MaxinePPT: Using 3D Virtual Characters for Natural Interaction

Francisco Serón, Sandra Baldassarri, Eva Cerezo

Advance Computer Graphics Group (GIGA), Computer Science Department
University of Zaragoza, Spain
Engineering Research Institute of Aragon, Zaragoza (I3A), Spain
{seron, sandra, ecerezo}@unizar.es
http://giga.cps.unizar.es/

Abstract. New ways of interaction are arising from the Ambient Intelligence areas. Multimodal interaction with virtual characters is changing the way we communicate and interact with applications and other users. This paper describes an animation engine developed with open source tools and based in the use of scripts that allows the management of scenes and the work with virtual actors. This animation engine is used for the development of an application for the generation of PowerPoint-like presentations enriched with a 3D virtual character. The presentation allows to shows the content of the slides and uses the virtual avatar to communicate with the audience using its voice, its expressions and lip-synch. At the end of the presentation, at the time of questions, there can be interaction between the audience and the user via the virtual avatar.

1 Introduction: Virtual Presenters for Ambient Intelligent Applications

Ambient intelligence emphasises on greater user-friendliness, more efficient services support, user-empowerment, and support for human interactions. In this vision, in a near future, people will be surrounded by intelligent and intuitive interfaces embedded in everyday objects around us and an environment recognising and responding to the presence of individuals in an invisible way. Ambient Intelligence is built on ubiquitous computing plus social and intelligent user interfaces. The intelligent user interfaces allow social interaction and communication with the environment in a natural and personalised way.

Current research into ambient intelligence deals with multimodality and social interaction [1, 2]. Moreover, most of the research on social interfaces is related to the design of embodied conversational agents (ECAs) [3]. Embodied agents are agents that are visible in the interface as animated cartoon characters or animated objects resembling human beings. Sometimes they just consist of an animated talking face, displaying facial expressions and, when using speech synthesis, having lip synchronization. Or sometimes they have 3D graphical representation, with complex corporal movements, etc…

In this context, virtual presenters are not new. Several examples of presentations by life-like agents have been reported in the existing literature: the virtual weather reporter of Noma and Badler [4], the TV reporter of Thalmann and Kalra [5], the André et al presenter from the Personalized Plan-based Presenter (PPP) project [6], the 3D avatar of Baus et al [7] or more recently, the work on virtual presenters for virtual meetings of Nijholt et al [8], directly related with the system presented here. All such agents provide a promising option for interface development as they draw on communication and interaction styles humans are well familiar with [9].

Even though we focus here on virtual embodied agents for presentations, the technology and the models that are being developed and designed can find their way in all kinds of applications of smart environments, including tour guides for a visit to a museum [10, 11] or domotic applications [12].

The rest of the paper is organized as follows: Section 2 introduces the developed animation engine, with its basic capabilities and architecture; A detailed description of MAXINEPPT, our presentation system, is given in Section 3; Section 4 provides some images of the use of the system, and finally, conclusions and future work are discussed in Section 5.

2   Behind the stage: our open-source animation engine

Maxine is a script-directed engine for the management and visualization of 3D virtual worlds, written in C++ and based in a set of open source libraries.

In Maxine it is possible to load in real-time models, animations, textures, sounds, etc. as they are needed in the virtual representation. Even though it is a very generic engine, it has been oriented to the work with characters in virtual scenarios. The specification of contents is done through an editor, but the direct communication with the low-level tools is also possible.

The engine manages scene graphs that are built in real-time, dynamically creating and manipulating their elements by means of a simple command interface. These commands are executed via script-files when initiating the application, during execution, or can be introduced through the text console at any moment.

2.1   Basic features

Maxine is an engine for the creation, organization, manipulation and animation of the different elements that constitute a virtual scenario, through a scene graph. The following elements can be represented in a scene graph:

- Images and texts, that can be shown and positioned as rectangles oriented in space. The main graphic formats (bmp, gif, jpeg, pic, png, rgb, tga, tiff) are allowed and the alpha-channel is supported.
- Simple geometric primitives but also complex geometric models. Maxine imports the most popular formats (3DS, flt, lwo, md2, obj, osg), allowing to create complex virtual scenarios with high level of detail.

