Computers & Graphics 32 (2008) 430–437

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Chaos and Graphics

Maxine: A platform for embodied animated agents

Sandra Baldassarri *, Eva Cerezo, Francisco J. Seron

Advanced Computer Graphics Group (GIGA), Computer Science Department, University of Zaragoza, Engineering Research Institute of Aragon (I3A), Zaragoza, Spain

ARTICLE INFO

Article history:
Received 20 December 2007
Received in revised form
18 March 2008
Accepted 7 April 2008

Keywords:
Virtual worlds
Animated characters
Natural interaction
Multimodal interfaces

ABSTRACT

This paper presents a powerful animation engine for developing applications with embodied animated agents called Maxine. The engine, based on open source tools, allows management of scenes and virtual characters, and pays special attention to multimodal and emotional interaction with the user. Virtual actors are endowed with facial expressions, lip-synch, emotional voice, and they can vary their answers depending on their own emotional state and the relationship with the user during conversation. Maxine virtual agents have been used in several applications: a virtual presenter was employed in MaxinePPT, a specific application developed to allow non-programmers to create 3D presentations easily using classical PowerPoint presentations; a virtual character was also used as an interactive interface to communicate with and control a domotic environment; finally, an interactive pedagogical agent was used to simplify and improve the teaching and practice of Computer Graphics subjects.

1. Introduction

Current research in ambient intelligence deals with multimodality and social interaction. In addition, most research on social interfaces is related to the design of embodied conversational agents (ECAs) [1]. ECAs are agents which are visibly present in the interface—sometimes as an animated talking face—may display facial expressions and, when using speech synthesis, may also be endowed with lip synchronization. Sometimes they are also endowed with 3D graphical representation and display complex body movements, etc. Virtual characters equipped with these new features can be used in a wide range of contexts [2,3], including education and learning [4–8], sign language interpretation [9], therapy [10], persuasion [11,12], entertainment [13,14], among others.

In order to achieve a more natural and realistic interaction, virtual agents must be capable of appropriately responding to users with affective feedback [15]. Therefore, within the field of ECAs, our research is focused on interactive virtual agents that support multimodal and emotional interaction in order to establish more effective communication with the user. Special emphasis is placed on capturing the user's emotions through images, as well as on synthesizing the virtual agent's emotions through facial expressions and voice modulation.

In the system we developed, called Maxine, the virtual agent is endowed with the following distinctive features:

- It supports interaction with the user through different channels: text, voice, peripherals (mouse, keyboard, web-cam, ...). This makes the use of the applications generated available to a wide range of users in terms of communication ability, age, etc.
- It gathers additional information on the user and his or her surroundings, i.e., in relation to the noise level in the room, the position of the user (in order to establish visual contact), image-based estimates of the user's emotional state, etc.
- It supports voice communication with the user in natural language, in Spanish.
- It has its own emotional state, which may vary depending on the relationship with the user. This emotional state modulates the presenter's facial expressions, its answers and its voice.

The paper is organized as follows. An overview of the animation engine, Maxine, is described in Section 2. Section 3 presents the system's input and its management by the sensory/perception modules, while Section 4 presents the agent's possible reactions through the deliberative and generative modules. Section 5 presents the motor module and the system's outputs. Section 6 deals with the presentation of different applications developed with the engine. Finally, Section 7 offers some conclusions and comments on future work.

2. Overview of Maxine: the animation engine

Maxine is a script-directed engine for the management and visualization of 3D virtual worlds. Maxine can load models, animations, textures, sounds, etc. into the virtual representation in real-time whenever they are needed. Despite being a very generic engine, it is oriented towards working with characters in virtual scenarios. One of its advantages is that it allows specification of contents via an editor; however, it also makes direct communication with low-level tools possible. It has been written in C++ and employs a set of open source libraries.

* Corresponding author. Tel.: +34 976 762357; fax: +34 976 761914.
E-mail address: sandra@unizar.es (S. Baldassarri).
URL: http://giga.cps.unizar.es (S. Baldassarri).

0097-8493/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2008.04.006

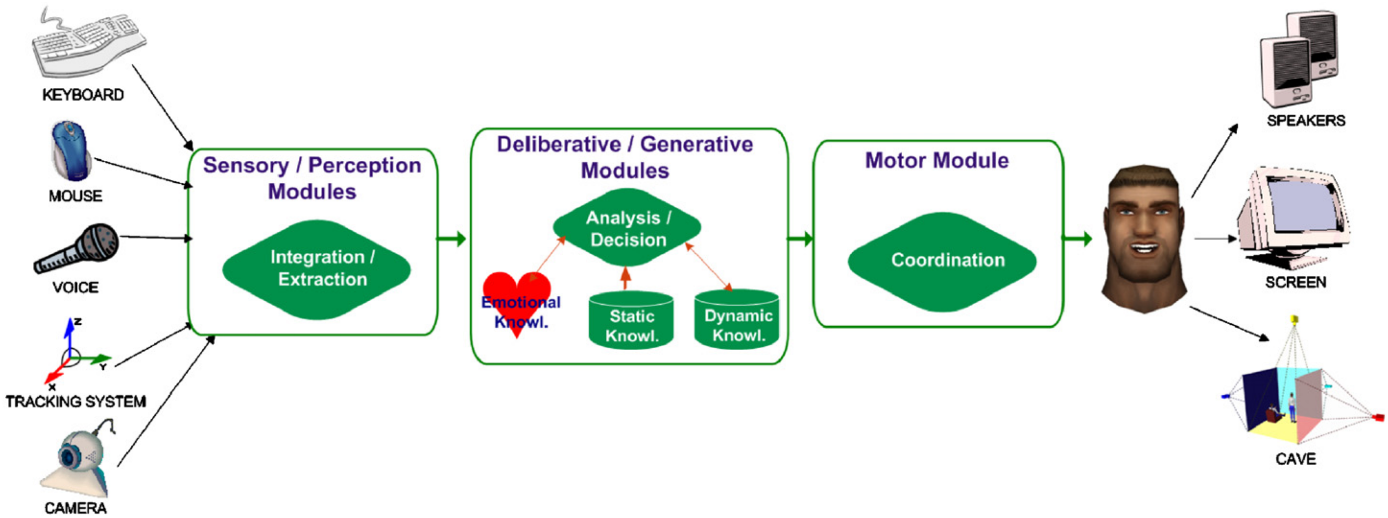

Fig. 1. Maxine's architecture.

