Greta: a SAIBA compliant ECA system

Maurizio Mancini[1] Radosław Niewiadomski[2] Elisabetta Bevacqua[2] Catherine Pelachaud[2]

[1]InfoMus Lab, DIST – University of Genova
via Causa 13
I-16145, Genova - Italy
mauman1974@gmail.com

[2]TELECOM ParisTech, Université Paris 8
46 rue Barrault
75013, Paris - France
radoslaw.niewiadomski, elisabetta.bevacqua, catherine.pelachaud@telecom-paristech.fr

ABSTRACT
We present a modular, distributed, and real-time architecture of an Embodied Conversational Agent (ECA) called Greta. It follows SAIBA and MPEG4 standards. We describe also a language to describe the communicative intentions of our agent.

Keywords
ECA architecture, multimodal behavior, SAIBA, BML, FML

1.INTRODUCTION
In this paper we describe the architecture of an embodied conversational agent (ECA) called Greta. Greta is 3D agent that is able to communicate with the user using verbal and a set of nonverbal channels like gaze, head and torso movements, facial expressions and gestures. It follows the SAIBA framework that defines modular structure, functionalities and communication protocols for ECA systems. Moreover our agents follows the MPEG4 standard of animation. As a consequence our agent could be integrated with existing softwares, and, furthermore, each of its modules could be replaced by other ones. Last but not least it is optimized to be used in applications in a real-time to interact with users.

2.BACKGROUND
SAIBA [8] 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 Fig. 1), from the computation of the agent's communicative intention, to behavior planning and realization.

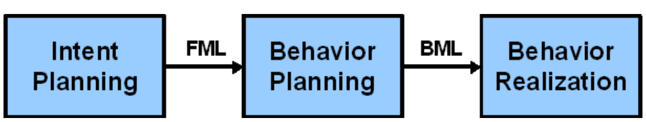

Figure 1. SAIBA framework (from [8]).

The *Intent Planning* module decides the agent's current goals, emotional state and beliefs, and encodes them into the *Function Markup Language* (FML)[1] [3]. 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 [8]. 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. Finally the task of the third component of the SAIBA framework, **Behavior Realization**, is to realize the behavior scheduled by the **Behavior Planning**.
There exists several implementations like *SmartBody* [6] and *BMLRealizer* [1] that are SAIBA compatible. In the following sections we present our agent system.

3.FML-APML LANGUAGE
Our agent uses the FML-APML language [3]. FML-APML is based on APML [2], allowing the specification of the agent's *communicative intentions*, that is, what the agent aims to communicate to the user: for example its beliefs and goals. The duration of each communicative intention can be specified explicitly in seconds or in relation to a speech act. The other novelty is the possibility to define not only the speaker's intentions but also the listener's ones. Finally, in FML-APML information on the world can be specified to communicate some physical or abstract properties of objects, persons, events.

4.GRETA ARCHITECTURE
Figure 2 illustrates the architecture of Greta agent. The **Behavior Planning Module** receives as input the agent's communicative intentions written in FML-APML and generates as output a list of BML signals. These signals are sent to the **BML Engine** (corresp. *Behavior Realization)* that generates the MPEG4 FAP-BAP files. Finally, the animation is played in the **FAP-BAP Player**.
Our architecture has also a **Speaker and Listener Intent Planning** that belongs to the SAIBA Intent Planning module. Such a planner is able to generate in real-time the agent's communicative intention while in the role of the speaker or of the listener. For the moment we have implemented only the listener part.

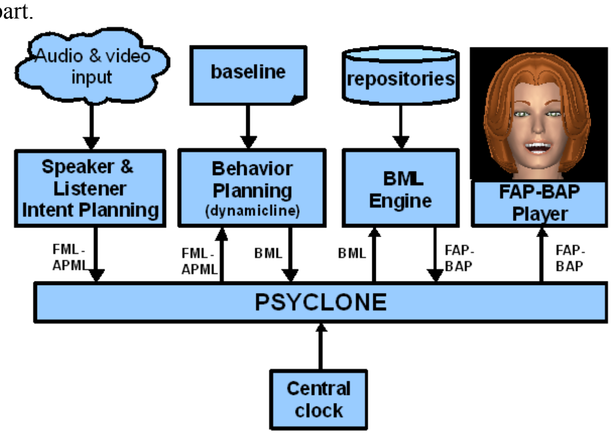

Figure 2. Greta architecture.

[1]This language is still being defined.

All modules in the Greta architecture are synchronized through a central clock and communicate with each other through a *whiteboard*. For this purpose we use Psyclone messaging system [7] which allows modules and applications to interact together, even if they are running on separate machines connected through TCP/IP. The system has a very low latency time that makes it suitable for real-time applications.

In the following subsections we describe each module of Greta.

