Reactive behaviors in SAIBA architecture

Elisabetta Bevacqua
Telecom ParisTech
46 rue Barrault, 75013
Paris, France
bevacqua@telecom-paristech.fr

Ken Prepin
Telecom ParisTech
46 rue Barrault, 75013
Paris, France
prepin@telecom-paristech.fr

Etienne de Sevin
Telecom ParisTech
46 rue Barrault, 75013
Paris, France
de-sevin@telecom-paristech.fr

Radosław Niewiadomski
TELECOM ParisTech
46 rue Barrault, 75013
Paris, France
niewiado@telecom-paristech.fr

Catherine Pelachaud
Telecom ParisTech
46 rue Barrault, 75013
Paris, France
pelachaud@telecom-paristech.fr

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.

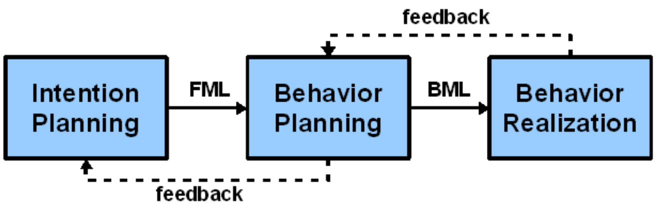

Figure 1: Saiba

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 laugth or she 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 replan 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. In Section 2 we present some scenarios/applications that can be realized with the exptended architecture. In Section 4.1 we present how adaptation (viewed here as reaction) between interactants is possible in the new SAIBA architecture. In Section 4.2 we will argue the importance of an *Action Selection* module that selects the appropriate behavior the agent should display. We also suggest in Section 4.3 that, for real-time purpose, FML input files should not be sent as a whole but in chunks. Finally, we will describe some examples that make

Cite as: Reactive behaviors in SAIBA architecture, E. Bevacqua, K. Prepin, E. de Sevin, R. Niewiadomski and C. Pelachaud, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

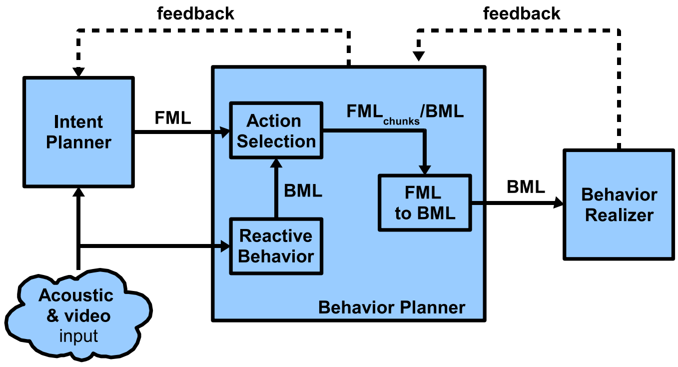

Figure 2: Proposed extension of the current SAIBA architecture. Three elements are added to the Behavior Planner module: Reactive Behavior (see Section 4.1), Action Selection (see Section 4.2)and the FMLchunk (linking 'Action Selection' to 'FML to BML' elements, see Section 4.3).

use of this architecture.

2. REALTIME INTERACTIONS

The SAIBA architecture is primarily dedicated to verbal-driven interactions. In these interactions, the speech is used to transmit sense and meaning to a partner. FML represents these meanings by mean of communicative intentions. However speech alone is not enough to enable realtime verbal communication to take place between speaker and listener. The speaker needs feedbacks from the listener, and the listener needs that the speaker adapts its speech depending on these feedbacks. There is a need of realtime adaptation of the agents to both, the context and the reaction of their partner.

To be involved in natural verbal interactions with humans, we believe that the Behavior Planner module needs to be modified. This module should be able to receive visual and acoustic input signals (described with BML tags) and to influence agent's actions in a very reactive way. We have added three elements to the Behavior Planner module (see Figure 2): one element to compute reactive response (Reactive Behavior), another to select between this 'on the fly' reaction and the preplanned behavior (Action Selection), and finally one element offering the capacity to replan the behavior whenever necessary (thanks to a Chunked FML representation).

This new version of the Behavior Planner module will be both influenced by the higher level communicative intentions conveyed by FML and be reactive to physical events.

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 suppose 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 contains 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 behavioral 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.