Table 1
Libraries used in Maxine

Function	Library	Comment
Scripting	Lua, LuaBind [16]	Real-time creation and manipulation of objects
3D management and representation	OpenSceneGraph [17]	Scenegraph management
	Producer [17]	3D visualization and camera management
Animation	CAL3D [18]	Local animation of virtual characters
	OpenSteer [19]	Steering behaviors
Audio	OpenAL + + [20]	3D sound
Speech synthesis	SAPI [21]	Voice synthesis (free but not open source)
	Loquendo [22]	Voice synthesis (commercial)
General	OpenThreads [23]	Basic Thread manipulation tools
	STLPort [24]	Multiplatform C + + Standard Library

The engine manages scene graphs that can be built in real-time, thus dynamically creating and manipulating its elements 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.

The overall architecture of our system is shown in Fig. 1. In the sections which follow, all of the modules that comprise Maxine are explained in detail: the sensory/perception modules, which process the system's inputs, the deliberative/generative modules, which manage the appropriate reactions according to the inputs, and the motor module, which generates the final outputs of the system. In this paper, we will be focusing on system inputs and outputs.

The driving philosophy behind the generation of our system was the use of Open source libraries to the largest extent possible in order to concentrate our efforts on truly interesting and challenging problems, particularly those related to emotional and multimodal user interaction. Table 1 displays the main libraries used, as well as specifying their functionality and briefly describing their purpose within Maxine.

3. System's inputs: the sensory/perception modules

During the development of the system, special attention was paid to creating multimodal user interaction via text, voice, image and movement. This broadens the spectrum of potential users of the system by allowing interaction with disabled users (for example, the hearing-impaired or paraplegics) and people of different ages and with different levels of education (people with or without computer knowledge). Stress was also placed on collecting the largest possible amount of information on the user through body language or facial expression without requiring him or her to enter data, with the ultimate aim of enhancing interaction and establishing emotional communication between the user and the virtual character. The sensory/perception modules integrate all the information from the different inputs into the system.

3.1. *Interaction via consola/mouse*

Advanced users can fully control the scene by resorting to the scripting languages used: Lua and Luabind (see libraries in Table 1), which allow control of execution time as well as the extension and modification of the behavior of its components. An example of a Lua script file for the initialization of an actor follows:

```
- Creating the object "actor"
act = actor();

-- File paths
DIR = './data/facemaker/';

-- Loading the actor files: skeleton, mesh, materials
act:loadSkeleton(DIR .. "facemaker.csf");
act:loadMesh(DIR .. "facemaker.CMF");
act:loadMaterial(DIR .. "facemaker.xrf");
act:loadTextures(DIR);

--Loading the animation files
act:loadAnimation("e_closed", DIR .. "e_closed.CAF");

-- Loading the animations of visemes
act:loadVisemeAnim("vis_silence", DIR .. "default.CAF");
...
act:loadVisemeAnim("vis_s",  DIR .. "s.CAF");
act:setVisemeAnim ("vis_sh", "vis_s");
...

-- Creating the actor from the loaded files
act:create();

-- Setting actor's initial posture
act:startPose("stand", 0.5, 0.5, 1.0);

-- Setting actor's periodic animation (actor blink)
act:startPeriodic("blink",0.05,0.05,1,3);
act:setPeriodicVar("blink",2);
```

However, non-programmer users can also associate the execution of a given command to the pressing of certain key or mouse click. Due to the power of some of the functions available for the elements and, in particular, of the scripting language used, there is a wide range of options. Besides managing commands through a Lua script file, each class also has a set of commands that can be controlled via console. Some commands control the basic actions of the animation engine and others allow one to modify the virtual environment, to load objects and actors, and also to execute animations.

3.2. Audio speech recognition (ASR)

The user formulates any order, question or sentence that might be used in conversation in natural language. The sound or audio generated by the user is picked up by the microphone and the sound card, as seen in Fig. 2.

One of the main requisites of our system is that it must be able to *understand* and speak Spanish. This constraint prevented us from using the existing open source libraries, all of which are in English.

Therefore, in order to obtain a text chain from the words delivered in Spanish by the user, a voice recognition engine was built using the commercial Loquendo ASR software [22]. However, certain problems specific to Spanish had to be solved during the development of the recognizer: specifically, Loquendo ASR cannot distinguish between words with or without 'h' (a letter which is not pronounced in Spanish), 'b' or 'v', or with 'y' or 'll' (these letter pairs apply to single phonemes). Neither is it of any use when it comes to detecting the difference between words that are spelt the same but pronounced differently, as would be the case with the verb form 'está' and the pronoun 'ésta', the difference between which is marked by an accent. This modification solves problems without reducing the efficiency of the system. However, this must be taken into account when writing the words of the Java Speech Grammar Format (JSGF) grammar and generating Artificial Intelligence Markup Language (AIML) [25] files, thanks to both of which, depending on the context of the conversation, we will be able to know which word is being referred to—as well as its meaning—in order to come up with a proper answer.

3.3. Image interaction

A webcam takes pictures of the user's face. The purpose of these pictures is to gather additional information on the user and his or her emotional state in particular. Fig. 3 shows an overview of the stages involved in the image interaction process.

