A cross-platform, remotely-controlled mobile avatar simulation framework for AmI environments

Emmanouil Zidianakis
Foundation for Research and Technology – Hellas (FORTH),
Institute of Computer Science &
University of Crete, Department of Computer Science
Heraklion, GR-70013, Greece
zidian@ics.forth.gr

George Papagiannakis
Foundation for Research and Technology – Hellas (FORTH),
Institute of Computer Science &
University of Crete, Department of Computer Science
Heraklion, GR-70013, Greece
papagian@ics.forth.gr

Constantine Stephanidis
Foundation for Research and Technology – Hellas (FORTH),
Institute of Computer Science &
University of Crete, Department of Computer Science
Heraklion, GR-70013, Greece
cs@ics.forth.gr

Abstract

Nowadays, users are able to interact with digital content using their mobile devices almost everywhere and anytime due to the increased power, portability, and ubiquitous connectivity of mobile devices. This paper presents the design and implementation of a novel, remotely controlled for the purposes of edutainment and instructor-student interaction, three-dimensional full body avatar gamification framework. The main innovation introduced focuses on multi-presence gamified educational scenarios in multiple desktop computers and mobile devices. Thus the remotely-controlled avatar can act as a guide, assistant or information presenter for novel, cross-platform Ambient Intelligence (AmI) edutainment scenarios. In detail, the avatar's role depends on the requirements of the AmI client applications as these are propagated remotely (using remote procedure calls). Examples of remote function invocations include real-time 3D biped skinned animations, text-to-speech, producing facial expressions and presenting multimedia content.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Animation

Keywords: Performance, Design, Experimentation, Human Factors virtual assistant, ambient intelligence, text to speech, lip-synchronization, biped skeletal animation, presence, virtual character simulation

1. Introduction and main concept

Nowadays users can interact with digital content using their mobile devices almost everywhere and anytime due to their increased power, portability, and ubiquitous connectivity. In the context of Ambient Intelligence (AmI), as Information Communication Technologies (ICT) get integrated to the environment processing power, input and output devices are constantly present and available to fulfill several human needs. According to [Prendinger et al 2005], Ambient Intelligence is described as electronic environments that are sensitive and responsive to the presence of people. More and more electronic devices are being constantly integrated in the environments surround people e.g. houses.

The top level goal is to achieve natural communication with the environment imitating aspects of human communication. People usually communicate with other people through the senses of sight and hearing. A natural interface should be capable of imitating this behavior though speech and gestures. This paper focuses on the design and implementation of a three-dimensional avatar that can be displayed almost in any device including desktop computers, tablets and smart phones. Avatars are virtual characters which make communication between user and machine more natural and interactive.

Figure 1: *The cross platform three-dimensional avatar*

Max is a remote controlled full body three-dimensional avatar that supports multi-presence almost in any available device in an AmI environment (see Figure 1). The role of Max depends on the client-application's requirements. Typical examples include acting as a guide, assistant or information presenter. In order to achieve natural communication channels both, non-verbal and verbal behavior is essential. Non-verbal communication includes full body animation and facial expressions. For example, when idle, the virtual character is never motion-less due to an undulating body animation and eyes blinks randomly, giving the illusion that is alive. The facial animation is strengthened by raising the eyebrows. Max can also present multimedia content on the television contained in the scene.

Max currently accompanies an augmented interactive table (see Figure 2) suitable for young children from 3 to 7 years old [Zidianakis et al 2012] called Beantable. The objectives of Beantable are to: a) integrate AmI technologies into children's playtime, b) support child development through playing and c) provide intuitive tools that monitor and enhance the child's playing experience. In this setup, an iOS mobile device (iPad), is

used to present the child's virtual partner supporting various modes of interaction such as playful or didactic. Each game requests Max to act as the friend or the opponent of the child, giving instructions or insisting on the completion of a task.

Figure 2: *Max acting as virtual partner for the needs of an augmented interactive table for young children.*

Within the same setup, Max employs a large display to create the so called "Mimesis game". In this game, Max is presented in the large display that integrates a Windows 7 operating system and interacts with the child in a very natural way using both verbal and no-verbal communication channels. More specifically, he requests the child to assume various body postures he does. Using a sensory infrastructure presented in [Zidianakis et al 2014], the game measures the quality and performance of the body posture that the child assumes and extracts indications of the archived maturity level and skills of the child.