3.1 Mimicry

Several researches have shown that in human-human interactions people tend to imitate each other. This copying behaviour, called *mimicry*, has been proven to play an important role during conversations. For example, when fully engaged in an interaction, mimicry of behaviors between interactants may happen [5]. Mimicry behavior can be performed consciously or unconsciously. During the interaction a person could decide to imitate the other party's smile in order to show that he shares his appreciation. To generate this type of behavior, the architecture proposed in Figure 2 would generate a FML containing the communicative intention of mimicry; afterwards the Behavior Planner would translate it in behavioral signals according to the behavior performed by the other party, in this example, the chosen signal would be a smile.

On the other hand one could be completely unaware of mimicking the person he is interacting with. Such a behavior, called by Lakin "chameleon effect" [4], helps to create affiliation and rapport. To generate this type of reactive and unconscious behavior, we propose that the Behavior Planner should include a sub module, the Reactive Behavior in Figure 2. Such a module, triggered by the user's acoustic and non verbal signals, generates the mimicry behavior in BML format. No need for FML in this situation since the agent's behavior is unintentional and since, being a reactive behavior, its generation should be as faster as possible.

3.2 Empathy

Empathy is commonly defined as the capacity to "put your-self 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, getting directly input signals, i.e. the BML coming from the human (see Figure 2), as well as the currently planned actions, i.e. the BML produced at the output of the BehaviorPlanner.

With these two sources of information, the Reactive Behavior module will propose to the Action Selection module (see Section 4.2) two different types of data. It can propose adaptation of the current behavior. By comparing its own actions to the actions of the speaker at a very low level, among other thing tempo or rhythm of signal production; for example it can propose to slow down or speed up behaviors. This type of propositions may enable synchronisation, or similarity of tempo with the user. The second type of data proposed by the Reactive Behavior are full actions. By extracting from the user's behavior salient events, it will propose actions such as performing a backchannel, imitating the user or following its gaze.

Finally the Reactive Behavior will be able to propose real-time reactions or adaptations to the user's behavior 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 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 Section 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, ie 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-211486 (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

[1] B. D. Carolis, C. Pelachaud, I. Poggi, and M. Steedman. APML, a mark-up language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, *Lifelike Characters. Tools, Affective Functions and Applications*. Springer, 2004.

[2] U. Dimberg, M. Thunberg, and S. Grunedal. Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition & Emotion*, 16(4):449–471, 2002.

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

[4] J. L. Lakin and T. L. Chartrand. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, 14:334–339(6), July 2003.

[5] J. L. Lakin, V. A. Jefferis, C. M. Cheng, and T. L. Chartrand. Chameleon effect as social glue: Evidence for the evolutionary significance of nonconsious mimicry. *Nonverbal Behavior*, 27(3):145–162, 2003.

[6] R. M. Maatman, J. Gratch, and S. Marsella. Natural behavior 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.

[7] M. Mancini and C. Pelachaud. Distinctiveness in multimodal behaviors. In L. Padgham, D. C. Parkes, J. Müller, and S. Parsons, editors, *Proceedings of Conference on Autonomous Agents and Multi-Agent Systems (AAMAS08)*, 2008.

[8] M. Mancini and C. Pelachaud. Distinctiveness in multimodal behaviors. In L. Padgham, D. C. Parkes, J. Müller, and S. Parsons, editors, *Proceedings of Conference on Autonomous Agents and Multi-Agent Systems (AAMAS08)*, 2008.

[9] R. Niewiadomski, M. Ochs, and C. Pelachaud. Expressions of empathy in ecas. In H. Prendinger, J. C. Lester, and M. Ishizuka, editors, *in Proceedings of 8th International Conference on Intelligent Virtual Agents, IVA 2008*, volume 5208 of *Lecture Notes in Computer Science*, pages 37–44, Tokyo, Japan, 2008. Springer.

[10] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine interaction. In *Autonomous Agent and Multi-Agent Systems (AAMAS)*, 2008.

[11] E. Pacherie. *L'empathie*, chapter L'empathie et ses degrés, pages 149–181. Odile Jacob, 2004.

[12] W. Sato and S. Yoshikawa. Spontaneous facial mimicry in response to dynamic facial expressions. In *Proceedings of the 4th International Conference on Development and Learning*, pages 13–18, 2005.

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

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