We used Ekman's emotional classification [26], which distinguishes between six basic emotions: happiness, sadness, anger, fear, surprise, disgust, plus a neutral category. The emotional classification system developed is based on the detection of ten points and five distances. The face classification implemented is an effective method based on the theory of evidence [27]. The classification results derived from the five distances extracted from the user image (as well as on the basis of a series of thresholds taken from the analysis of a broad image database) are acceptable, and range from a 63% success rate in recognizing surprise to a rate of 100% in the case of happiness or disgust. In the case of the other three emotions, values approximated 90%.

3.4. Other possible inputs: adaptation to the CAVE-like facility

The system is being used for the presentation of previous projects in different languages in our CAVE-like system [28] (see Fig. 4).

New modes of input/interaction have been set up in this environment:

- *Positioning system*: The positioning system (mixed inertial-ultrasound) calculates the position and orientation of the user's head and hand. This information enables the virtual presenter to 'look at' or 'speak to' the user, or the user to point at and select objects in the presentation. The 'main' user can also use a wand tracker (special mouse) equipped with four programmable buttons for the purpose of reproducing the interaction via keyboard/mouse on console. Special care has been taken to discard incorrect information (due to drifts or incorrect measurements) and also to detect any appreciable changes in the position or orientation of the user's head or hand, as well as to identify gestures such as nodding or shaking the head, raising the hand to signal, point, etc.

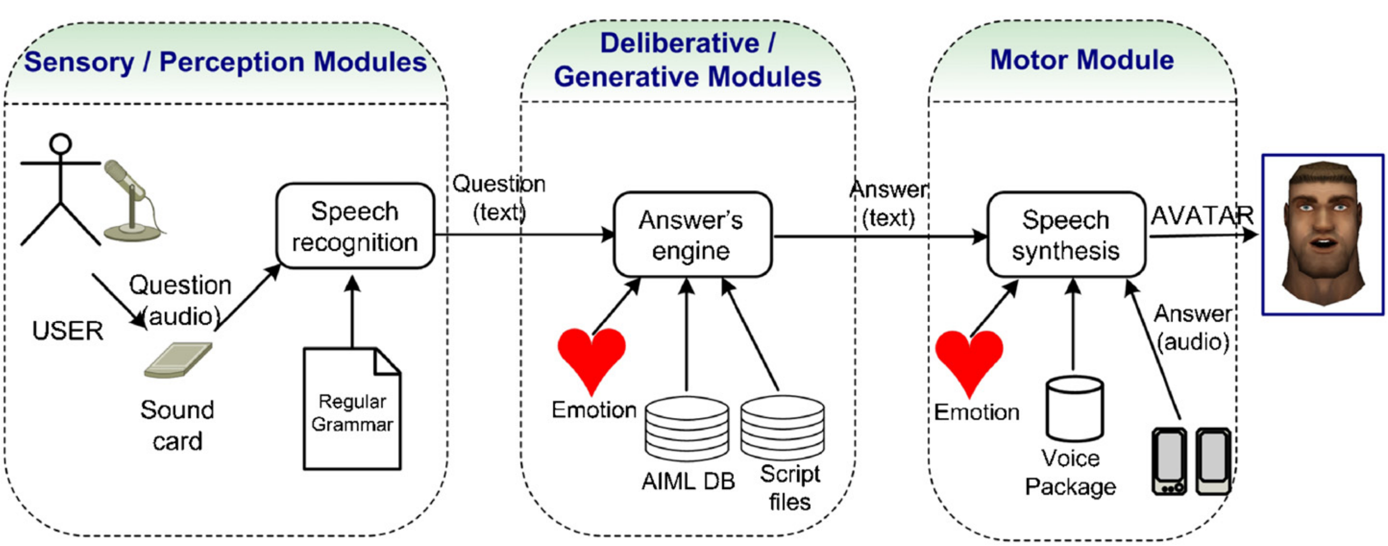

Fig. 2. Stages of the user–avatar voice communication process.

