Suggestions for Extending SAIBA with the VIB Platform

Florian Pecune[1], Angelo Cafaro[1], Mathieu Chollet[2], Pierre Philippe[2], and Catherine Pelachaud[1]

[1]CNRS-LTCI, Télécom ParisTech and [2]Institut Mines-Télécom, Télécom ParisTech, CNRS-LTCI

1 Main Research Themes

Virtual Interactive Behavior (VIB) is a SAIBA compliant platform which supports the creation of socio-emotional ECAs. It takes as input utterances augmented with communicative functions and emotional states specified in FML-APML (Carolis et al., 2004). The ECA spoken utterances are enriched by nonverbal behaviors NVB (gaze, facial expression, gesture). The choice of NVB can be modulated by the definition of the *dynamicline* that is associated to each ECA (Mancini and Pelachaud, 2008). The dynamicline specifies the preferred modalities and the expressive parameters of each modality (Huang and Pelachaud, 2012). These parameters act on the quality of execution of a behavior such as its speed and acceleration, its fluidity and amplitude.

VIB allows the agent to be an active interactant (speaker or listener). The ECA can decide which social attitudes to display towards its conversation partners (Pecune, 2013). These social attitudes can be shown by the choice of its intentions (Callejas et al., 2014), the type of the ECAs reactions (e.g. exhibit a polite or amused smile (Ochs and Pelachaud, 2013)) and its capacity to be temporally aligned (Prepin et al., 2013). For example, an interpersonal attitude can be chosen at intentional level and the supporting utterances (Chollet et al., 2014) and multimodal behaviors (Ravenet et al., 2013) are produced. While most of the behavior models integrated within VIB are procedural, Ding has used machine learning techniques to drive the multimodal behaviors of the agent when saying emotional speech (Ding et al., 2013) and laughing (Ding et al., 2014).

We are currently working on endowing the agent with the capacity to show its engagement during the interaction by choosing its conversation topics (Glas in (Ochs et al., 2013)) and by making use of hetero-repetition, on expanding the expressivity model by analysing a large database of motion capture data of expressive multimodal behaviors (Fourati and Pelachaud, 2014), and on modeling group behavior during conversations.

2 Current Architectures and Standards

VIB has a modular and extensible architecture. Each module represents an ECA's functionality, therefore one ECA is defined by a particular set of modules as depicted with three examples in Fig. 1. Different formats are adopted to describe the actions that an ECA can perform, these formats represent information ranging from the cognitive level (eg. communicative intentions) to the physical level (eg. skeleton animations). These information are exchanged in the form of events and represent the input and output of the modules. Each module in VIB automatically processes the received events.

We developed a graphical user interface (GUI) that allows a user to design, instantiate and connect the modules (e.g. FML/BML reader, behavior planner, 3D engine player, gesture editor, etc.).

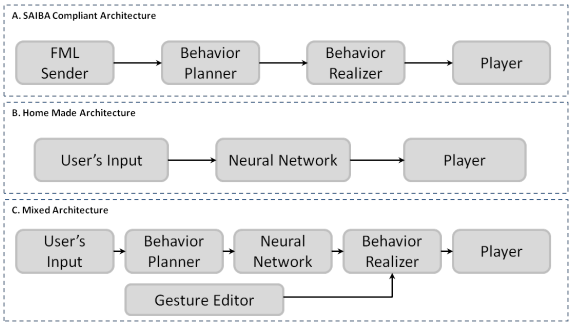


Figure 1: Three example architectures created in VIB

Core Modules These modules include a SAIBA compliant *Behavior Planner* and *Realizer* that work with our own FML-APML specification and extended BML file formats respectively. The Behavior Realizer outputs are keyframes performers used to compute the real-time animation of the agent's body and face. An external player (i.e. 3D engine) can be plugged into the system. This supports the rendering of the 3D environment with different engines (e.g. Ogre3D or Unity3D). Different agents or users can be represented in VIB and share the same player, therefore the same 3D environment.

