ABSTRACT
In this paper we propose an extension of the current SAIBA architecture. The new parts of the architecture should manage the generation of Embodied Conversational Agents’ reactive behaviors during an interaction with users both while speaking and listening.

General Terms
1. INTRODUCTION
SAIBA [13] is an international research initiative whose main aim is to define a standard framework for the generation of virtual agent behavior. It defines a number of levels of abstraction (see Figure 1), from the computation of the agent’s communicative intention, to behavior planning and realization.

The Intent Planning module decides the agent’s current goals, emotional state and beliefs, and encodes them into the Function Markup Language (FML) [3] (this language is still being defined). To convey the agent’s communicative intentions, the Behavior Planning module schedules a number of communicative signals (e.g., speech, facial expressions, gestures) which are encoded with the Behavior Markup Language (BML). It specifies the verbal and nonverbal behaviors of ECAs [13]. Each BML top-level tag corresponds to a behavior the agent is to produce on a given modality: head, torso, face, gaze, body, legs, gesture, speech, lips.

In a previous work we proposed a first approach to the FML: the FML-APML language [7]. 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: its emotional states, beliefs and goals. It originates from the APML language [1] which uses Isabella Poggi’s theory of communicative acts. 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.

However, we believe that FML alone cannot encompass all the behaviors that people perform during an interaction. Some of them do not derive uniquely from a communicative intention, they appear rapidly as a dynamic reaction to external or internal events. For example, a person engaged with friends in conversation will respond to their laughter or he could react to an unexpected shift of the other party’s gaze and look (unconsciously) in the same direction.

We think that, to perform these behaviors type, the ECAs must be able, when a new event occurs (expected or not), to compute immediate reaction (Reactive Behavior module), to select between this reaction and the previously planned behavior (action-selection module), and if necessary, to re-plan behavior dynamically (FML chunked representation). In the next Section, we propose an extension of the current SAIBA architecture that should manage these tasks. Then we will explain how this architecture allows us to generate both speaker’s and listener’s behaviors.

cesses needed for dialogue generation. However, even while the Intent Planner module to execute all the cognitive processes triggered by external and internal events. We support that, while speaking, the system will go mainly through the Intent Planner module to execute all the cognitive processes needed for dialogue generation. However, even while speaking, the agent could perform some reactive behaviors, like smiling back to the listener’s smile. On the other hand, while in the role of the listener, the agent’s behavior could be mainly reactive, since previous research has shown that the listener’s behaviour is often triggered by the verbal and nonverbal signals performed by the speaker [6, 14]. However, even while listening, the agent can intentionally display some signals to show the other party what it thinks about the speech, for example that it agrees or not, believes or not and so on. In conclusion, in both interactive roles, the ECA system must be able to generate cognitive and reactive behaviors.

In particular, when going through the cognitive process, some information in the FML can help the system to generate the right behavior according to the current role of the agent. In fact, that during a human-human communication, participants know exactly where they stand into the interaction. They know when they are speaking or listening, if they aim to give the turn to elicit an answer from the other party. They recognize when they can take the turn or when they have to insist to obtain it. Such a knowledge drives the interlocutors’ behavior. For example, if a participant wants to communicate his agreement towards the content of the speech, he will just nod the head if he is listening otherwise he will express his agreement with a full sentence if he is speaking. To fit well in an interaction with users, a conversational agent should know which is its role at any moment of the communication in order to show the right behavior. That is why the FML should contain tags for the turn management. This type of tag would not only influence the choice of the appropriate behavior to convey a certain communicative intention, like in the example described above, but also determine the generation of particular behavior signals. For example, if the agent wants to take the turn, it can open its mouth and emit short sounds to make the user let him the floor.

3. REALTIME APPLICATIONS

The proposed architecture in Figure 2 can be easily applied to generate the agent’s behavior both while speaking and listening. In both roles the agent can perform behaviors derived from its communicative intentions and reactive responses triggered by external and internal events. We support that, while speaking, the system will go mainly through the Intent Planner module to execute all the cognitive processes needed for dialogue generation. However, even while speaking, the agent could perform some reactive behaviors.