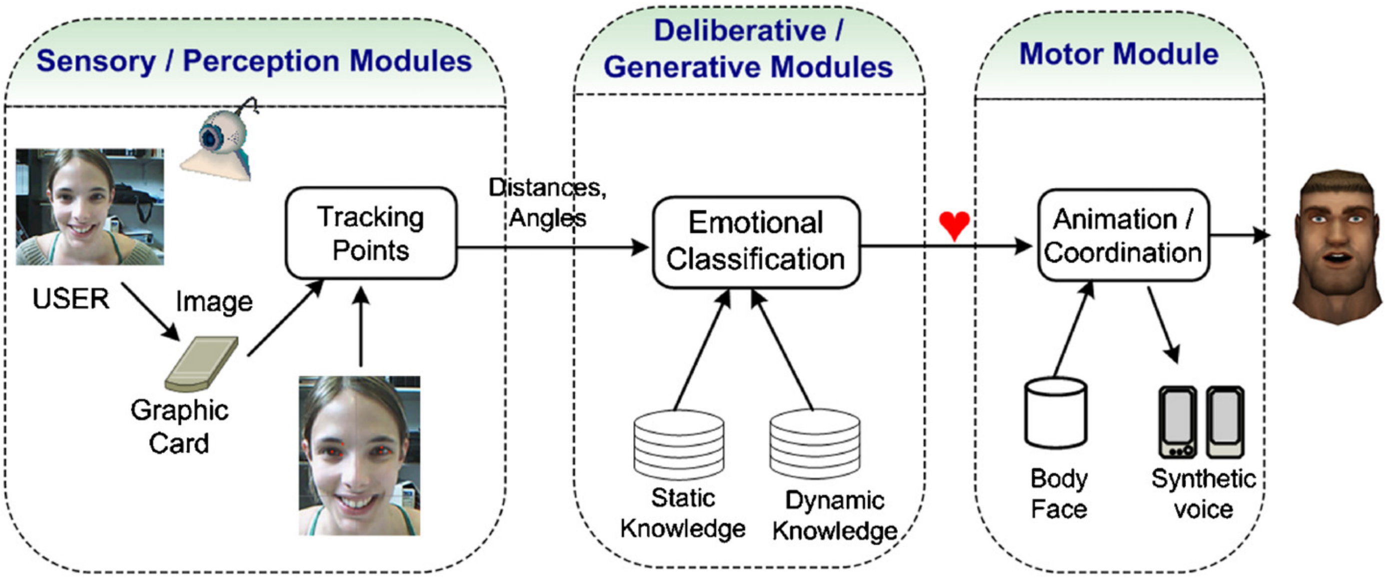

Fig. 3. Stages of the user–avatar image interaction process.

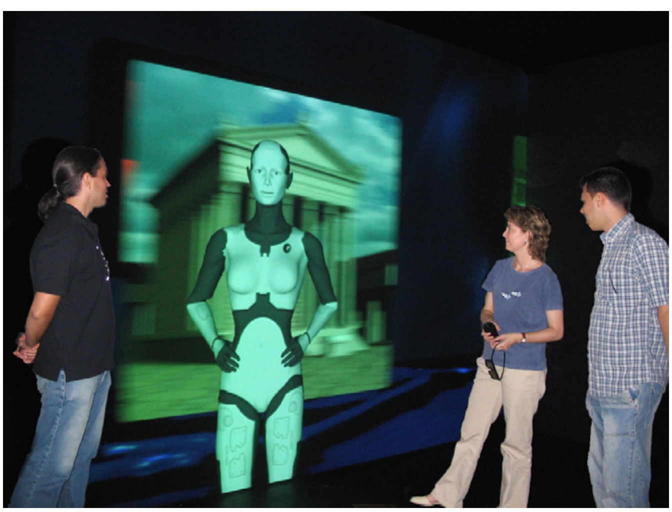

Fig. 4. A Maxine's presentation in our immersive environment.

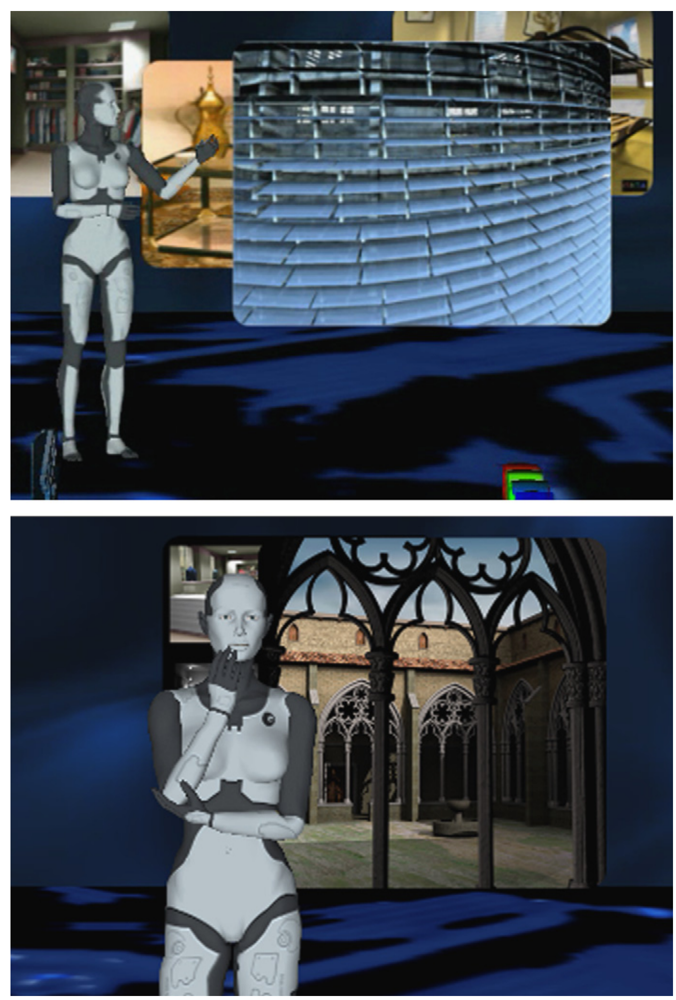

Fig. 5. Automatic position changes of the virtual character during a presentation.

- *Background microphones*: These microphones monitor the sound level (and consequently, the level of attention) in the room.

4. Agent reactions: the deliberative and generative modules

After extracting the input information, the system has to manage the appropriate reactions. Two kinds of actions must be distinguished:

- *Purely reactive*: For example, should the user key something in, the virtual presenter will interrupt the presentation; should the user change his or her position, the presenter's look/orientation will change; if a lot of background noise is detected, it will request silence, etc. Such reactions are managed in the generative module.
- *Deliberative*: The choice of the virtual character's reaction calls for a more complex analysis. This analysis is performed in the deliberative module, which elicits an answer through the user's voice interaction or estimates the user's emotional state through image interaction.

These modules basically generate an answer for the user's questions in text mode (see Fig. 2), and they are based on the recognition of patterns associated to fixed answers (static knowledge). These answers, however, vary according to the virtual character's emotional state, the emotional state of the user, or may undergo random variations so that the user does not get the impression of repetition should the conversation goes on for a long time (dynamic knowledge). The development of this part of the system is based on chatbot technology under GNU GPL licenses: ALICE [29] and CyN [30]. However, since CyN is only designed to hold conversations in English, we had to modify the code in order to enable it to support dialogues in Spanish. The main differences lie in the code's capacity to work with written accents, diaeresis and the 'ñ' character, as well as enabling the use of opening question and exclamation marks.