Other Modules The libraries of movements used are not fixed and may be updated continuously. For this purpose, "editors" modules can be used to define the facial expressions by action units (AU) (Ekman and Friesen, 1978), the gestures (Kendon, 2004; McNeill, 1992), the mapping between AU and FAP (MPEG-4, 2014), the

shape of hands, etc. Network communications modules have been developed to interface with external software or to exchange events between different instances of the platform over several machines. Currently, different APIs are used such as ActiveMQ and Thrift. A Neural Network module provides an opportunity for the user to create neurons and connect them via a specific GUI. It can be used to create real-time expressions by motor resonance.

3 Future Architectures and Standards for IVAs

User Behavior Perception Many ECA systems attempted to recognize user behavior during the interaction (e.g. the SEMAINE project (Bevacqua et al., 2012)). There exist commercial and academic software for behavior recognition (e.g. head tracking with OpenCV) and mental states (e.g. emotion recognition with SHORE). However, these are often heterogeneous and require *ad-hoc* development of interfaces for integration into ECA systems. We advocate for a standard representation of the ouptut provided by these applications. The Perception Markup Language (PML) (Scherer et al., 2012), represents a first attempt. However, in addition to the certainty value and the sensory layer, PML redefines behavioral and functional levels (i.e. *behavior* and *function* layers), which could simply reuse the BML and FML specifications.

Multi-party and Multi-floor Interaction In face-to-face communication a person might be engaged in more than one conversation at the same time with different roles. For example, someone could listen to a talk and speak to the person seated next by. This moves from dyadic settings towards more complex multi-party scenarios that should be represented both at functional (i.e. FML) and behavioral (i.e. BML) levels. An FML specification addressing this aspect has been proposed by (Cafaro et al., 2014) but it has not been adopted yet by any ECA system. We also think that representing more complex configurations at functional level and not only with BML may be important to affect the subsequent produced multimodal behavior. A few examples are *1-to-many* (e.g. describing a public speech) and *many-to-many* (two groups interacting as a whole with each other).

Transforming from FML to BML in SAIBA SAIBA currently does not specify how FML should be transformed into BML. We believe that this aspect cannot be left aside of the framework with individual researchers providing their *"home made"* solutions, as this may critically impact the flexibility of integration into other systems. Previous ECA systems have mainly adopted two strategies to solve this problem that are broadly categorized as **data-driven** or **procedural** approaches (cf. introduction of Chapter 6 in (Cafaro, 2014) for a review). In general, our vision is that SAIBA should not only provide standardized interface languages but also techniques and modules that enable to properly transfer between SAIBA components the information represented by these interface languages.

4 Suggestions for Discussion

Raw Perception vs. Attention We emphasized the importance of separating low level user's behavior recognition and its interpretation. An interesting aspect is also the distinction between raw environment perception and the actual information processed by an ECA depending on its attention level. In (Balint and Allbeck, 2013), agents' perceptions are limited by their senses (i.e. solely within their field of view). However, the agent's attention level might filter out some raw information. In a crowded scene, for example, an ECA in face-to-face interaction might exclude some auditory or visual raw perceptions (e.g. other agents walking by) since its attention is focused on the interactant, but another agent or an object (e.g. a car moving fast close by) might trigger an attention shift. We suggest a discussion on how to separate raw perception and attention with emphasis on how to model attention level.

Dealing with Reactive Behaviors Reactive behaviors are part of the interaction. For instance, an agent might bend over to avoid a ball coming to him, or scream after seeing a spider landing on its shoulder. According to (Scherer, 2001), a quicker response is needed for those behavior, may be bypassing the functional planning part of the SAIBA pipeline. The question is whether these low-level quick reactions would fit in SAIBA or should be modeled externally. (Bevacqua et al., 2008) attempted to model this aspect, but on behavioral level (e.g. BML). Our question is what happens at the functional level? If an immediate reaction is required, should the pending intention be canceled, or re-scheduled to be accomplished later?