Figure 3: *Max in Mimesis game; asks the child to assume various body postures he does.*

2. Related work

A growing number of research projects have begun to investigate the use of animated life-like characters in natural user interfaces because they present a priori a lot of advantages, which have also been validated by many authors. The authors of [Arita et al 2004] describe interaction via an avatar and show that communication via an avatar can be useful. In another research, an avatar is used as a personal assistant to interact with the television [Ugarte et al 2007]. According to many authors, the use of animated life-like characters in natural user interfaces presents a lot of advantages such as: a) social interaction [Nass et al 1994][Nijholt 2003][Prendinger et al 2005], b) user attention [Hongpaisanwiwat et al 2003][Kim et al 2007], c) naturalness, d) more information in the transmitted message [Mehrabian 1968] and e) trustworthiness and believability [Koda et al 1996]. Previously we had explored remotely-controlled virtual humans for educational scenarios [Papagiannakis 2013] as well as interactive virtual characters for presence in mixed reality [Egges et al 2007][Papagiannakis 2013][Papagiannakis et al 2014] but not in a cross-platform, mobile, multi-device AmI environment.

3. Architecture and main novelties

The novel AmI architecture presented in this section mainly focuses on the development of a uniquely cross-platform, mobile three-dimensional avatar that could be remotely controlled by any client application in the context of an AmI environment. To achieve this, a top level goal was to support AmI applications, games, information kiosks, navigation guides, etc. in their communication with Max. This is carried out through a networking middleware as depicted in Figure 1. The middleware network is designed to facilitate the communication of systems that are deployed on diverse platforms as presented in [Georgalis et al 2009]. Max accepts remote procedure calls from diverse clients to animate, to read some text, to assume a posture, to generate a specific facial expression, or to present multimedia content. As shown in Figure 4, these functions are served by the following software components: a) Camera Path Animations, b) Skeleton Animations, c) Text to Speech, d) Media Presentation and e) Facial Expressions.

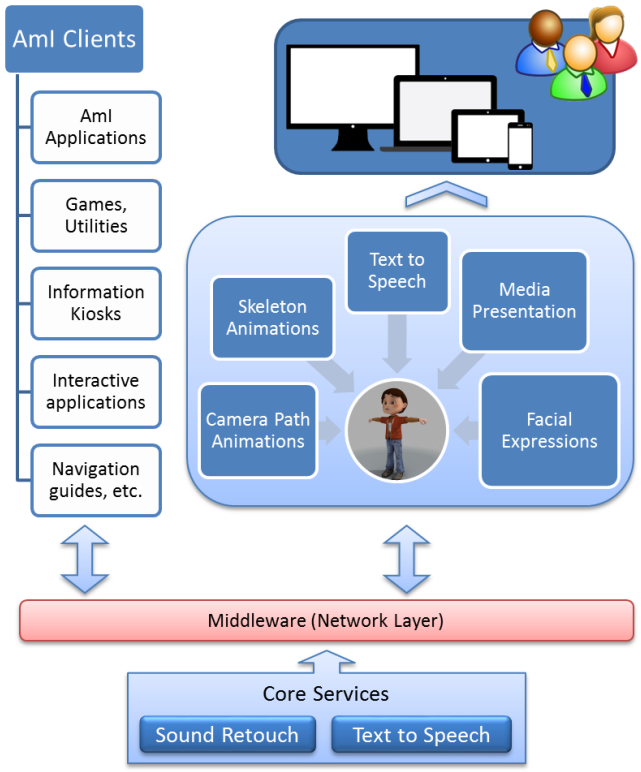

Figure 4: *Architecture of Max: A three-dimensional remotely controlled cross-platform avatar*

As mentioned before, Max enhances natural communication using verbal communication channels. Therefore, it was of the utmost importance to adopt a generic solution that could be able to convert text into computer-generated voice output. Unfortunately, there is no yet available a cross platform text to speech (TTS) system capable to run in any device or platform. To overcome this difficulty and make Max able to talk regardless the selected device or platform, a set of TTS services were developed. This set includes: a) a sound retouch service, and b) a text to speech service. Both services are hosted by a windows server that

integrates the Microsoft Speech API (SAPI). As a result, text to speech service provides access to the functionality of an installed speech synthesis engine and therefore is able to turn requested text into speech using various free or commercial TTS products like Acapella or Loquendo. Through middleware the text to speech service accepts requests from any remote client application, i.e. Max, and returns back (through the network) the speech audio stream accompanied with the phonemes and visemes reached. If necessary, the sound retouch service can be used to change the tempo, pitch and rate of the speech audio stream received as input to match custom defined speaker characteristics, i.e. the voice age.

4. Implementation

Max is a six year old low polygon three-dimensional character (i.e. 26K polygons) which originally was purchased from Turbosquid and modified in order to achieve increased rendering performance (reducing the number of polygons to 23K). The model was already rigged with biped system and was animation ready including some facial expressions. Minor changes were made where necessary in the context of skinning and texturing.