The knowledge possessed by the virtual character is specified in AIML [25]. AIML's power lies in three basic aspects:

- AIML syntax enables the semantic content of a question to be extracted easily, so that the appropriate answer can be quickly supplied.
- The use of labels to combine answers lends greater variety to the answers and increases the number of questions that can be answered.
- The use of recursion allows answers to be provided for inputs for which, in theory, there is no direct answer.

The AIML interpreter has been modified to include commands or calls to script files within the AIML category, so that these commands are executed and their results are returned to the user as part of the answer. This makes it possible, for example, to check the system time, log on to a website to check what the weather is like, etc.

5. System's outputs: the motor module

5.1. Facial and body animation

The skeletal animation technique is used for both facial and body animation. The animations which the system works with are derived from two sources: animations from motion capture (the group has access to an optical system) and animations generated by means of commercial software. In both cases, these animations are included in the scene and can be used at any moment by calling them via script file.

Similar to general animations, which may be cyclic or not, and pose animations, which are postures maintained for a certain period of time, the system automatically inserts background and periodic animations. The former are secondary animations that gradually modify the character's posture and certain parts of its body (see Fig. 5). Perlin's algorithms, based on coherent noise, were used for their implementation. Background and periodic animations are animations automatically executed at certain established intervals. One of the most typical examples of this type of animation is their use to get actors to automatically blink or to register their breathing motions. However, strict movement periodicity makes the actor

Viseme ID.	Animation ID	Sound (English)	Sound (Spanish)
0	"VIS_SILENCE"	<silence>	<silencio>
1	"VIS_AE"	ae, ax, ah	-
2	"VIS_AA"	aa	a
3	"VIS_AO"	ao	-
4	"VIS_EY"	ey, eh, uh	e
5	"VIS_ER"	er	-
6	"VIS_IY"	iy, ih, ix	i
7	"VIS_W"	w, uw	u
8	"VIS_OW"	ow	o
9	"VIS_AW"	aw	-
10	"VIS_OY"	oy	-
11	"VIS_AY"	ay	-
12	"VIS_H"	h	j
13	"VIS_R"	r	r, rr
14	"VIS_L"	l	l
15	"VIS_S"	s, z	s
16	"VIS_SH"	sh, ch, jh, zh	ch
17	"VIS_TH"	th, dh	z
18	"VIS_F"	f, v	f
19	"VIS_D"	d, t, n	d, n, ñ, t
20	"VIS_K"	k, g, ng	c, g, k, q
21	"VIS_P"	p, b, m	b, m, p, v

Fig. 7. Viseme/sound correspondence in English and Spanish.

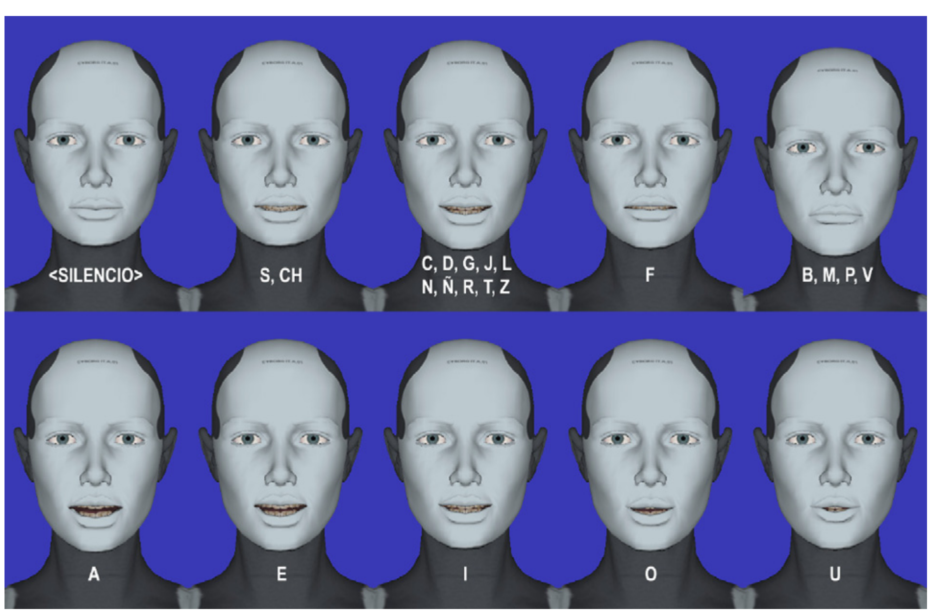

Fig. 6. Visemes designed for lip-synch in Spanish.

unnatural, so random variation is included in the execution frequency of these animations. For facial animation, we work with the six basic expressions defined by Ekman [26] and the nomenclature followed is that of the VHML standard [31].

In all these cases, animation blending was achieved with the help of the Cal3D library [18].

5.2. Expression coordination: lip-synch

A lip-synch module specially developed for Spanish language was implemented. In order to solve the problem of labial synchronization, this module uses SAPI5 shareware software [21] to gather information about the visemes (visual phonemes) that occur when delivering the phrase to be synthesized. Visemes were used to model the actor's mouth movements. Visemes used for labial synchronization are shown in Fig. 6, while Fig. 7 shows the correlation between viseme identifications, animation identifications and the sounds related to them. As may be appreciated, some of the visemes are used to represent more than one phoneme.