From BML to Animation Similarly to transforming FML to BML, transforming planned BML into low level parameters ready to be executed as animations by BML Realizer might be problematic. Currently there is no guarantee that playing the same BML block on two different realizers will lead to the same result. Is there a way for designers to be assured that the behaviors they are creating will be played the same way, whatever the agent could be? One solution to address this problem might be to set an additional standardization layer between the Planner and the Realizer components in SAIBA which could provide lower level information (e.g. joint rotations or animation parameters).

References

T. Balint and J. Allbeck. 2013. Whats going on? multi-sense attention for virtual agents. In R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, editors, *Intelligent Virtual Agents*, volume 8108 of *Lecture Notes in Computer Science*, pages 349–357. Springer Berlin Heidelberg.

E. Bevacqua, K. Prepin, E. de Sevin, R. Niewiadomski, and C. Pelachaud. 2008. Reactive behaviors in saiba architecture. In *AAMAS 2009 Workshop Towards a Standard Markup Language for Embodied Dialogue Acts*, pages 9–12.

E. Bevacqua, E. Sevin, S. Hyniewska, and C. Pelachaud. 2012. A listener model: introducing personality traits. *Journal on Multimodal User Interfaces*, 6(1-2):27–38.

A. Cafaro, H. Vilhjálmsson, T. Bickmore, D. Heylen, and C. Pelachaud. 2014. Representing communicative functions in saiba with a unified function markup language. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents (to appear)*, IVA '14.

A. Cafaro. 2014. *First Impressions in Human-Agent Virtual Encounters*. Ph.D. thesis, Center for Analysis and Design of Intelligent Agents, Reykjavik University, Iceland.

Z. Callejas, B. Ravenet, M. Ochs, and C. Pelachaud. 2014. A computational model of social attitudes for a virtual recruiter. In *International Conference on Autonomous Agent and Multi-Agent Systems (AAMAS)*.

B. D. Carolis, C. Pelachaud, I. Poggi, and M. Steedman. 2004. Apml, a markup language for believable behavior generation. In H. Prendinger and M. Ishizuka, editors, *Life-like characters*, Cognitive Technologies, pages 65–86. Springer.

M. Chollet, M. Ochs, and C. Pelachaud. 2014. Mining a multimodal corpus for non-verbal signals sequences conveying attitudes. In *Language Resources and Evaluation Conference (LREC)*.

Y. Ding, C. Pelachaud, and T. Artires. 2013. Modeling multimodal behaviors from speech prosody. In *13th International Conference of Intelligent Virtual Agents - IVA*.

Y. Ding, K. Prepin, J. HUANG, C. Pelachaud, and T. Artires. 2014. Laughter animation synthesis. In *International Conference on Autonomous Agent and Multi-Agent Systems (AAMAS)*.

P. Ekman and W. Friesen. 1978. *The Facial Action Coding System: A Technique For The Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, USA.

N. Fourati and C. Pelachaud. 2014. Emilya: Emotional body expression in daily actions database. In *Language Resources and Evaluation Conference (LREC)*.

J. Huang and C. Pelachaud. 2012. Expressive body animation pipeline for virtual agent. In *proceedings of 12th International Conference of Intelligent Virtual Agents - IVA*, pages 355–362.

A. Kendon. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

M. Mancini and C. Pelachaud. 2008. Distinctiveness in multimodal behaviors. In *7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 159–166, Estoril, Portugal.

D. McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

MPEG-4. 2014. `http://mpeg.chiariglione.org/standards/mpeg-4`.

M. Ochs and C. Pelachaud. 2013. Socially aware virtual characters: The social signal of smiles. *IEEE Signal Process. Mag.*, 30(2):128–132.

M. Ochs, Y. Ding, N. Fourati, M. Chollet, B. Ravenet, F. Pecune, N. Glas, K. Prpin, C. Clavel, and C. Pelachaud. 2013. Vers des agents conversationnels anims socio-affectifs. In *Interaction Humain-Machine (IHM'13)*.

