is done via the BML language. The *Behavior Planner* outputs FAPs for the agent and Aibo commands for the robot.

Figure 4: A user interacts with Greta and AIBO.

## 4.3 Applications in CALLAS

Another application of our agent that is being developed within the CALLAS project[7] is called The Interactive Storyteller. In this scenario a computer application presents a story content and displays a sequence of still images (with high emotional impact) related to the story. A web site at BBC is dedicated to public presentation of news items. Photos accompanied by explicative captions relate emotional content. A particular story reports images on the earthquake of Schezuan.
In the Interactive Storyteller application interaction takes the form of a guided conversation between the virtual agent and the user. The ECA is a storyteller. Its role is to interact with the user. It first asks her to comment each displayed image; it then provides some explanation. To enhance the emotional experience of the user, the agent shows affective empathy toward her showing its participation toward the story. The user expresses her opinion about the images. In the background its speech and gestures are analysed by system to detect her emotional states. For this purpose the MKS Keyword Spotting[8] and Gesture Expressivity Recognition[9] software provided by CALLAS partners are used. Next, the agent explains what the image is about. It uses various nonverbal signals like emotional facial expressions and gestures to emphasize the message and guide the emotional reactions of the user. Detected information about user's affective state is used by the system to influence the agent's

[7] Work done in collaboration with Marc Price, BBC.
[8] Jérôme Urbain and Thierry Dutoit, Faculté Polytechnique de Mons, Belgium
[9] Stelios Asteriadis and Amaryllis Raouzaiou, Image, Video and Multimedia Systems Laboratory, National Technical University of Athens, Greece

affective behaviour. When the ECA begins conveying the scene $s_{i+1}$ it expresses the same affective state as measured from the user in the previous segment of the story $s_i$, and then, throughout the duration of the scene, it gradually changes its expressed emotional state to the $s_{i+1}$ target affective state.

# 5 CONCLUSION

We have presented an interactive system of a virtual conversational companion. It models agent while being a speaker and a listener in an interaction.
In both roles, the agent's behaviours are derived from its own communicative intentions and emotional state. The agent is endowed with expressive capabilities; it can communicate its mental and emotional using various modalities (facial expression, gaze, gesture, and head). Behaviours can be of different varieties to simulate the agent's general behaviour tendency. Being able to display appropriate backchannels allows the agent to show its engagement in the interaction (or the way around, its disengagement). Reactive responses are triggered by external and internal events. These actions arise from the dynamism between both interactants. They allow synchrony and even mimicry to emerge and to be a sign of strong engagement between the conversational companions. Thus Expressivity, responsiveness and reactivity are importance features to embed in a virtual companion.

**References**


Adamson, L. B., Frick, J. E., The still face: A history of a shared experimental paradigm. *Infancy*, 4(4):451–473. 2003.

Allwood, J., Nivre, J. and Ahlsén, E. On the Semantics and Pragmatics of Linguistic Feedback. Journal of Semantics 9, 1-26. 1992.

Árnason, B.P., Porsteinsson, A., The CADIA BML realizer. http://cadia.ru.is/projects/bmlr/.

Auvray, M., Lenay, C., Stewart, J., Perceptual interactions in a minimalist virtual environment. *New ideas in psychology*, 27:32–47. 2009.

Baklouti, M., Couvet, S., Monacelli, E., Intelligent Camera Interface (ICI): A Challenging HMI for Disabled People. In Advances in Computer-Human Interaction, 2008 First International Conference on, pages 21—25, 2008.

Ball, G., Breese, J., Emotion and personality in a conversational agent. In Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors, Embodied Conversational Characters, Cambridge. MIT Press. 2000.

Bevacqua, E., Heylen, D., Tellier, M., Pelachaud, C., Facial feedback signals for ECAs. In AISB'07 Annual convention, workshop "Mindful Environments", pages 147--153, Newcastle upon Tyne, UK. 2007.