Figure 2: Proposed extension of the current SAIBA architecture. Three elements are added to the Behavior Planner module: Reactive Behavior (see Section 4.1), Action Selection (see Section 4.2) and the FMLchunk (linking ‘Action Selection’ to ‘FML to BML’ elements, see Section 4.3).
3.2 Empathy

Empathy is commonly defined as the capacity to “put yourself in someone else’s shoes to understand her emotions” [11]. To be empathic assumes one is able to evaluate the emotional dimension of a situation from the point of view of another person.

Magalie Ochs et al. [10] have proposed a model of empathic emotions elicitation during a dialog. From the subjective evaluation of the interlocutor’s speech, the Intent Planner generates the FML representing the empathic responses to be displayed by the agent. These empathic responses can be simple as well as complex expressions (e.g. superposition of empathic and egocentric emotions) [9]. This FML is sent to the Behavior Planner which translates it in behavioral signals.

The empathic expressions should be distinguished from the mimicry of emotional expressions [2, 12]. While the first may result in various emotional responses, the second consists in unconscious imitation of the facial expressions of the interlocutor. According to Dimberg et al. [2] these facial expressions are difficult to inhibit voluntary. This type of emotional expressions can not be generated by the Intent Planner. They ought to be specified more reactively. We believe these mimicry of emotional expressions have to be computed directly by the Reactive Behavior process.

4. MODIFICATION

In the next subsections we present the modifications we have brought to the SAIBA platform.

4.1 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, the emotions of the agents can change throughout the dialogue. However this mutual adaptation is also, in some other way, mostly reactive, just as a dynamical coupling with the partner: the listener will give backchannels, the partners may imitate each other, they may synchronise, or slow down or speed up their rhythms of production.

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: this dynamical coupling needs reactivity (realtime perception) and sensitivity (realtime adapted actions). For this reason, the ReactiveBehavior module has a certain autonomy from the rest of the architecture. It will short-cut the Intent Planner, much more reactive and 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 for instance an Action Selection module.

4.2 Action Selection

The Action Selection receives propositions of actions from the intention planner in FML and the Reactive Behavior module in BML (see Figure 2) and send the chosen action (in FML or BML) to the FMLtoBML module. The Action Selection allows the agent to adapt interactively to the user’s behaviors by choosing between actions coming from the Reactive Behavior module and from the intention planner. That is the Action Selection module chooses between a more cognitive-driven or a more reactive-driven behavior.

More precisely, the intent planner module and the Reactive Behavior module can propose conflicting actions. The action selection module has to decide which action is the most appropriate. This selection is made by considering the user’s interest level as well as the intentions and emotional states of the ECA. To enable the Action Selection module to make a choice, actions are associated to priorities. These priorities are computed depending on the importance the ECA gives to communicate a given intent. Importance of a communicative intent is represented by the importance tag of APML-FML [8].

4.3 FML chunk

To interact with users, the ECA system must generate the agent’s behavior in real-time. Computing the agent’s animation from a large FML input file, is that contains several communicative intentions, could create an unacceptable delay that would slow down the agent’s response, making the whole interaction unnatural. That is why we think that FMLs should be cut in smaller chunks when needed.

Therefore, we suggest that the FML language should contain additional information to specify if a FML command belongs to a larger FML, which is its order in the subset and how long is the original FML. Knowing the duration of the original FML would help the process of behavior planning. For example, a non verbal signal, bound to a minimum duration time, could start in a FML chunk if the original FML is long enough to allow its whole animation.

The decomposition of FML in a subset of chunks asks for the implementation of a feedback system between the modules of the SAIBA architecture. In order to plan or re-plan the agent’s intentions, the Intention Planner module needs to be informed about the current state of the FML that it has generated. Possible states of a FML are: “playing”, “completely played”, “discarded”, “interrupted”.

5. CONCLUSIONS

In this paper we discussed how some aspects of interactions can be managed within SAIBA. In our opinion reactive behaviors during an interaction cannot be managed properly
in the current architecture. Thus we proposed its extension as well as some examples of scenarios/applications of it. The new nodules of the architecture allows Embodied Conversational Agents for reactive behaviors during an interaction with users both while speaking and listening.

6. ACKNOWLEDGMENTS

This work has been funded by the STREP SEMAINE project IST-211480 (http://www.semaine-project.eu), the IP-CALLAS project IST-034800 (http://www.callas-newmedia.eu) and the project MyBlog-3D, contract ANR/06 TLOG 023

7. ADDITIONAL AUTHORS

8. REFERENCES