F. Pecune. 2013. Toward a computational model of social relations for artificial companions. In *Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII*, pages 677–682.

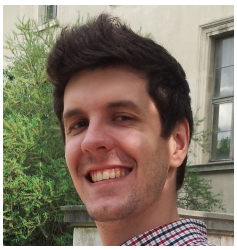
K. Prepin, M. Ochs, and C. Pelachaud. 2013. Beyond backchannels: co-construction of dyadic stancce by reciprocal reinforcement of smiles between virtual agents. In *International Conference CogSci (Annual Conference of the Cognitive Science Society)*.

B. Ravenet, M. Ochs, and C. Pelachaud. 2013. From a user-created corpus of virtual agent's non-verbal behavior to a computational model of interpersonal attitudes. In *13th International Conference of Intelligent Virtual Agents - IVA*, pages 263–274.

S. Scherer, S. Marsella, G. Stratou, Y. Xu, F. Morbini, A. Egan, A. Rizzo, and L.-P. Morency. 2012. Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In Y. Nakano, M. Neff, A. Paiva, and M. Walker, editors, *Intelligent Virtual Agents*, volume 7502 of *Lecture Notes in Computer Science*, pages 455–463. Springer Berlin Heidelberg.

K. R. Scherer. 2001. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92:120.

Biographical Sketches



Florian Pecune is a PhD candidate at the LTCI laboratory of Telecom Paristech. His research activities are focused on designing Embodied Conversational Agents able to adapt their decision making in particular social contexts. As part of the ANR MoCA project, which aims at creating a connected world of artificial companions, he proposed a model to compute the social attitude of an agent according to its goals and beliefs.

Angelo Cafaro is a postdoctoral researcher at CNRS-LTCI, Telecom ParisTech. He is doing research in the area of embodied conversational agents and serious game environments with emphasis on social interaction, group behavior and expression of social attitudes. Angelo is part of the EU FP7 Verve project, which aims at developing serious games to support the treatment of elderly people who are at risk of social exclusion. He obtained his Ph.D. from Reykjavik University in 2014. His dissertation dealt with analyzing and modeling human nonverbal communicative behavior exhibited by a virtual agent in a first greeting encounter with the user. In his dissertation he also proposed a SAIBA compliant computational model featuring a unified specification for the Function Markup Language (FML). More information is available on his personal webpage: `www.angelocafaro.info`.

Mathieu Chollet is a PhD candidate at the LTCI laboratory of Telecom Paristech. His research activities are focused on Embodied Conversational Agents and their applications for social skills training. As part of the EU FP7 TARDIS project, which aims at improving youngsters' job interview skills, he proposed behavior models of attitude expression for virtual recruiters. Additionally, he was involved as a Visiting Researcher at the Institute for Creative Technologies where he designed an interactive virtual audience architecture for public speaking training. Personal webpage: `http://perso.telecom-paristech.fr/~mchollet/`

Pierre Philippe is a research engineer at the LTCI laboratory of Telecom ParisTech. He received a Master degree in Computer Science and a M.S. degree in Artificial Intelligence from Brussels Polytechnic School. He worked as a consultant in Computer Associates, then as a research engineer in IRIDIA lab at Brussels University (Belgium) and in COIN lab at Skövde University (Sweden). He is currently modelling and programming modules for the Greta platform. His research interest includes Embodied Conversational Agents, emotions modelling and cognitive architectures.

Catherine Pelachaud is a Director of Research at CNRS in the laboratory LTCI of Telecom ParisTech. Her research interest includes embodied conversational agent, nonverbal communication (face, gaze, and gesture), expressive behaviors and socio-emotional agents.

E-Mail Contacts (`@telecom-paristech.fr`)

Florian Pecune: `florian.pecune`
Angelo Cafaro: `angelo.cafaro`
Mathieu Chollet: `mathieu.chollet`
Pierre Philippe: `pierre.philippe`
Catherine Pelachaud: `catherine.pelachaud`