- Simple lights.
- 3D and ambient sound.
- Animated characters. They use the format of the Cal3D animation library [13], but they can be generated with a commercial 3D modelling and animation toolkit. Different types of animations are available, including secondary animations to increase the characters' expressivity and realism.
- Animated Actors. They are a specialization of the previous ones and are provided with voice synthesis and facial animation with lip-synch, following the VHML standard [14] and the MPEG4 specification [15] for facial animation.
- Synthetic voices. The voices are provided via text, either by the user or by file, and may be associated to an element. Several tags are used to configure and control the audio characteristics like voice selection, volume control, speech speed, insertion of pauses, tone, word emphasis, pronunciation specification, etc.

The engine also manages several auxiliary elements like: cameras, group of elements, animators (for animating group of elements), …

The developed system allows the inclusion of animations provided by capture systems that, in our particular case, come from a VICON 460 optical motion capture system [16]. The exportation of the VICON system is done through the Character Studio Marker (CSM) format that contains the recorded positions of optical markers attached to the actor.

The developed system allows the inclusion of animations coming from capture systems. In our particular case, the animations come from a VICON 460 optical motion capture system [16], exported in the Character Studio Marker (CSM) format.
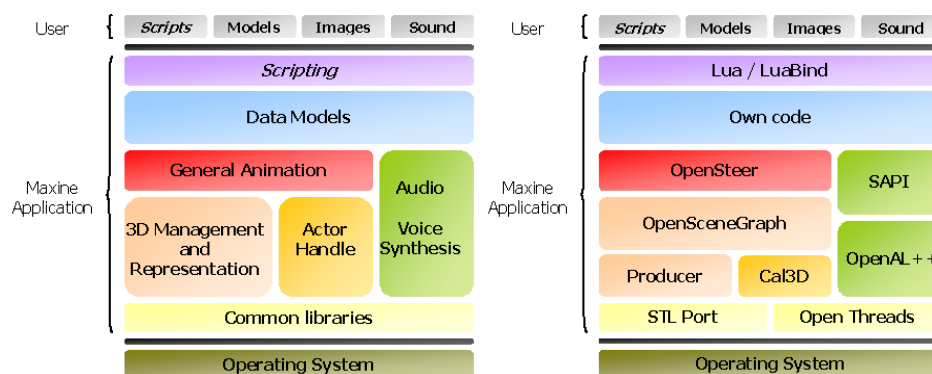
2.2 Engine's architecture

Fig. 1. Maxine general architecture (left). Detail of the libraries used (right).

The main blocks that compose the general architecture of Maxine are shown in the left graph of Figure 1. The files of models, images and sounds allow the construction of the virtual environment. The load and the management of the scene are done via scripts, either through files or by typing commands.

In the right graph of Figure 1 there is a detail of the libraries used in each block:

- OpenSceneGraph [17]: It is the core of Maxine and is in charged of managing the scene graph and the graphical representation of the virtual environment. The render is based on OpenGL.
- Cal3D [13]: This library manages the local animation of the virtual characters. It is based on skeletal mesh deformations.
- OpenAL++ [18]: It is used for handling the 3D and the ambient sound.
- SAPI [19]: It performs the voice synthesis and generates the information necessary for the lip animation and synchronization. Due to real-time requirements this is the only non open source library that has been used (it is freely distributed but owned by Microsoft).
- OpenSteer [20]: This library enriches the navigation of the scene elements by adding steering behaviours.
- Lua and LuaBind [21]: They make possible the creation and manipulation of objects through the scripting language Lua.
- Producer [17]: This library is in charge of 3D visualization and camera management.
- STL port [22]: This is a multiplatform C++ Standard Library
- OpenThreads [23]: It is a standard library with basic programming and thread manipulation tools.

