Dublin Institute of Technology
ARROW@DIT

9th. IT & T Conference

School of Computing

2009-10-01

Intelligent Virtual Agent: Creating a Multi-modal 3D Avatar Interface

Mark Dunne
Dublin Institute of Technology, mark.dunne@dit.ie

Brian Mac Namee
Dublin Institute of Technology, brian.macnamee@dit.ie

John Kelleher
Dublin Institute of Technology, john.d.kelleher@dit.ie

Follow this and additional works at: http://arrow.dit.ie/ittpapnin

Part of the Computer Sciences Commons

Recommended Citation

Dunne, M., Mac Namee, B., Kelleher, J. (2009) Intelligent virtual agent: creating a multi-modal 3D avatar interface. *Ninth IT & T Conference,* Dublin Institute of Technology, Dublin, Ireland, 22- 23, October

This Conference Paper is brought to you for free and open access by the School of Computing at ARROW@DIT. It has been accepted for inclusion in 9th. IT & T Conference by an authorized administrator of ARROW@DIT. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie.

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

Intelligent Virtual Agent: Creating a Multi-Modal 3D Avatar Interface

Mark Dunne, Brian Mac Namee and John Kelleher

DIT AI Group, Dublin Institute of Technology, Dublin, Ireland
firstname.lastname@dit.ie

Abstract

Human-computer interactions can be greatly enhanced by the use of 3D avatars, representing both human users and computer systems in 3D virtual spaces. This allows the human user to interface with the computer system in a natural and intuitive human-to-human dialog (*human face-to-face conversation*). Hence, continuing to blur the boundaries between the real and virtual worlds. This proposed avatar system will go a step further and will use a camera to track the user's head and eye movements during the dialog. This information will help to build rapport between the user and computer system by registering the user's emotional state and level of interest. The system will adjust the dialog according to this information paying special attention to the user's feedback. For example, one obvious benefit for head and eye tracking will be to allow the avatar to make and keep realistic eye contact with the user, but there is lots of room to expand on these techniques in this research.

Keywords: Avatar, IVA, Interface.

1 Introduction to the LOK8 and Avatar Module

1.1 LOK8

The goal of the LOK8 (pronounced locate) project is to create a new and innovative approach to human-computer interactions. With LOK8 a person will be able to engage in meaningful interaction with a computer interface in a much more natural and intuitive way than we are used to. A virtual character will be displayed in numerous locations depending on the user's position and context. Users will be able to communicate with this virtual character through speech and gestural input/output, which will be processed and controlled by the dialog management component of the system. This will allow "face-to-face" interactions with the LOK8 system. The LOK8 system will deliver content to the user in a variety of context-specific ways with the aim of tailoring content to suit the user's needs. In addition to screens and projectors displaying the avatar, the user's mobile device, as well as speakers within the environment, will be used to deliver focus-independent content. Ultimately the goal is to replace a human-computer interface with a *human-virtual human interface*. (See Figure 1)

1.2 Avatar Module

The use of avatars as visual computer interfaces is long been established and the Avatar Module will continue researching in this area. Done well, they contribute to human-computer interactions. Done badly, the interaction becomes tedious, leading the user to switch them off, for example, the infamous *Microsoft Clippy* [1]. Avatars should be used to enrich the user's experience, not frustrate it.
The interaction style should be intuitive and natural, contributing to the overall immersive experience. Computer game technology has been wide spread over the past decade, so people are familiar with avatars. In games, avatars tend to be user-controlled, embodying the user in a virtual world. Also being

customisable, users strive to make them look unique, representing themselves or one of their personae in the virtual world. In the video game industry, computer-controlled avatars are called non-player characters (NPC) but in the world of academia they are known as intelligent virtual agents (IVA). The LOK8 Avatar module is focused on further development in the area of IVA research, where IVAs have been used in many applications [2, 3].

