Towards an Intelligent Affective Multimodal Virtual Agent for Uncertain Environments

Isabelle Hupont[1], Rafael Del-Hoyo[1], Sandra Baldassarri[2],
Eva Cerezo[2], Francisco J. Serón[2] and Diego Romero[1]
[1] Instituto Tecnológico de Aragón, Zaragoza, Spain
[2] Computer Science and Systems Engineering Department, University of Zaragoza, Spain
ihupont@ita.es, rdelhoyo@ita.es, sandra@unizar.es,
ecerezo@unizar.es, seron@unizar.es, dromero@ita.es

ABSTRACT
Affective interaction between a user and a virtual agent must be believable. The virtual actor has to behave properly, to have the capacity to talk in natural language and to express some affectivity. For achieving this goal, it is basic to provide the agent with intelligence to let him make all kind of real-time decisions in complex situations. In this paper, an architecture for the user interaction with intelligent affective multimodal virtual agents in uncertain environments is proposed. It is based on the integration of Maxine, a powerful multimodal animation engine for managing virtual agents and 3D scenarios, and PROPHET, a cognitive intelligent system based on Fuzzy Logic. The outstanding features of the proposed architecture are its affective and intelligence capabilities that make the system able to cope with decisions based on the analysis and learning of information and perceptions coming from uncertain environments.

Categories and Subject Descriptors
D.3.3 [**Models and Principles**]: User/Machine Systems – *human factors*, h*uman information processing;* I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – *Intelligent agents;*

General Terms
Algorithms, Experimentation, Human Factors.

Keywords
Multimodal interfaces, affective interaction, expressive agents, intelligent agents, real-time animation.

1. INTRODUCTION
Most research on social interfaces is related to the design of virtual agents, since human face-to-face communication has been proved to be an ideal model for designing a multimodal human-computer interface [1, 2]. A virtual agent must be believable: it has to move properly, paying special attention to its facial expressions [3] and body gestures [4], and have the capacity to talk in natural language [5]. Besides its external appearance, it must possess some affectivity, an innate characteristic in humans, for which reason it is necessary to carefully manage the emotions of the virtual agent [6]. Finally, it is basic to provide the agent with intelligence, in order to allow him to take real-time decisions in complex situations [7].

To make the virtual agent intelligent, it is important to analyze how the human mind works for correctly "modeling" the virtual agent's reasoning mechanisms. The human brain is characterized by its capacity to handle and store uncertain and confuse perceptions. People usually face problems with great uncertainty and partial, context-dependent and contradictory information. Fuzzy Logic makes it possible to model those types of problems and to find solutions similar to the ones taken by human beings. In doing so, it is possible to develop a more "cognitive" computation that tackles effectively the interaction among persons and virtual agents, how they communicate and act through words and perceptions [8].

In this paper we focus on the presentation of an architecture for integrating a multimodal engine for managing 3D virtual scenarios and characters (section 2) and a cognitive intelligent system for automatic decision-making with the ability of learning on the basis of experience (section 3). This integration will make the virtual agents intelligent, affective-aware and able to take decisions in real-time (section 4).

2. MAXINE: A PLATFORM FOR EMBODIED ANIMATED AGENTS
Maxine is a powerful animation engine for managing virtual environments and virtual actors [9]. The system is capable of controlling virtual 3D agents for their use as new interfaces in a wide range of applications. For example, it has been used in different domains such as virtual humans for PowerPoint-like presentations in real-time, control of domotic environments or interactive pedagogical agents for teaching Computer Graphics (see Figure 1).

Maxine is mainly based on the adaptation and assembly of different open source libraries for the solution of several problems related to real-time management and interaction with 3D scenarios. In particular, special attention has being paid to the development and animation of the virtual 3D agents: the actors are endowed with body and facial animation, lip-synch, and synthesized voice. Agents' animation elements can be

AFFINE '09, November 6, 2009 Boston, MA.

dynamically created and manipulated by means of a simple command interface. These commands can be executed via script-files when initiating the application or during execution, or they can be introduced through the text console each time.

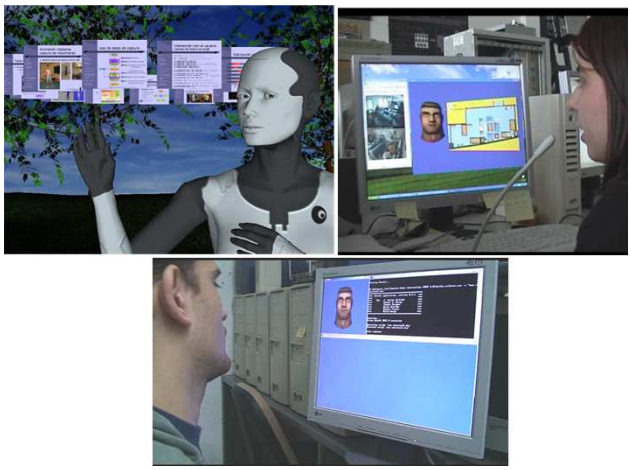

Figure 1. Screenshots from Maxine's applications.