Bevacqua, E., Mancini, M., Pelachaud, C., A listening agent exhibiting variable behaviour. In Prendinger, H., Lester, J. C., and Ishizuka, M., editors, Proceedings of 8th International Conference on Intelligent Virtual Agents, volume 5208 of Lecture Notes in Computer Science, pages 262--269, Tokyo, Japan. Springer, 2008.

Bickmore, T., Mauer, D., Brown, T., Context Awareness in Mobile Relational Agents. 7th International Conference on Intelligent Virtual Agents, Paris, pp354-355, 2007.

Cassell, J., Sullivan, J., Prevost, P., Churchill, E., Embodied Conversational Characters. MIT Press, Cambridge, MA. 2000a.

Cassell, J., Bickmcre, T., Campbell, L., Designing Embodied Conversational Agents. Embodied Conversational Agents. 2000b.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., Yan, H., Embodiment in Conversational Interfaces: Rea. *Proceedings of the CHI'99 Conference,* pp. 520-527. Pittsburgh, PA. 1999.

Chartrand, T. L., Bargh., J. A., The Chameleon Effect: The Perception-Behavior Link and Social Interaction. Personality and Social Psychology, 76:893–910, 1999.

DeCarolis, B., de Rosis, F., Pelachaud, C., Poggi, I., A reflexive, not impulsive Agent, The Fifth International Conference on Autonomous Agents, Montreal, Canada, pp. 186-187, May 2001.

Douglas-Cowie, E., Cowie, R., Cox, C., Amir, N., Heylen, D., The sensitive artificial listener: an induction technique for generating emotionally coloured conversation. In LREC2008 - Workshop on Corpora for Research on Emotion and Affect, Marocco, May 2008.

Gustafson, J., Lindberg, N., Lundeberg, M., The August spoken dialog system. Proceedings of Eurospeech'99, Budapest, Hungary, 1999.

Gratch, J., Marsella, S., A domain-independent Framework for modeling emotion. Journal of Cognitive Systems Research, 5(4), 269-306, 2004.

Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy R. Creating Rapport with Virtual Agents. 7th International Conference on Intelligent Virtual Agents, Paris, 125-138, 2007.

Hall, L., Vala, M., Hall, M., Webster, M., Woods, S., Gordon, A., Aylett, R., FearNot's appearance: Reflecting Children's Expectations and Perspectives. In J Gratch, M Young, R Aylett, D Ballin and P Olivier, eds. 6th International Conference, IVA 2006, Springer, LNAI 4133, 407-419, 2006.

Hartmann, B., Mancini, M., Pelachaud, C., Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In Proc. of Int. Gesture Workshop. S. Gibet, J.-F. Kamp, N. Courty (eds.), Lecture Notes in Computer Science 3881 Springer 2006, pp. 188-199, 2006.

Heylen D., Head gestures, gaze and the principles of conversational structure. International Journal of Humanoid Robotics (IJHR), 3(3), September, 2006.

Heylen, D., Bevacqua, E., Tellier, M., Pelachaud, C., Searching for prototypical facial feedback signals. In Proceedings of 7th International Conference on Intelligent Virtual Agents IVA 2007, pages 147--153, Paris, France, 2007.

Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., Vilhjalmsson, H., Why conversational agents do what they do? Functional representations for generating conversational agent behaviour. the first Functional Markup Language workshop. The Seventh International Conference on Autonomous Agents and Multiagent Systems Estoril, Portugal, 2008.

Johnson, W. L., Vilhjálmsson, H., Marsella, S., Serious Games for Language Learning: How Much Game, How Much AI? 12th International Conference on Artificial Intelligence in Education, Amsterdam, The Netherlands, July. 2005.

Kopp, S., Wachsmuth, I., Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds,* 15(1), 39-52, 2004.

Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., Stocksmeier, T., Modeling Embodied Feedback with Virtual Humans. Lecture Notes in Computer Science, 4930:18. 2008.