Now, Max is available in Collada, which is an open standard XML schema for exchanging digital assets among various graphics software applications. Max "came to life" using OpenSceneGraph 3.0.1; an open source cross platform 3D graphics application programming interface. OpenSceneGraph is used by application developers in fields such as visual simulation, computer games, virtual reality, scientific visualization and modelling. Max is currently built for Apple's mobile operating system (iOS) and for Windows operating system, as well as tested in iPhone, iPad and PC. Some of the technical challenges that had to be addressed for the final functional solution were: a) the expansion of the middleware framework to include and support [Georgalis et al 2009] iOS devices b) various hardware imposed limitations, and c) the development of a text to speech solution able to support diverse platforms. The frame rate in mobile devices was measured at 24 fps without shadows and at 60 fps with shadows enabled in PC using an ordinary graphics card. According to the aforementioned architecture, Max is comprised of different modules that are responsible for executing remote requests that are received from client applications. These modules are described in detail in the following sections.

4.1 Text to Speech

Max supports lip-synchronization using the aforementioned text to speech service. In detail, when Max receives some text to read, he uses the text to speech service (via middleware) to get back the equivalent audio stream accompanied with a list of the reached visemes. According to the request's requirements, Max may submit the audio stream to the sound retouch service to change voice characteristics i.e. voice tone, pitch or tempo. That list contains information about each reached viseme in the following format: *AudioPosition\Duration\Emphasis\NextViseme\Viseme*.

Given that information, Max before starting reading, prepares an animation consisting of various float linear animation channels. As shown in Figure 5, each animation channel smoothly activates or deactivates the morph target corresponding to each viseme contained in the visemes list. In detail, a morph target begins to fade in from the previous viseme's audio position added by its duration and fades out until the audio position of the next one. Random eyes blinking is implemented in the same way. Various tests indicated that this technique has better results even if some animation key-frames are lost due to rendering instability. Every received request can be identified by a unique keyname. The purpose of the latter is to avoid unnecessary requests to the text to speech and sound retouch service. When TTS data already exist in Max's documents folder these are recycled to reduce network traffic unless client forces cache replacement. Max sends an event to the remote client in order to notify that reading has finished.

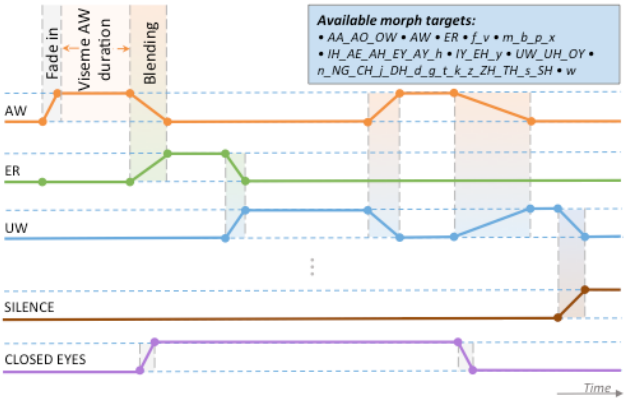

Figure 5: *Blending formula among various animation channels for each morph target.*

4.2 Facial Expressions

Max shows emotions through facial expressions. Typical examples of morph targets used in facial animation include a smiling mouth, a closed eye, and a raised eyebrow. Apart from making a face using only a morph target, Max is able to animate among different facial expressions in a timeline. To this end, facial expression manager adopts the same blending functionality, as described in the previous section, to animate among different morph targets. Additionally, a remote client can submit a 3D model in .dae format that contains a morph geometry which corresponds to a new facial expression. Max sends an event back to the remote clients in order to notify that a facial expression or animation has completed.

4.3 Camera Path Animations

The camera's translation and rotation matrix can change dynamically depending on the needs of remote clients. For example, during multimedia presentation, the camera moves to focus the television. As a result, Max can present multimedia content in full screen mode on the running device. To this end, Camera Path Animation module is responsible to animate the camera's position and orientation according to the path received remotely from the client application. Additionally, it can provide information about the current's camera position and orientation as well as to receive a text file that contains a timeline of paths. Each timeline is identified and stored locally by its unique keyname for caching. Max sends an event to the remote clients in order to notify that a camera path animation has completed.

4.4 Media Presentation

Multimedia content presentation is done by the television next to Max. The Media Presentation module uses ffmpeg to open and play almost any known format of videos or images while audio playback is currently supported only for .wav audio streams. To reduce network traffic, Media Presentation module is able to receive and store locally a multimedia file and play it when needed. Furthermore, it sends an event to notify that a multimedia

content presentation has ended. In case of pictures, that event is sent right after the appearance of the image.

4.5 Skeleton Animations

Max implements body animation through biped skeletal animation using key-frame interpolation. Typical examples of biped animations include walking, clapping, running, dancing and a bowing. Remote clients may receive a list of available animations or add a new one by submitting a 3D model in .dae format containing an animation. Clients can also set the number of loops as well as the playing mode (once, stay, loop or ping pong). Skeleton Animation module notifies remote clients when the animation has finished.