A big effort has been invested in the integration and communication of all these libraries. In this way, to work with different libraries implies to know their capabilities but, in many cases the existent documentation was so poor, that a direct study of the code was essential to understand the correct features of each tool. In other cases, it was indispensable to modify the tool code in order to adapt it functionality to our requirements. The use of dynamic libraries greatly facilitates the development of huge applications like ours, but unlikely, the resulting code is very complicated to be debugged. For example, the use of different coordinate systems in the libraries or the different algorithms for calculus (matrix rotations, quaternion, etc) imply the addition of code for doing the appropriate transformations between them.

3   Our presentation system: MaxinePPT

The general animation engine has been used in a specific application that allow to present PowerPoint information through a like-life character on a graphical display. In this way, the traditional human presenter in Power Point presentations is replaced by a 3D virtual character, and this is especially useful when the same presentation must be repeated several times.

The interaction between the virtual presenter and the user is performed in real-time through the avatar's voice (by text to speech translation), its gestures and expressions and through lip synchronization. The virtual avatar can also answer the questions of the audience but, till this moment, the answers must be typed so that the avatar can "tell" them through its voice and gestures.

However, as the environment includes a recently acquired tracking software and speech recognition software, in a near future, the user communication with the pre-

senter will be done entirely by voice through a microphone or detecting the user position with the tracking system.

3.1 General description

MaxinePPT is an application that makes possible to use the information contained in a PowerPoint file to perform a virtual 3D presentation (see Figure 2).
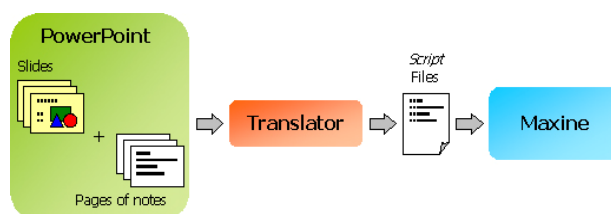

Fig. 2. Overview of the MaxinePPT application

The most important features of the application are:
- It is able to create and perform a virtual presentation in a 3D virtual scenario from a classical PowerPoint file enriched with virtual actors and additional information like videos, images, etc.
- All the aspects of the virtual presentation are controlled by a XML-type language, called PML (Presentation Markup Language). The PML instructions are added to the page notes of the PowerPoint slides in order to determine, for example, the text to be said by the avatar.
- Once the presentation is created, the intervention of the user is not necessary. The presentation is performed automatically and verbally, either in Spanish or English, by using a virtual character created by the user, or the default avatar included in the system. The virtual presenter doesn't read the slide; it says what has been previously specified.

3.2 Generating a 3D presentation

In Figure 3 the general process to create a presentation is shown.
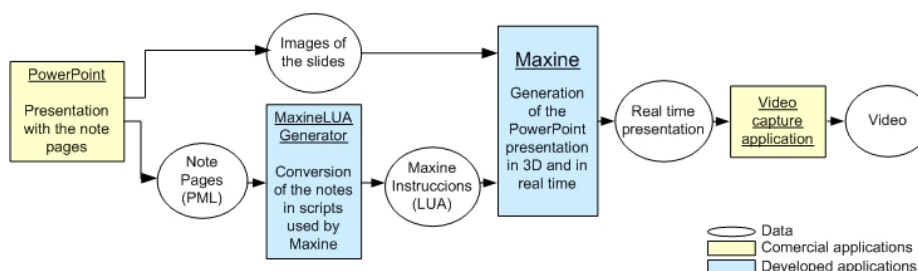

Fig. 3. Global view of the process to generate a 3D presentation.

A more detailed explanation of the steps to be carried out follows.