One of the most outstanding features of the system is its affective capabilities. Maxine is able to analyze in real-time the user's facial expressions for recognizing his emotional state [10]. Virtual agents can also express their emotions by showing different facial expressions and by modulating their voices.

The system supports real-time multimodal interaction with the user through different channels (text, voice, mouse/keyboard and image). The agent reacts to user actions: for example, should the user key something in, the virtual presenter will interrupt the presentation; should the user change his position, the 3D actor look/orientation will change; if a lot of background noise is detected, it will request silence, etc. Agents can also be minimally deliberative: it can elicit an answer through the user's voice interaction according to the recognition of a set of patterns associated to fixed answers and is able to change the emotional state of the virtual agent depending on a set of keywords detected in the conversation with the user (e.g. if the user insults the actor, the later will reply with an angry voice and facial expression).

For achieving a more realistic human-computer interaction through the virtual agents, it turns out necessary to provide Maxine with more intelligence, in order to enrich the virtual character's behaviour by moving its responses from reactive to cognitive schemes.

3. PROPHET: A COGNITIVE INTELLIGENT BRAIN

PROPHET is a knowledge-based intelligent system that enables automatic decision-making and self-learning. The system has already been used successfully in different domains such as logistics decision-making systems [11], real-time networking management [12] and natural language automatic analysis.

The main distinguishing feature of the system is its ability to integrate different softcomputing technologies. Among other techniques, it makes use of Fuzzy Logic as a real-time method to achieve computation with words and concepts. Computation with words, through the use of Fuzzy Logic, allows to model systems' inputs in a human-like natural language. Thanks to it, PROPHET is able to automatically define rules for making decisions in terms of perceptions (e.g. "feasible", "extremely good", "optimal", etc.).

PROPHET consists of a set of modules for pre-processing, integrating, extracting information and making decisions in a flexible way under uncertain contexts (see Figure 2). The system is based on a state machine in order to increase its scalability: each module generates events that are treated asynchronously inside the state machine. The database management is done by means of Hibernate [13], a powerful, high performance object/relational persistence and query service framework. The different modules that compose the system are presented in the following sections.

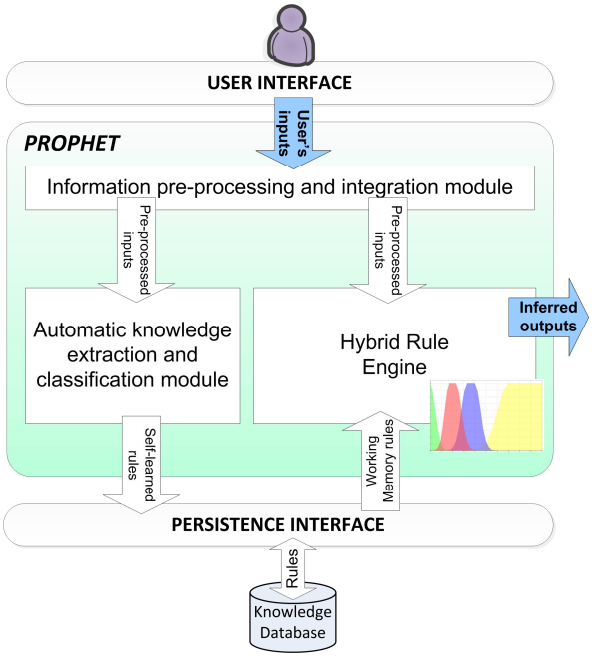

Figure 2. PROPHET's general architecture.

3.1 Information Pre-processing and Integration Module

This module is in charge of inputs' pre-processing and integration. The inputs come from any source of information: real-time sensors, databases, etc. The system has several pre-defined filters (e.g. data normalization filters), but also allows the free definition of any kind of expert pre-processing rules (e.g. truncate an input value if greater than a given threshold, accumulate data values…).

3.2 Automatic Knowledge Extraction and Classification Module

This module extracts knowledge from input data, by means of softcomputing-based algorithms. Thanks to the use of Neuro-Fuzzy techniques [14], the module has the capability of self-extracting and self-learning new fuzzy decision rules from historical data. The softcomputing algorithms (Neural Networks, Genetic Algorithms and Support Vector Machines) allow to predict and classify possible future variables' behaviours in view

of the incoming inputs, which are taken into account in the decision rule definition process.