Lakin, J. L., Jefferis, V. A., Cheng, C. M., Chartrand, T. L., Chameleon effect as social glue: Evidence for the evolutionary significance of nonconsious mimicry. Nonverbal Behaviour, 27(3):145—162. 2003.

Lee, J., Marsella, S., Nonverbal behaviour generator for embodied conversational agents. In Proceedings of 6th International Conference on Intelligent Virtual Agents, volume 4133 of Lecture Notes in Computer Science, pages 243–255, Marina Del Rey, CA, USA, 2006. Springer.

Maatman, R. M., Gratch, J., Marsella, S., Natural behaviour of a listening agent. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, and T. Rist, editors, Proceedings of 5th International Working Conference on Intelligent Virtual Agents,
volume 3661 of Lecture Notes in Computer Science, pages 25--36, Kos, Greece, 2005. Springer.

Mancini, M., Pelachaud, C., Distinctiveness in multimodal behaviors, Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'08, Estoril Portugal, May 2008.

McNeill, D., Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago, IL. 1992.

Morency, L., Sidner, C., Lee, C., and Darrell, T., Contextual recognition of head gestures. In Proceedings of the 7th International Conference on Multimodal Interfaces, pages 18{24. ACM New York, NY, USA. 2005.

Morency, L.-P., de Kok, I., and Gratch, J., Predicting listener backchannels: A probabilistic multimodal approach. In Prendinger, H., Lester, J. C., and Ishizuka, M., editors, *Proceedings of 8th International Conference on Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, Tokyo, Japan. Springer. 2008.

Moreno, R., Animated software pedagogical agents: How do they help students construct knowledge from interactive multimedia games? In R. Lowe and W. Schnotz, eds. Learning with Animation. Cambridge University Press, in press.

Moubayed, S. A., Baklouti, M., Chetouani, M., Dutoit, T., Mahdhaoui, A., Martin, J.-C., Ondas, S., Pelachaud, C., Urbain, J., Yilmaz, M., Generating Robot and Agent Backchannels During a Storytelling Experiment. In ICRA, Pasadena, California, 2008.

Murray, L., Trevarthen, C., Emotional regulation of interactions between two-month-olds and their mothers. *Social perception in infants*, pages 101–125. 1985.

Nadel, J., Prepin, K., Okanda, M., Experiencing contigency and agency: first step toward self-understanding ? In Hauf, P., editor, *Making Minds II: Special issue of Interaction Studies 6:3 2005*, pages 447–462. John Benjamins publishing company. 2005.

Nass, C. I., Moon, Y., Morkes, J., Kim, E. Y., Fogg, B. J., Computers are social actors: a review of current research. Pages 137--162. 1997.

Ostermann, J., Face animation in MPEG-4. In Pandzic, I. And Forchheimer, R., editors, MPEG-4 Facial Animation - The Standard Implementation and Applications, pages 17--55. Wiley, England, 2002.

Peters, C., Pelachaud, C., Bevacqua, E., Poggi, I., Mancini, M., Chafai, N. E., A model of attention and interest using gaze behaviour. In Proceeding of IVA'05: Intelligent Virtual Agents, Kos, Greece, 2005.

Pelachaud, C., Multimodal expressive embodied conversational agent, ACM Multimedia, Brave New Topics session, Singapore, November, 2005.

Poggi, I., Backchannel: from humans to embodied agents. In Conversational Informatics for Supporting Social Intelligence and Interaction – Situational and Environmental Information Enforcing Involvement in Conversation workshop in AISB'05. University of Hertfordshire, Hatfield, England. 2005.

Prepin, K., Revel, A., Human-machine interaction as a model of machinemachine interaction: how to make machines interact as humans do. *Advanced Robotics*, 21(15):1709–1723. 2007.

Pure Data. http://www.puredata.org.

Reeves, B., Nass, C., The media equation: how people treat computers, television, and new media like real people and places, Cambridge University Press, 1996.