PowerPoint Presentation Generation: Basically, a MaxinePPT presentation consists on two parts: the images taken from the PowerPoint slides and the pages of notes, which are used to control the presentation. The page of notes contains two types of PML tags:

- The initialization and settings tags, which allow to: select the virtual presenter, select the scenario, configure the voice, configure the illumination, and set the global arrangement and motion of the slides.
- The presentation control tags, which allow to: control the visualization, flow, voice and sound aspects of each slide, in particular: camera selection, presentation pauses, presenter speech, presenter movements, sort of the slides and sound reproduction.
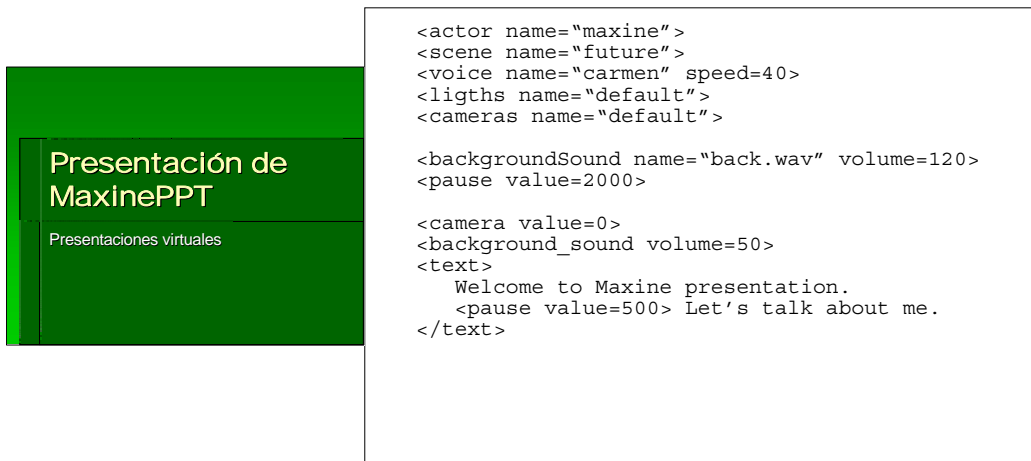
Figure 4 shows an initial slide with its page of notes.


```
<actor name="maxine">
<scene name="future">
<voice name="carmen" speed=40>
<ligths name="default">
<cameras name="default">

<backgroundSound name="back.wav" volume=120>
<pause value=2000>

<camera value=0>
<background_sound volume=50>
<text>
    Welcome to Maxine presentation.
    <pause value=500> Let's talk about me.
</text>
```

Fig. 4. Initial slide of a presentation and the corresponding note page.

Conversion from PML to LUA: The pages of notes written in PML are translated with the MaxineLUAGenerator into the LUA scripting language in order to be understood by the animation engine.

Real-time generation of presentations: The results of the previous steps (the slide images and the LUA instructions) are interpreted by the animation engine Maxine, which carries out the virtual presentation in real-time. So that, enjoy with the presentation is the only thing that remains to be done!

As well as projecting the presentation in a traditional PC screen, it can be projected in an advanced visualization system called CLS (CAVE-Like System) [24] developed by our group and that constitutes an alternative low-cost approach to the classical CAVE [25].

Capturing video: Real-time generation of high quality presentations involves powerful equipment. Therefore, the presentation can be captured so that it can be viewed in any equipment and without needing to install Maxine.

4 Results

The use of MaxinePPT is very simple. The user defines the presentation analogously as it were a film, by means of a script that specifies what is going to happen at every moment. More advanced users can generate their own presenters and scenarios.

The presentation generated with this application can be used for exposing a theme to an audience, showing the contents of the slides and using the virtual avatar to communicate with people using its movements, its voice, its expressions,… However, also the audience can interact with the avatar. A person of the audience can make questions, and the answers to the questions can be typed so that the application makes that the virtual avatar can "tell" the answers through its voice and gestures.

Figure 5 shows the virtual presenter "explaining" the contents of a typical index slide. In Figure 6, a full-length presenter is showing and explaining some images.

Figure 7 presents part of the script file used in the presentation for the creation of the virtual presenter.