3.3 Hybrid Rule Engine

The rule inference engine is the main sub-system of PROPHET. It is in charge of rule-based decision-making tasks. It is a hybrid rule inference engine since it can both deal with crisp rules (applied to exact inputs' values) and execute inference from rules that handle fuzzy concepts. The elements in the Working Memory are not only the rules pre-defined by an expert, but also the set of automatically self-learned decision rules created by the knowledge extraction and classification module.

4. A NOVEL ARCHITECTURE TO MAKE MAXINE INTELLIGENT AND AFFECTIVE-AWARE

Today, we are working on providing Maxine with PROPHET's cognitive intelligent brain. That way, virtual agents will be able to generate and learn cognitive -and not mainly reactive- responses in real-time and will be provided with both general and affective artificial intelligence. Fuzzy Logic will allow the agents to manage information and learn in high-uncertainty environments in a way similar to humans.

The overall architecture of the proposed system is shown in Figure 3. Following PROPHET's philosophy, the whole system will be controlled by a state machine (generation of asynchronous events). In the sections which follow, all of the modules that comprise the architecture are explained in detail.

4.1 Perception Module

The perception module allows the multimodal interaction between the user and the virtual agent: via keyboard, mouse, voice (natural language conversation), webcam, background microphones, positioning tracker, etc.

Besides from recognizing the user's facial expressions, more modules are being added to the system for the extraction of affective information. In particular, a keyboard pulses and mouse clicks analyzer is now being developed to detect states of boredom, confusion, frustration and nervousness of the user. We are also working on an affective analyzer of speech contents and considering the analysis of prosody in speech in a near future. Another source of affective information we are considering is the one provided by an Eye Tracker system (mainly eye gaze analysis).

4.2 Cognitive Module

The brain of the cognitive module will be PROPHET. This module is in charge of the generation of an appropriate response (motor parameters) for the user at each moment and in real-time. Fuzzy Logic and other softcomputing techniques will allow handling simultaneously a large amount of uncertain information. A knowledge database will store both the rules self-learned by PROPHET (dynamic knowledge) and the rules pre-established by an expert (static knowledge). The different sources of affective information will be integrated, the natural language (dialogue) will be processed, the commands (inserted either by keyboard or mouse) and the user position will be analyzed… all this in terms of perceptions, in order to achieve similar solutions to the ones taken by human beings. That way, Maxine will be provided with more flexibility, intelligence and ability of autonomous learning.

4.3 Motor Module

Finally, the motor module -completely implemented at present- is in charge of generating the system's outputs and the final animation of the 3D virtual agent. Script-files-based commands containing the facial expression, speech and body parameters will be generated and executed in real-time for achieving the appropriate body animation, facial animation (lip-synch and facial expressions) and emotional voice synthesis (prosody and speech). Note that an inherent feedback exists between the motor module and the cognitive, since the evolution of the animation parameters must be considered by the cognitive module to ensure a temporal consistency in the agent's behaviour.

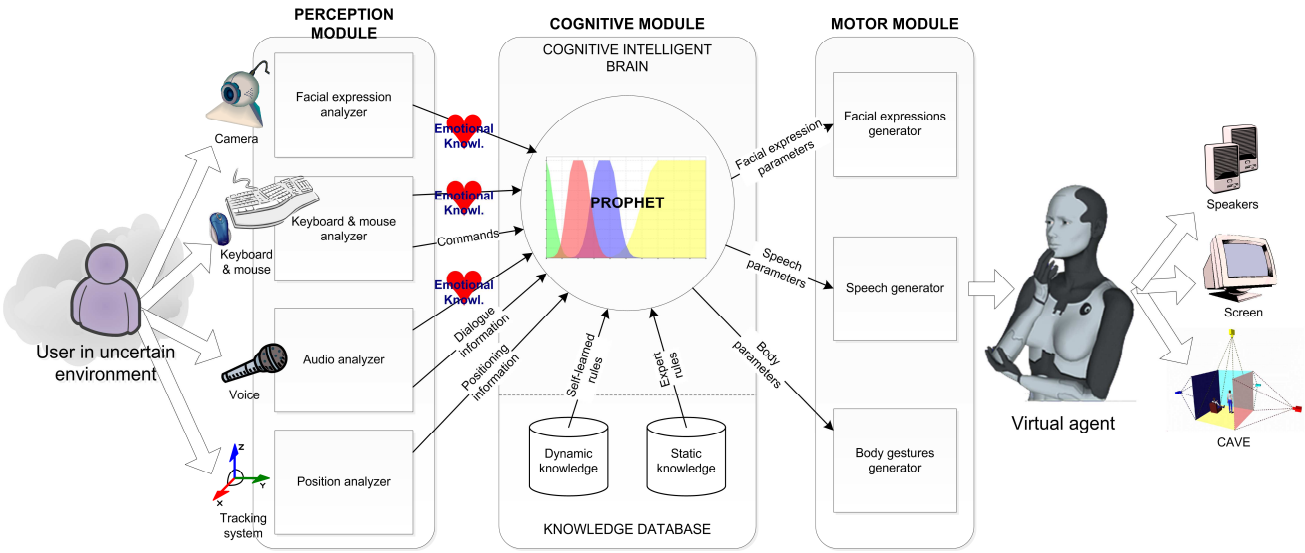

Figure 3. Proposed architecture for the interaction with intelligent affective multimodal virtual agents in uncertain environments.