Schegloff, E., Sacks, H., Opening up closings. Semlotica, 8(4), 289-327., 1973.

de Sevin, E., Thalmann, D., "A motivational Model of Action Selection for Virtual Humans", In Computer Graphics International (CGI), IEEE Computer SocietyPress, New York, 2005.

Soussignan, R., Nadel, J., Canet, P., Girardin, P., Sensitivity to social contingency and positive emotion in 2-month-olds. *Infancy*, 10(2):123–144. 2006.

Striano, T., Henning, A., Stahl, D., ensitivity to social contingencies between 1 and 3 months of age. *Developmental Science*, 8(6):509–518. 2005.

Thiébaux, M., Marsella, S., Marshall, A., Kallmann, M., SmartBody: behaviour realization for embodied conversational agents. In Proceedings of 7th Conference on Autonomous Agents and Multi-Agent Systems, pages 151—158, 2008.

Thórisson, K. R., List, T., Pennock, C., Dipirro, J., Whiteboards: Scheduling blackboards for semantic routing of messages & streams. In AAAI-05 Workshop on Modular Construction of Human-Like Intelligence, pages 8--15, 2005.

Tyrrell, T., Defining the action selection problem. In Associates, Lawrence, ed. : the fourteenth annual conf. of the Cognitive Society. 1992.

Van baaren, R. B., Mimicry: a social perspective, 2003.
http://webdoc.ubn.kun.nl/mono/b/baaren_r_van/mimi.pdf[10.02.2006].

Vilhjálmsson, H. H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thórisson, K. R., van Welbergen, H., van der Werf, R. J., The Behaviour Markup Language: Recent developments and challenges. In Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., and Pelé, D., editors, Proceedings of 7th International Conference on Intelligent Virtual Agents, volume 4722 of Lecture Notes in Computer Science, pages 99--111, Paris, France. Springer, 2007.

Ward, N., Tsukahara, W., Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics, 23:1177--1207, 2000.

Warner, R. M., Malloy, D., Schneider, K., Knoth, R., Wilder, B., Rhythmic organization of social interaction and observer ratings of positive affect and involvement. Journal of Nonverbal Behavior, 11(2):57–74, 1987.

**Elisabetta Bevacqua** received her Ph.D. at the University of Paris 8 in 2009. She got her master in Computer Science at the University 'La Sapienza' of Rome in 2002. Since 2001, she works on Embodied Conversational Agents; her research field includes verbal and non verbal communication, human-machine interaction and the implementation of models to simulate humans' behaviour for virtual agents, particularly while listening to a user.

**Ken Prepin** received his PhD degree in Computer Science from the University of Paris6 in 2008. He is currently doing a Post-Doc at TELECOM ParisTech in Paris. His research interest includes human-robot interaction, imitation, dynamical and real-time systems.

**Radoslaw Niewiadomski** received his PhD degree in Computer Science from the University of Perugia, Italy, in 2007. He is currently working at TELECOM ParisTech in Paris. His research interest includes embodied conversational agents, nonverbal communication of emotions and multimodal interfaces.

**Etienne de Sevin** received his PhD degree in Computer Science from VRLab EPFL, Suisse in 2006. He currently works at TELECOM ParisTech in Paris. His research interest includes decision-taking of embodied conversational agents according to internal and external factors such as perceptions and user's interest level.

**Catherine Pelachaud** is Director of Research at CNRS in the laboratory LTCI, TELECOM ParisTech. She received her PhD in Computer Graphics at the University of Pennsylvania, Philadelphia, USA in 1991. Her research interest includes representation language for agent, embodied conversational agent, nonverbal

communication (face, gaze, and gesture), expressive behaviors and multimodal interfaces. She has been involved and is still involved in several European projects related to multimodal communication (EAGLES, IST-ISLE), to believable embodied conversational agents (IST-MagiCster, FP5 PF-STAR), emotion (FP5 NoE Humaine, FP6 IP CALLAS, FP7 STREP SEMAINE) and social behaviors (FP7 NoE SSPNet).