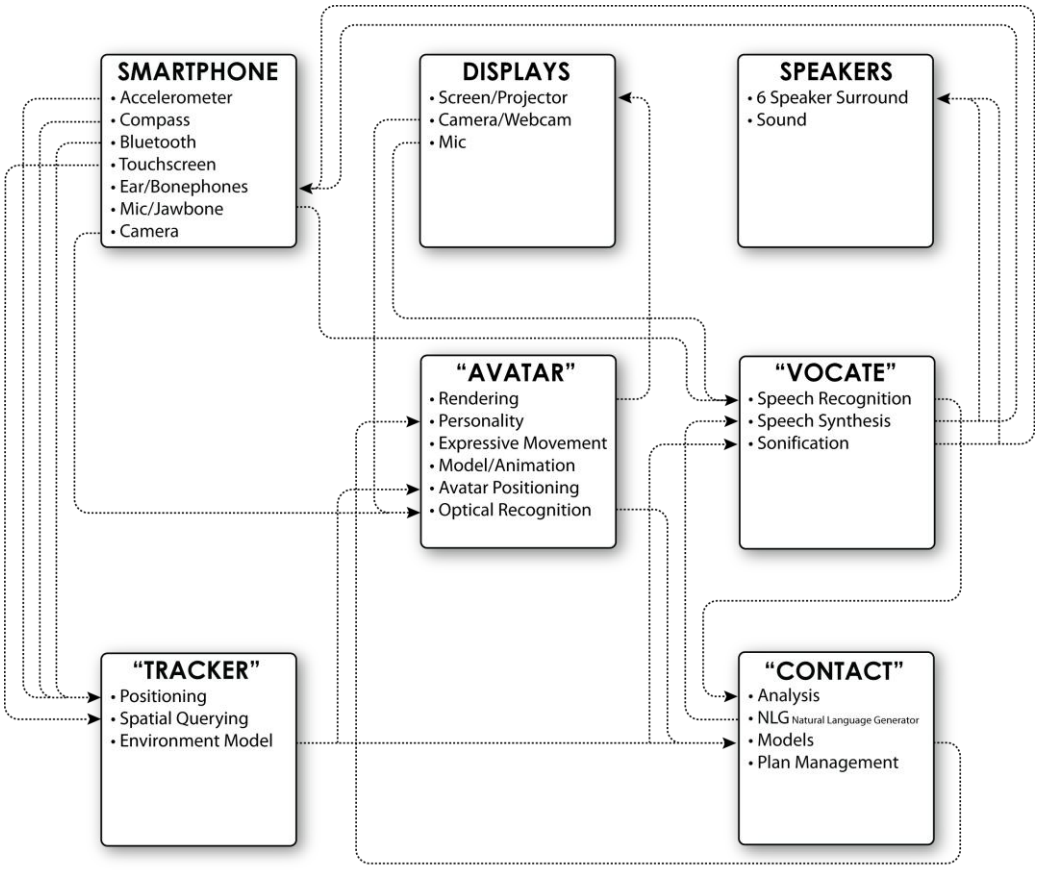

Figure 1: LOK8 Overview Diagram: *Avatar, Contact, Tracker* and *Vocate* Modules.

At the core of the Avatar Module is the development of a scalable visual interface for the LOK8 system using a variety of display modalities, where the avatar can migrate seamlessly across modalities with little or no effort on the user's behalf, adding to the immersive user experience? In order to achieve this, the Avatar Module will not only research some obvious display modality approaches like using wall mounted displays (projectors, LCD screens) but investigate novel display modalities (spatial augmented reality, movable projectors and mobile phones) helping to blur the boundaries of real and virtual worlds. Furthermore, an investigation will be carried out into how an IVA should behave in a multi-user environment. The Avatar Module has two main objectives:

- **Investigate the design of an intelligent avatar/IVA**
- **Contribute to an overall multi-modal interface**

2 Related Work

IVA research requires a multi-disciplinary approach with the main disciplines being as follows: Human figure animation; facial recognition and animation; perception; cognitive modeling; emotions and personality; natural language processing; speech recognition and synthesis; non-verbal communication; distributed simulation and computer games [4]. For human-level intelligence to occur, a core number of capabilities are required, including planning, belief representation, communication ability, emotional reasoning, and a way to integrate these capabilities [5]. The Avatar

module will focus on four key areas, while maintaining strong ties with the other three LOK8 modules to create an intelligent system:

- **Face-to-Face Conversation**
- **Human Face and Figure Animation/Modeling**
- **Emotions and Personality**
- **Integration**

2.1 Face-to-Face Conversation

Non-verbal communication (NVC) (body language, hand and head gestures, facial expressions and eye gaze) can contain lots of information, which can be beneficial to all parties involved in human face-to-face conversations. Eye gaze alone can indicate interest, attention and involvement [6]. If the avatar is to act in a humanistic way, it will be worthwhile and important for the LOK8 system to interpret such NVC displayed by the user during their interaction. The avatar could then in turn display complimentary NVC back to the user (feedback), when and where appropriate.
Eye/Face tracking will monitor the user's gaze [7], whilst head tracking will monitor the user's orientation and position in the environment. An obvious benefit for such tracking techniques in the LOK8 system will be the avatar's ability to make realistic eye contact with the user.