In order to avoid problems arising from maintaining synchronism, a special procedure was implemented in two particular cases: when the character's speech speed is too high and when system performance is not high enough. If one considers the real case of human beings speaking very quickly, it is impossible to read their lips because there are mouth positions (visemes) that last for a very short time or do not take place at all. The improvement detects visemes that do not last long enough to be reproduced because the frequency with which they are checked is insufficient. These visemes are discarded, thus improving synchronization between the animation and the audio. Fig. 8 graphically represents some examples of this situation through the synthesis of the sentence "I can speak". The diagram shows the range of possible cases. The viseme that applies to 's' is preceded by a 'silence' viseme that will not be heard (and will not be executed) because it is too close. The second case arises when the update frequency decreases from 20 to 10 (these values are a little low for the human eye but suited to the graphic representation in the example), which means that more visemes are ignored.

5.3. Speech synthesis

In the system we developed, speech synthesis is performed with the Spanish voice packages offered by Loquendo TTS [22], although SAPI5 [21] is used for gathering information regarding the visemes.

In order to avoid the voice sounding artificial, it was equipped with an emotional component. Voice emotions fall into Ekman's six universal categories: joy, sadness, anger, surprise, disgust and fear, plus the neutral category. SAPI5 enables us to modify tone, frequency scale, volume and speed, which is why we used it as a basis. In order to represent each emotion, fixed values were assigned to the parameters that enable the relevant emotion to be evoked. These emotional parameters were configured in reference to several studies [32–34]; the process followed to find the values at which these parameters must be fixed was implemented through voice assessment by the users according to three assessment paradigms: Forced Choice, Free Choice and Modified Free Choice [35].

6. Maxine's applications

The system previously described has been used in the different applications, such as:

- Virtual humans for PowerPoint-like presentations;
- Control of a domotic environment;
- Interactive pedagogical agent for teaching Computer Graphics.

Details regarding each application are set forth below.

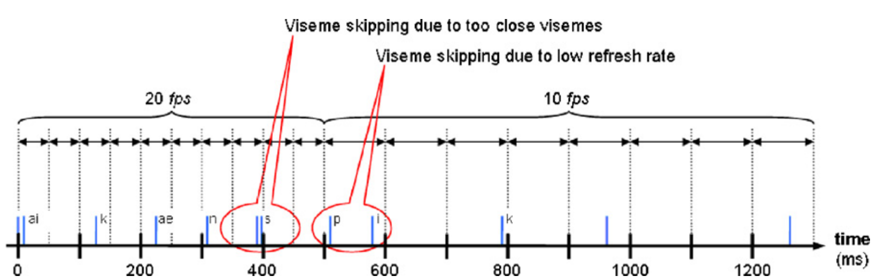

Fig. 8. Elimination of visemes in order to maintain synchronism.

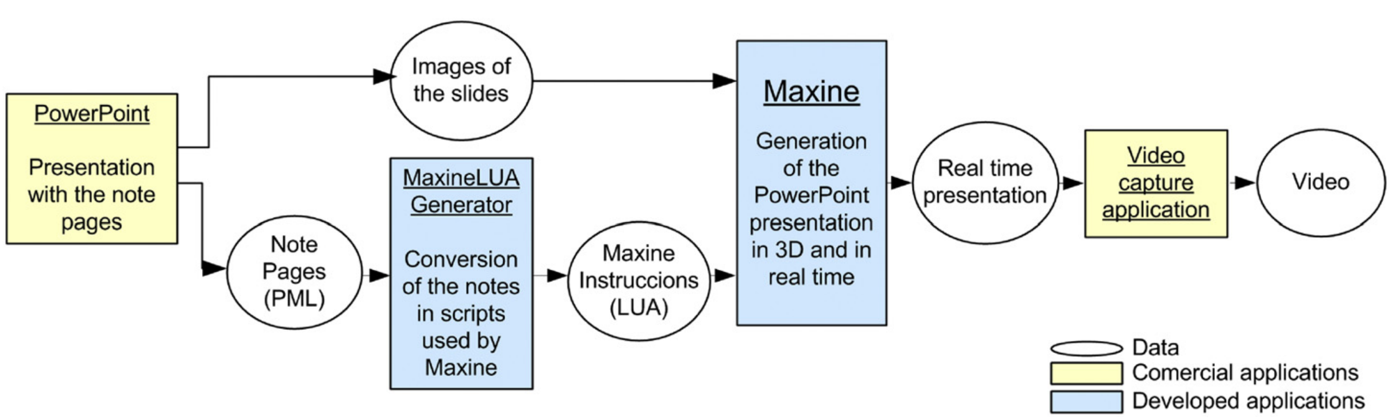

Fig. 9. Overview of the process for generating a 3D presentation.

6.1. MaxinePPT: virtual humans for PowerPoint-like presentations

This application allows the presentation of PowerPoint information through a like-life character in a graphic display. This kind of presenter has proved to be particularly helpful when the same presentation has to be repeated several times or delivered in a different language (for example, in English by a non-fluent English speaker).

The most important features of the MaxinePPT application [36] are:

- It can create and perform virtual presentations in 3D virtual scenarios enriched with virtual actors and additional information such as videos, images, etc. on the basis of a classical PowerPoint file.
- All the aspects of the virtual presentation are controlled by an XML-type language called PML (Presentation Markup Language). PML instructions are added to the page notes of the PowerPoint slides in order to determine, for example, the text to be delivered by the avatar.

Once the presentation has been created, user intervention is unnecessary. The presentation is automatically and verbally delivered, either in Spanish or English, by using a virtual character specially created by the user, or by the default avatar included in the system. An overview of the process involved in the creation of a presentation is offered in Fig. 9.

Fig. 10 shows some screenshots of a virtual presentation.