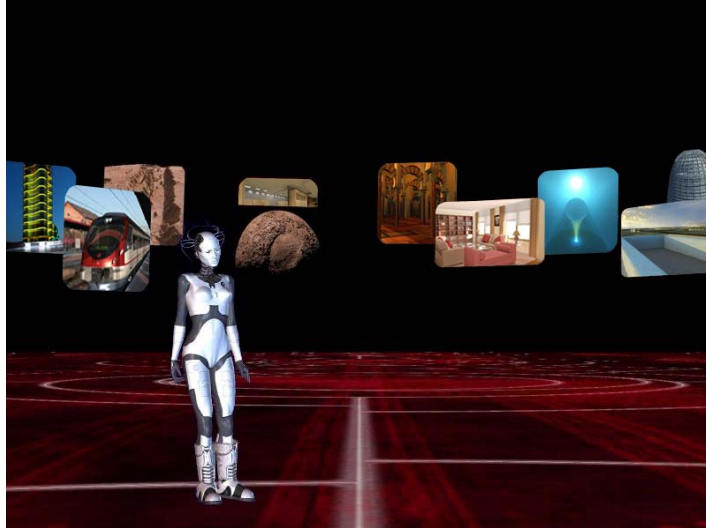**Fig. 5.** Virtual presenter "explaining" an index slide

Fig. 6. Presentation performed by a complete body virtual presenter

```
-- Creation of the actor element
act = actor();

-- File loads
DIR = './data/maxine/';
act:loadSkeleton(DIR .. "woman.csf");
act:loadMesh(DIR .. "woman_body3.CMF");
act:loadMesh(DIR .. "woman_leye.CMF");
act:loadMesh(DIR .. "woman_reye.CMF");
act:loadMaterial(DIR .. "woman_suit.xrf");
act:loadMaterial(DIR .. "woman_eyes.xrf");
act:loadTextures(DIR);

-- Animations
act:loadAnimation("relax", DIR .. "relax_pose.CAF");
act:loadAnimation("breath", DIR .. "relax_breath.CAF");
act:loadAnimation("show", DIR .. "show_pose.CAF");
act:loadAnimation("blink", DIR .. "blink.CAF");
...
act:loadVisemeAnim("vis_silence", DIR .. "visemeDEFAULT.CAF");
act:loadVisemeAnim("vis_aa", DIR .. "visemeAA.CAF");
act:setVisemeAnim ("vis_ae", "vis_aa");
...

-- Creation of the actor from the files
act:create();
act:startPeriodic("blink",0.05,0.05,1,3);
act:setPeriodicVar("blink",2);
act:startPose('relax',0,0,1);
```

Fig. 7. Script file for the creation of the virtual presenter

5   Conclusions and Future Work

An open source animation engine, called Maxine, has been developed for the management of virtual environments and characters. Some of its potentials have been shown through a specific application called MaxinePPT. The animation engine has been designed to satisfy the following requirements:

- Natural motion with presentation skills.
- Real-time motion generation of virtual actors synchronized with speech.
- Proper inputs for representing presentation scenarios.
- Wide applicability of the system.

In order to improve the interaction between the presenter and the spectators we are working now in the following aspects:

- The use of a voice recognition system so that all interaction can be made through voice.
- Recently, a tracking system based on ultrasound, with wand button interaction, has been acquired. The exact information of position and orientation and a suitable tracking of the movement of a person will be essential to achieve immersion feeling in real time simulations.
- Also the character would be provided the character with face-to-face conversation capabilities. Deploying the ALICE framework for programming chat robots [26], pattern matching rules will be applied to derive request for presentation goals and new settings of presentation parameters from a user's input message.

To accomplish these new functionalities a new application is being developed: an avatar acting as an assistant to help the user controlling a domotic system for handicapped users.

Acknowledgments

Bibliography

1. Nijholt A. "Multimodal interactions in ambient intelligence", in W. Verhaegh, J. Korst, E. Aarts (eds). Algorithms in ambient intelligence, Chapter 2. Phillips Research Book Series. Dordrecht: Kluwer Academic Publishers, 2003.
2. Nijholt A. "Where computers disappear, virtual humans appear"; Computer & Graphics N° 28, pp. 467-476, 2004.
3. Casell J., Sullivan J., Prevost S., Churchill E (eds). "Embodied Conversational Agents", Cambridge: MIT Press, 2000.