5. Conclussion and lessons learned

This paper presented a novel, mobile, cross-platform remotely controlled full body three-dimensional avatar supporting multi-presence almost in any available device in an AmI environment. To the best knowledge of the authors, such a platform does not exist yet in an open, research-oriented environment and could be achieved only with closed game-engines and COTS components that do not allow diversity, and extensibility. To this end a service oriented distributed architecture was employed to achieve device independence and the universal provision of facilities such as TTS and rendering of multimedia content. Hence among the major achievements of this research work are: a) open, cross-platform, mobile framework for avatar simulation, b) remote control of virtual characters via multi-device (iOS and Windows) AmI environment that facilitate c) novel edutainment scenarios. Currently Max is hosted in a children's playroom setup within FORTH-ICS's AmI research facility. Regarding future improvements it is considered crucial to extend the support for more platforms such as Android phones to holistically address the mobile computing market. Additionally, an evaluation strategy will be designed with the active contribution of the end users and research questions will be formulated around user experiences.

6. Acknowledgments

This work is supported by the FORTH-ICS internal RTD Programme 'Ambient Intelligence and Smart Environments'.

7. References

ARITA, D., AND TANIGUCHI, R. I. 2004. Real-time human proxy: An avatar-based interaction system. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 419-425). Springer Berlin Heidelberg.

EGGES, A., PAPAGIANNAKIS, G., AND MAGNENAT-THALMANN, N. 2007. Presence and interaction in mixed reality environments. In *The Visual Computer*, *23*(5), 317-333.

GEORGALIS, Y., GRAMMENOS, D., AND STEPHANIDIS, C. 2009. Middleware for ambient intelligence environments: Reviewing requirements and communication technologies. In *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments* (pp. 168-177). Springer Berlin Heidelberg.

HONGPAISANWIWAT, C., AND LEWIS, M. 2003. Attentional effect of animated character. In *Proceedings of the human-computer interaction*, 423-430.

KIM, Y., BAYLOR, A. L., AND SHEN, E. 2007. Pedagogical agents as learning companions: The impact of agent emotion and gender. In *Journal of Computer Assisted Learning*, *23*(3), 220-234.

KODA, T., AND MAES, P. 1996. Agents with faces: The effect of personification. In *Robot and Human Communication, 1996., 5th IEEE International Workshop on* (pp. 189-194).

MEHRABIAN, A. 1968. Communication without words. In *Psychological today*, *2*, 53-55.

NASS, C., STEUER, J., AND TAUBER, E. R. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78). ACM.

NIJHOLT, A. 2003. Disappearing computers, social actors and embodied agents. In *Cyberworlds, 2003. Proceedings. 2003 International Conference on* (pp. 128-134). IEEE.

PAPAGIANNAKIS, G. 2013. Geometric algebra rotors for skinned character animation blending. In *SIGGRAPH Asia 2013 Technical Briefs*, 11.

PAPAGIANNAKIS, G., PAPANIKOLAOU, P., GREASSIDOU, E., AND TRAHANIAS, P. 2014. glGA: an OpenGL Geometric Application framework for a modern, shader-based computer graphics curriculum. In *Eurographics 2014, Education Papers*, 1–8.

PONDER, M., HERBELIN, B., MOLET, T., SCHERTENEIB, S., ULICNY, B., PAPAGIANNAKIS, G., MAGNENAT-THALMANN, AND THALMANN, D. 2002. Interactive scenario immersion: Health emergency decision training in JUST project. In *VRMHR2002 Conference Proceedings*.

PRENDINGER, H., MA, C., YINGZI, J., NAKASONE, A., AND ISHIZUKA, M. 2005. Understanding the effect of life-like interface agents through users' eye movements. In *Proceedings of the 7th international conference on Multimodal interfaces* (pp. 108-115). ACM.

RUYTER, B. DE, AND AARTS, E. 2004. Ambient intelligence: visualizing the future. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 203-208). ACM.

UGARTE, A., GARCÍA, I., ORTIZ, A., AND OYARZUN, D. 2007. User interfaces based on 3D avatars for interactive television. In *Interactive TV: a Shared Experience* (pp. 107-115). Springer Berlin Heidelberg.

ZIDIANAKIS, E., ANTONA, M., PAPAROULIS, G., AND STEPHANIDIS, C. 2012. An augmented interactive table supporting preschool children development through playing. In *Proceedings of the AHFE International 2012*, July 21-25, 2012 - San Francisco, California, USA

ZIDIANAKIS, E., PARTARAKIS, N., ANTONA, M., AND STEPHANIDIS, C. 2014. Building a Sensory Infrastructure to Support Interaction and Monitoring in Ambient Intelligence Environments. In *Distributed, Ambient, and Pervasive Interactions* (pp. 519-529). Springer International Publishing.