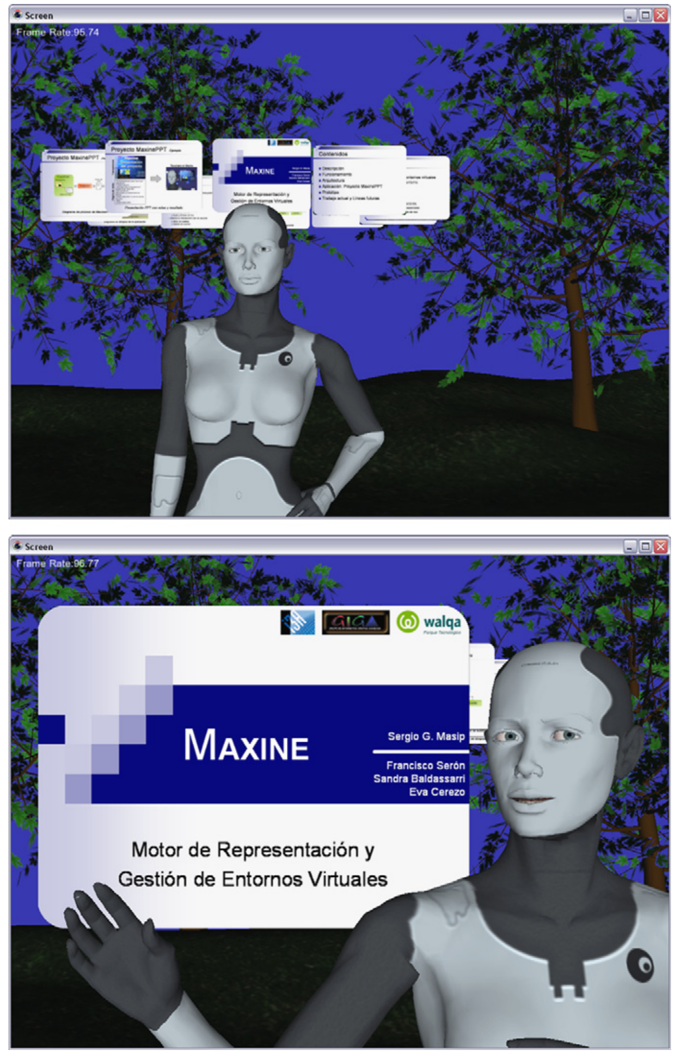

Fig. 10. Screenshots from a virtual presentation.

6.2. Control of a domotic environment

A virtual avatar called Max was created and used as an interactive interface for the accessing and remote control of an intelligent room [37]. The user can communicate with Max through natural language in Spanish and ask him to perform different tasks within the domotic environment, as well as to execute orders relating to the different devices present in the intelligent room (see Fig. 11).

In addition, in order to take full advantage of the emotional communication possibilities of the Maxine system, Max incorporates lip synchronization, while voice modulation takes the emotional state of the virtual actor into account. Given an order or entry from the user, the answer engine will select the appropriate answer. These can vary dynamically according to the avatar's emotional state and the type of conversation taking place with the user.

Fig. 11. User interacting with the domotic environment through Max, the virtual actor.

6.3. Interactive pedagogical agent for teaching Computer Graphics

Maxine was also used to develop a learning platform for simplifying and improving the teaching and practice of Computer Graphics subjects [38]. By carefully orchestrating the different possibilities offered by our animation engine—facial expression, body placement, arm movements, hand gestures and conversational signals—the embodied pedagogical agents may encourage students to pay more attention to their own progress, convey enthusiasm for the subject and simply make learning more fun.

The interactive pedagogical agent can help students to expound certain specific topics by acting as a virtual teacher, and help them understand difficult topics involved in CG subjects by allowing interaction with and handling of a 3D environment (when a relevant concept is being explained, the teacher or student may manipulate the environment by modifying different parameters, points of view, lights, etc., and interactively visualizing the results).

The use of Maxine has proved to be a very useful tool for improving the comprehension of the most complex concepts involved in our courses, and according to the students who have worked with the system, the results are promising.

7. Conclusions and future work

A powerful animation engine for managing virtual environments and virtual actors called Maxine is presented. The system allows the development of new applications in which interaction is based on virtual agents that support multimodal and emotional interaction.

The most outstanding features of our system are that it supports real-time interaction with the user through different channels (text, voice, mouse/keyboard, image); the voice interface enables communication in natural language and in Spanish; discourse is reinforced with referential acts (pointing gestures and/or looking at objects) and interaction (the virtual actor follows the user's movements with its eyes and body orientation, allows the user to interrupt, etc.); the system is capable of adding character movements that increase the virtual actor's naturalness and credibility (blinking, position and expression changes, etc). Communication between the user and the 3D character is mediated by the virtual character's facial expressions and its 'emotional voice', since the answers to questions and voice modulation are adapted to the emotional state.

The potential of the animation engine is presented through the use of different applications that use virtual humans in order to enhance interaction: through virtual presenters in PowerPoint-like presentations, through an interface for controlling a domotic environment and through virtual teachers for improving the theory and practice of a Computer Graphics course.

However, other lines of research, most of which focus on enriching the interaction between the virtual actor and the audience, remain open:

- Inclusion not only of emotion but also of personality models for the virtual character.
- Enrichment of virtual character behavior by moving from reactive to cognitive schemes.

At the present time, we are working on the validation of Maxine's system, its virtual characters and interaction with the users.

Acknowledgments

This work has been partially financed by the Spanish 'Dirección General de Investigación', Contract Number TIN2007-63025 and by the Regional Government of Aragon through the WALQA agreement Ref. 2004/04/86.

The authors thank David Anaya for his collaboration in this work.