BEHAVIOR PLANNING. The Behavior Planning takes as input both the agent's communicative intentions specified in FML-APML and the agent's *baseline*.

The agent's baseline contains information on the preference the agent has in using each available communicative modality (head, gaze, face, gesture, and torso) and on the expressive quality of each of these modalities. *Expressivity* is defined by a set of parameters that affect the qualities of the agent's behavior: e.g. wide vs. narrow gestures, smooth vs jerky movements and so on. Depending on the agent's baseline, the system computes the *dynamicline* associated with the each intention: for example an agent that tends to perform smooth and slow gestures, as defined in its baseline, could perform fast and jerky movements if it aims to communicate anger. The dynamicline, together with the current communicative intention is then used to select the multimodal behavior that best conveys the given intention: e.g., an agent that is always low aroused (baseline), in a joy state (communicative intention) could produce just a light smile, without moving the rest of his body; on the other hand, a very expressive agent (baseline) with the same communicative intention could produce a combination of signals like smiling, stretching his arms and body, all at the same time.

BML ENGINE. This module generates the animation of Greta in MPEG4 format. The input to the module is specified in BML language. It contains a text to be spoken by Greta and/or a set of nonverbal signals to be displayed. BML used by Greta contains a number of extensions compared with the standard that allows us to define, among others, the expressivity and the intensity of each signal. It solves also eventual conflicts between the signals that use the same modality, and generates the smooth animation using interpolation algorithms. The module generates animation files in MPEG4 standard [5]. For this purpose it uses facial expressions, gestures, torso movements etc. which are predefined in repository files. The speech is generated by an external TTS and the lips movements are added to the animation. Finally, both the animation and audio file in *wav* format are sent to the FAP-BAP Player via Psyclone.

FAP-BAP PLAYER. The FAP-BAP Player receives the animation (a sequence of FAP frames) generated by the BML Engine and plays it by drawing the Greta's head and body in a graphic window. The animation is generated with the use of OpenGL library.

SYNCHRONIZATION. The synchronization of all modules in the distributed environment is ensured by a **Central Clock** which broadcasts regularly timestamps through Psyclone.

LISTENER INTENT PLANNING. Researches have shown that listener's responses (also called backchannel signals) are often emitted according to the listener's mental state as well as the verbal and non verbal behavior performed by the speaker [4]. The **Listener Intent Planning** computes the backchannels of the agent when it dialogs with a user. It takes as input data from video and audio streams coming out from a webcam and microphone and selects which user's behaviors (for example, a head nod or a variation in the pitch of the user's voice) that could elicit a backchannel from the agent. Once a backchannel must be emitted this module also decides which communicative intentions (agreement, refusal, interest...) the agent should convey. The choice is done according to the agent's mental state that describes how the agent is reacting to the user's speech (if it agrees, refuses, understands... what is being said).

5. CONCLUSION

We have described a SAIBA compliant architecture for our Embodied Conversational Agent, Greta. Our system can generate listener behavior in real-time. In the future we aim at expanding our system by adding the Speaker Intent Planning module.

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 by the ANR-MyBlog-3D(https://picoforge.int-evry.fr/cgi-bin/twiki/view/Myblog3d/Web/WebHome).

7. REFERENCES

[1] http://cadia.ru.is/projects/bmlr/

[2] De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M. 2004. APML, a Mark-up Language for Believable Behavior Generation, In: Prendinger, H., Ishizuka, M. (eds.), Lifelike Characters. Tools, Affective Functions and Applications, Springer.

[3] Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., Vilhjalmsson, H. 2008. Why Conversational Agents do what they do?Functional Representations for Generating Conversational Agent Behavior. The First Functional Markup Language Workshop. In 7th International Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal.

[4] Maatman, R. M., Gratch, J., Marsella, S. 2005. Natural behavior of a listening agent. In 5th International Conference on Interactive Virtual Agents. Kos, Greece.

[5] Ostermann, J. 2002. Face Animation in MPEG-4, in: Pandzic, I.S., Forchheimer, R., (eds.), MPEG-4 Facial Animation - The Standard Implementation and Applications, Wiley, England, pp. 17-55.

[6] Thiebaux, M., Marsella, S., Marshall, A. N., Kallmann, M. 2008. SmartBody: behavior realization for embodied conversational agents. In Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal, 151-158.

[7] Thörisson, K.R., List, T., Pennock, C., DiPirro, J. 2005 Whiteboards: Scheduling blackboards for interactive robots. In Twentieth National Conference on Artificial Intelligence.

[8] Vilhjalmsson, H., Cantelmo, N., Cassell, J., Ech Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thörisson, K. R., van Welbergen, H., van der Werf, R. 2007. The Behavior Markup Language: Recent developments and challenges. In 7th International Conference on Intelligent Virtual Agents.