5. CONCLUDING REMARKS

In this paper, an architecture for the interaction with intelligent affective multimodal virtual agents in uncertain environments has been proposed. The system is based on the integration of Maxine, a powerful animation engine for managing virtual agents and 3D environments, and PROPHET, a cognitive intelligent system. Thanks to the use of Fuzzy Logic, PROPHET will provide Maxine with complex real-time decision-making capabilities based on the analysis and learning of information and perceptions. Other outstanding features of the proposed architecture are its affective capabilities, either for recognizing the user's affective state as for making the agent look emotional, its multimodality and its capability for working in real-time.

However, since PROPHET and Maxine are pre-existing, a set of problems and decisions are expected to be faced when integrating both systems. In particular, the integration of the information coming from the different perception sources will be one of the main questions to solve due to the system's asynchronous nature. Another question that will arise is how to manage the scheduling and interruption of currently playing animations. To date, there has been no need to distribute PROPHET or Maxine since most of the information they manage is stored in memory to make the systems work faster. If the integration requires more persistence power, the possibility of making the different modules run on separate machines exists.

In the future, the system is expected to be used in a wide range of scenarios such as domotic, e-learning or teleassistance.

6. ACKNOWLEDGMENTS

This work has been partly financed by the CETVI project (PAV-100000-2007-307) funded by the Spanish Ministry of Industry, the Grupo de Ingeniería Avanzada (GIA) of the Instituto Tecnológico de Aragón, the Spanish DGICYT (contract NºTIN2007-63025), the Government of Aragon IAF Nº2008/0574 and CyT Nº2008/0486 agreements.

7. REFERENCES

[1] Cassell, J., Sullivan, J., Prevost, S. and Churchill E. 2000. Embodied conversational agents. Cambridge, MIT Press.

[2] Canny, J. 2006. The future of human-computer interaction. ACM Queue, Vol. 4, No. 6 (Jul./Aug. 2006).

[3] Pantic, M. and Bartlett, M.S. 2007. Machine Analysis of Facial Expressions. Face Recognition. K. Delac and M. Grgic, Eds. I-Tech Education and Publishing, Vienna, Austria, 377-416.

[4] Jaume-i-Capó, A., Varona, J. and Perales, F.J. 2006. Interactive applications driven by human gestures. Ibero-American Symposium on Computer Graphics (Santiago de Compostela, Spain, Jul. 2006).

[5] Cowie, R., Douglas-Cowie, E. and Schroeder, M. 2000. ICSA Workshop on Speech and Emotion: a Conceptual Framework for Research (Belfast).

[6] Egges, A., Kshirsagar, S. and Magnenat-Thalmann, N. 2004. Generic personality and emotion simulation for conversational agents. Computer Animation and Virtual Worlds, Vol. 15, No. 1, 1-13.

[7] Magnenat-Thalmann, N. 2003. Creating a Smart Virtual Personality. Knowledge-Based Intelligent Information and Engineering Systems 2003, LNCS, Springer, Heidelberg, 15-16.

[8] Zadeh, L. A. 2003. Computing with Words and Perceptions - A Paradigm. Computing and Decision Analysis and Machine Intelligence, Proceedings of the 2003 International Conference on Machine Learning and Applications, 3-5.

[9] Baldassarri, S., Cerezo, E. and Seron, F.J. 2008. Maxine: a platform for embodied Animated Agents. Computers & Graphics, Vol. 32, No. 4, 430-437.

[10] Hupont, I., Baldassari, S., Cerezo, E. and Del Hoyo, R. 2008. Effective emotional classification combining facial classifiers and user assessment. V Conference on Articulated Motion and Deformable Objects, LNCS, Springer, No. 5098, 431-440.

[11] Del Hoyo, R., Ciprés, D., Prieto, J., Del Barrio, M., Polo, L. and Calahorra, R. 2007. PROPHET: Herramienta para la Toma de Decisiones en Sistemas Complejos. II Simposio de Inteligencia Computacional 2007 (Zaragoza, Spain).

[12] Del-Hoyo, R., Martín-del-Brío, B., Medrano, N. and Fernández-Navajas, J. Computational Intelligence Tools for Next Generation Quality of Service Management. Neurocomputing, 2009, in press.

[13] https://www.hibernate.org/

[14] Lin, C.T., and Lee, C. S. G. 1996. Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems. Upper Saddle River, NJ: Prentice Hall.