2.2 Human Face and Figure Animation/Modeling

Modeling realistic human-like avatars can have a negative outcome. The more human they become in appearance, the more positive people respond to the model, until a certain point where the realism of the model causes repulsion. This point is known as the Uncanny Valley [8]. The main contributing factor for this is that the more life-like a model becomes, the more life-like it is expected to act. If an avatar looks human, then it should move and act human as well.
The LOK8 avatar will be human-like so that people will communicate with it in a normal manner. This avatar will also be able to display a full range of non-verbal communication. Falling somewhere on the visual scale (see Figure 2) between the Nintendo Wii avatar [10], which is simplistic but still portrays a human and the Second Life avatar [11] which is more realistic in appearance. The Xbox 360 avatar [12] falls in the middle; it has enough detail to show a wide variety of human-like movement and expressiveness. It is vital that the LOK8 avatar creates the illusion of life, and thus suspends the user's disbelief [9].

Figure 2: Nintendo Wii avatars [10], Second Life avatar [11] and Xbox 360 avatars [12].

2.3 Emotions and Personality

The idea that the mind is the driving force for action [9], adds importance to the addition of emotions in any intelligent conversational system. It has been asserted that a machine will be a more affective communicator if given the ability to perceive and express emotions [13]. An established rule from animation development has been adopted in IVA development and that states; *"for an avatar to be*

real, it must have an interesting personality" [14]. This can be elaborated on by suggesting that the avatar must be human enough for the user to understand and identify with it.
The addition of personality into the avatar should help create more believable agent, rooting the user in a more engaging dialog. This will be particularly important where the user will be interacting with the avatar on a daily-basis, for long periods of time, helping the user and agent build a rapport with each other. Without a personality and memory this would not be possible [15].

2.4 Integration

One solution to integrating all the aspects of the LOK8 avatar would be to have it run in a game engine. However, as the display modalities incorporate many different technologies, running a game engine on them all, might not be possible. A more favorable solution would be to stream data (visual and audio) through adapted web browsers (Chrome) using the new HTML 5 standard [16]. The development for an Application Programming Interface (API) could be beneficial to the passing of raw data between the user and system, as well as the core LOK8 modules.

3 Current Work

The current work been done in Avatar Module, is the production of a test environment where the user will interact with an avatar using head and eye movements. The user will be required to play a 3D Pong game (see Figure 3), where they control their paddle to hit a ball and score points, whilst blocking and preventing the opponent from scoring. Their opponent will be the 3D avatar, watching the avatar's eye and head movement will help the user predict their opponent's next move.

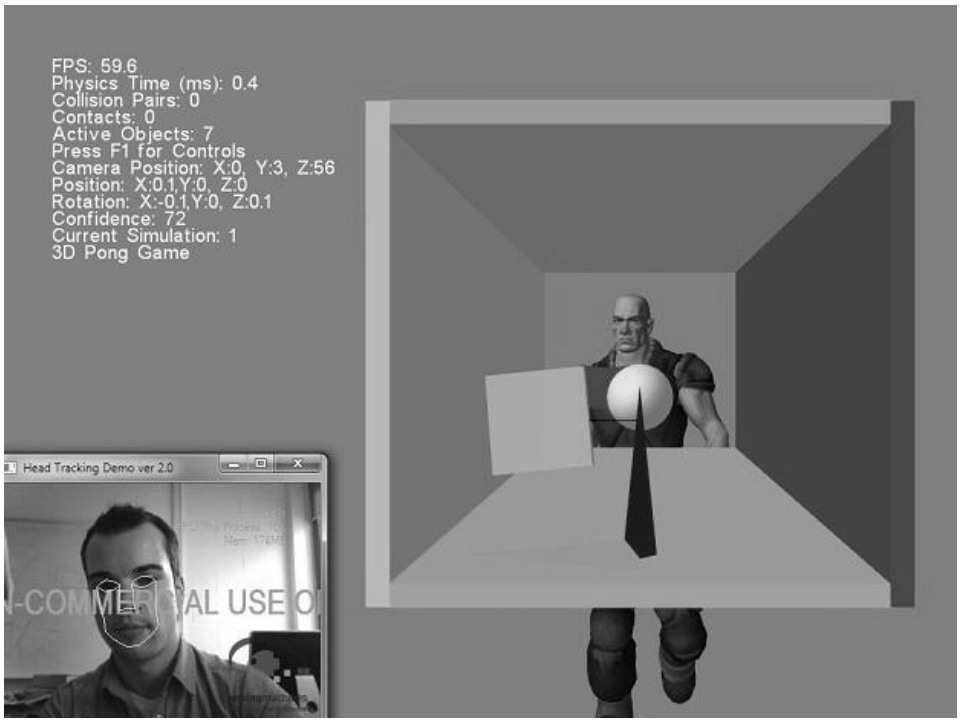

Figure 3: Screen shot of 3D Pong Game in action.