4. Noma T., Zhao L., Badler N. I. "Design of a Virtual Human Presenter"; IEEE Journal of Computer Graphics and Applications, Vol.20, No.4, July/August, 2000, pp. 79-85.

5. Magnenat-Thalmann N., Kalra P. "The Simulation of a Virtual TV Presenter", Proc. Pacific Graphics 95, World Scientific, Singapore, 1995, pp. 9-21.

6. André E., Rist T., Müller J. "Integrating reactive and scripted behaviours in a Life-Like presentation agent", Proceedings of the 2nd International Conference on Autonomous Agents, pp. 261-268, 1998

7. Baus J., Butz A., Krüger A. "Incorporating a Virtual Presenter in a Resource Adaptive Navigational Help System"; Workshop notes of the Workshop on Guiding Users through Interactive Experiences, Paderborn, Germany, April, 2000

8. Nijholt A., van Welbergen A., Zwiers J. "Introducing an embodied virtual presenter agent in a virtual meeting room"; Proc. of the 23rd IASTED International Multi-Conference Artificial Intelligence and Applications, February, 2005 Innsbruck, Austria, pp. 579-584.

9. Rist T., André E., Baldes S., Gebhard P., Klesen M., Kipp M., Rist P., Schmitt M. "A Review of the Development of Embodied Presentation Agents and Their Application Fields in Life-Like Characters: Tools, Affective Functions, and Applications"; Helmut Prendinger and Mitsuru Ishizuka (eds), Springer Series on Cognitive Technologies, pp. 377-404, 2003

10. Yuan X., Chee Y.S. "Design and evaluation of Elva: an embodied tour guide in an interactive virtual art gallery"; Computer Animation and Virtual Worlds 16, pp.109-119, 2005

11. Rocchi C., Stock O., Zancanaro M., Kruppa M., Krüger A. "The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices"; Proceedings of the 9th International Conference on Intelligent User Interfaces IUI'04, January 13-16, Madeira, Portugal (2004)

12. Focken D., Stiefelhagen R. "Towards vision-based 3-D people tracking in a smart room"; Proceedings 4th IEEE International Conference on Multimodal Interfaces. 2002.

13. Cal3D: Character Animation Library, http://cal3d.sourceforge.net/

14. Virtual Human Markup Language, VHML Working Draft v0.3, http://www.vhml.org/ 2003.

15. ISO/IEC JTC1/SC29/WG11 N4668, MPEG-4 Overview; http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm, 2002.

16. Vicon Motion Systems – Optical Motion Capture Technology for Science, Engineering and Entertainment Industries, http:// www.vicon.com

17. Open Scene Graph, http://www.openscenegraph.org/

18. OpenAL++, http://alpp.sourceforge.net/

19. Long B. "Speech Synthesis & Speech Recognition Using SAPI 5.1", http://www.blong.com/Conferences/DCon2002/Speech/SAPI51/SAPI51.htm

20. OpenSteer, http://opensteer.sourceforge.net/

21. LuaBind, http://www.rasterbar.com/products/luabind.html

22. STLPort, http://sourceforge.net/projects/stlport

23. OpenThreads, http://openthreads.sourceforge.net/

24. Gutierrez D., Seron F.J., Magallon J.A., Sobreviela E.J., Gutierrez J.A. "CLS: A Low Cost Visualization Environment for the Train Industry"; Journal Mécanique & Industries 5: 139-145, 2004.

25. Cruz-Neira C., Sandin D.J. DeFanti T. A. "Surround-screen projection-based virtual reality: The design and implementation of the CAVE"; Computer Graphics (SIGGRAPH '93 Proceedings), vol. 27: 135–142, 1993.

26. Alice v2.0: Learn to Program Interactive 3D Graphics, http://www.alice.org/