References

[1] Casell J, Sullivan J, Prevost S, Churchill E, editors. Embodied conversational agents. Cambridge: MIT Press; 2000.
[2] Mignonneau L, Sommerer C. Designing emotional, methaforic, natural and intuitive interfaces for interactive art, edutainment and mobile communications. Computers & Graphics 2005;29:837–51.
[3] Cassell J. Embodied conversational agents: representation and intelligence in user interface. AI Magazine 2001;22(3):67–83.
[4] Lester JC, Towns SG, Fitzgerald PJ. Achieving affective impact: visual emotive communication in lifelike pedagogial agents. International Journal of Artificial Intelligence in Education 1999;10:278–91.
[5] Garcia A, Lamsfus C. An eLearning platform to support vocational training centers on digital security training with virtual tutors and graphical spatial metaphores. In: Proceedings of the IADAT-e2005, international conference on education; 2005. p. 117–21.
[6] Boff E, Reategui E, Viccari R. An affecrive agent-based virtual character for learning environments. In: 12th international conference on artificial intelligence in education; 2005. p. 1–8.
[7] Graesser A, Chipman P, Haynes B, Olney A. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education 2005; 48(4):612–8.
[8] Ortiz A, Aizpurua I, Posada J. Some techniques for avatar support of digital storytelling systems. In: Proceedings of the TIDSE03: technologies for interactive digital storytelling and entertainment; 2003. p. 322–7.
[9] Rieger T. Avatar gestures. Journal of Winter School of Computer Graphics 2003;11(2):379–86.
[10] Marsella SC, Jhonson WL, LaBore C. Interactive pedagogial drama. In: Proceedings of the 4th international conference on autonomous agents; 2000. p. 301–8.
[11] de Rosi F, Pelachaud C, Poggi I, Carofiglio V, de Carolis B. From Greta's mind to her face: modelling the dynamics of affective status in a conversational embodied agent. International Journal of Human–Computer Studies 2003:81–118 (Special issue on applications of affective computing in HCI).
[12] Berry D, Butler L, de Rosis F. Evaluating a realistic agent in an advicedgiving task. International Journal in Human–Computer Studies 2005;63: 304–27.
[13] El-Nasr MS, Ioerger T, Yen J. PETEEI: a PET with evolving emotional intelligence. In: Autonomous agent'99; 1999. p. 9–15.
[14] Yuan X, Chee S. Design and evaluation of Elva: an embodied tour guide in an interactive virtual art gallery. Computer Animation and Virtual Worlds 2005;16:109–19.
[15] Cowie R, Douglas-Cowie E, Schroder M, editors. Proceedings of the ISCA workshop on speech and emotion: a conceptual framework for research, Belfast, 2000.
[16] LuaBind ⟨http://www.rasterbar.com/products/luabind.html⟩.
[17] Open Scene Graph ⟨http://www.openscenegraph.org/⟩.
[18] Character Animation Library ⟨http://cal3d.sourceforge.net/⟩.
[19] OpenSteer ⟨http://opensteer.sourceforge.net/⟩.
[20] OpenAL++ ⟨http://alpp.sourceforge.net/⟩.
[21] Long B. Speech synthesis and speech recognition using SAPI 5.1: In: European Borland conference 2002, 2002 ⟨http://www.blong.com/Conferences/DCon2002/Speech/SAPI51/SAPI51.htm⟩.
[22] Loquendo Audio Speech Recognition ⟨http://www.loquendo.com/⟩.
[23] OpenThreads ⟨http://openthreads.sourceforge.net/⟩.
[24] STLPort ⟨http://sourceforge.net/projects/stlport⟩.
[25] Artificial intelligence Markup Language. AIML Version 1.0.1 ⟨http://www.alicebot.org/TR/2001/WD-aiml/⟩.
[26] Ekman P. Facial expression, the handbook of cognition and emotion. New York: Wiley; 1999.
[27] Cerezo E, Hupont I. Emotional facial expression clasiffication for multimodal user interfaces. In: Lecture notes in computers science, vol. 4069. Berlin: Springer; 2006. p. 405–13.
[28] Seron FJ, Gutierrez D, Magallón JA, Sobreviela EJ, Gutierrez JA. A CAVE-like environment as a tool for full-size train design, Virtual reality. vol. 7. Berlin: Springer, 2004. p. 82–93.
[29] Artificial Intelligence Foundation ⟨http://www.alicebot.org/⟩.
[30] CyN Project ⟨http://www.daxtron.com/123start.htm?Cyn⟩.
[31] Virtual Human Markup Language. VHML Working Draft v0.3 ⟨http://www.vhml.org⟩.
[32] Boula de Mareuil P, Celerier P, Toen J. Generation of emotions by a morphing technique in English, French, and Spanish. In: Proceedings of the speech prosody 2002; 2002. p. 187–90.
[33] Francisco V, Gervas P, Hervas R. Expression of emotions in the synthesis of voice in contexts narrative. In: Proceedings of the UCAmI 2005; 2005. p. 353–60.
[34] Iriondo I, Guaus R, Rodriguez A, Lazaro P, Montoya N, Blanco J, et al. Validation of an acoustical modelling of emotional expression in Spanish, using speech synthesis techniques. In: Proceedings of the ICSA 2000; 2000. p. 161–6.
[35] Baldassarri S, Cerezo E, Seron F. Interaccion emocional con actores virtuales a traves de lenguaje natural. In: Proceedings of the VIII Congreso Internacional de Interaccion Persona-Ordenador Interacción 2007; 2007. p. 343–52 (in Spanish).
[36] Seron F, Baldassarri S, Cerezo E. Maxineppt: using 3D virtual characters for natural interaction. In: Proceedings of the 2nd international workxhop on ubiquitous computing & ambient intelligence; 2006. p. 241–50.
[37] Haya P, Montoro G, Alaman X. A prototype of a context-based architecture for intelligent home environments. In: Lecture notes in computers science, vol. 3290. Berlin: Springer, 2004. p. 477–91.
[38] Cerezo E, Baldassarri S, Seron F. The use of interactive animated agents for teaching. In: Proceedings of the ICIE 2007: 3rd international conference on interdisciplinary education, 2007.