4 Conclusions & Future Work

The purpose of the LOK8 project is to create a new and innovative approach to human-computer interactions. Using a 3D avatar will help the user engage with the system in a more natural and intuitive way. Building emotions and personality into the avatar can contribute to the overall experience, establishing rapport between the user and the avatar. Head and eye tracking will give the

effect of realistic eye contact between avatar and user, immersing the user even more into the interactive experience.

The main focus of the Avatar Module will be developing a scalable visual interface for the LOK8 system on variety of display modalities, many of which are novel. A key feature being the seamless migration of the avatar across these display modalities, i.e., the avatar will migrate from the user's mobile device to a large wall mounted display and then maybe to a mobile robotic platform.

Integrating all of the findings and solutions into the LOK8 test environment will be the key stone of the LOK8 project (Avatar, Contact, Tracker and Vocate Modules).

Future work for the Avatar Module will entail gathering data from the 3D Pong game, which will help in the analysis of this novel interaction style with an avatar discussed above. The ability to adjust the gameplay, game logic and enforce a set of rules will help to create a control experiment and multiple combinations of head and eye, or head only and eye only movement experiments. Test subjects will be required to play a series of these experiments and rate their experiences using a survey/questionnaire.

Acknowledgements: This work is funded by the Higher Education Authority (HEA) in Ireland, Technological Sector Research Strand III: Core Research Strengths Enhancement Programme.

References

[1] USA Today. "Microsoft banks on anti-Clippy sentiment". The Associated Press, 2/6/2002. http://www.usatoday.com/tech/news/2001-05-03-clippy-campaign.htm, 29/7/09

[2] Lim, Mei Yii, and Ruth Aylett. "Feel the Difference: A Guide with Attitude." Intelligent Virtual Agents - 7th International Conference, IVA 2007. Paris, France, September 2007: Springer, 2007. 317-330.

[3] Hung-Hsuan Huang, Aleksandra Cerekovic, Igor S. Pandzic, Yukiko Nakano and Toyoaki Nishida. "Toward a multi-culture adaptive virtual tour guide agent with a modular approach." AI & Society, June 2009, Springer, London.

[4] Gratch, Jonathan, Jeff Rickel, Elisabeth André, Justine Cassell, Eric Petajan, and Norman Badler. "Creating Interactive Virtual, Humans: Some Assembly, Required." IEEE Intelligent Systems. 2002.

[5] Swartout, William, et al. "Toward Virtual Humans." AI Magazine - Volume 27 , Issue 2, 2006: 96-108.

[6] Knapp, Mark L., & Hall, Judith A. (2007) "Nonverbal Communication in Human Interaction. (5th ed.)" Wadsworth: Thomas Learning. ISBN 0-15-506372-3

[7] Eicher, Tobias, Helmut Prendinger, Elisabeth Andre, and Mitsuru Ishizuku. "Attentive Presentation Agents." Intelligent Virtual Agents - 7th International Conference, IVA 2007. Paris, France, September 2007: Springer, 2007. 283-295.

[8] Mori, Masahiro. "The Uncanny Valley." Energy, 7(4), 1970: 33-35.

[9] Martinho, Carlos, and Ana Paiva. "It's All in Anticipation." Intelligent Virtual Agents - 7th International Conference, IVA 2007. Paris, France, September 2007: Springer, 2007. 331-338.

[10] Nintendo Wii Avatars called "Miis", 10/08/2009, http://www.crunchgear.com/wp-content/pbucket/450-300- scale-117364755138.jpg.

[11] Customisation of a Second Life Avatar, 10/08/2009, http://i.gleeson.us/gb/0607/2nd_life/Avatar_creation.jpg.

[12] Microsoft's Xbox 360 Avatars, 10/08/2009, http://scrawlfx.com/wp-content/uploads/2008/11/niko- Avatar-360.jpg.

[13] Picard, R. W. "Affective Computing." MIT Press, Cambridge, 1997.

[14] Thomas, Frank, and Ollie Johnston. Disney Animation: The Illusion of Life. New York: Abbeville Press, 1981.

[15] Ho, Wan Ching, and Kerstin Dautenhahn. "Towards a Narrative Mind: The Creation of Coherent Life Stories for Believable Virtual Agents." Intelligent Virtual Agents 8th International Conference, IVA 2008. Tokyo, Japan, September 2008: Springer, 2008. 59-72.

[16] HTML 5 - "A vocabulary and associated APIs for HTML and XHTML". Editor's Draft 29 July 2009. Latest Published Version: http://www.w3.org/TR/html5/