Scalable Multi-modal Avatar Interface for Multi-user Environments

Mark Dunne Brian Mac Namee John Kelleher
DIT Kevin Street, Dublin 2, Ireland
'firstname'.*'lastname'*@dit.ie
LOK8.dit.ie

Abstract
This research outlines an Intelligent Virtual Agent (IVA) interface, where multiple users will be able to interact with 3D avatars. This will take place in a distributed multi-modal environment where the LOK8 Avatar System (AS) will need to locate it's users from a crowd, using face tracking and novel 3D animation techniques.

Keywords: intelligent virtual agents, computer animation, face tracking, user immersion, agent transitions, Mona Lisa effect

1 Introduction

The field of research known as Intelligent Virtual Agents (IVA) covers a variety of disciplines from areas such as *emotions* and *personality* to realistic *human figure* and *facial* animation [1]. The proposed system in this research is focused primarily on two disciplines, *Perception* and *Distributed Simulation*, concentrating on the ability of the agent's 3D avatar to transition across 2D display modalities while maintaining a high level of immersion for the user. The overall design will be a scalable IVA interface system that will cope with multiple users within crowds, across a distributed multi-modal environment where users will be on the move. **Contribution:** This paper outlines agent transitions between 2D display modalities and proposes three animation techniques that will help maintain the level of immersion the user experiences while these transitions occur. These techniques will be implemented within the proposed LOK8 Test Environment (TE). **Overview:** In Section 2 background work is discussed in relation to this research, Section 3 outlines our approach specifying two main areas, Section 3.1 on Agents Transitions and Section 3.2 on User Immersion. Section 4 details the proposed system TE . Section 4.2 outlines the proposed evaluation process. Finally concluding with Section 5.

2 Related Work

There has been much work done to boost the level of realism in IVAs, such as animating realistic humanoid models [2, 3], adding human-like features such as emotions [2] and personality [4, 5], and even incorporating memory into agents [6, 7]. All these contribute to building believable IVAs and as a result create the *"illusion of life"* [8]. However, there has been less research on how these IVAs will integrate into users day-to-day lives in large crowded environments. Most research focuses on a single end user (e.g. [9, 10, 3]), considering them as a participant in controlled experimental type scenarios and not a real-life situation, where there are likely to be multiple users of the system in a crowded multi-modal environment. Some research has focused on agent transitions, where the agent is completely transported from one platform to the next, adapting and harnessing each platform's functionality as necessary [11]. Other research has used face tracking to create

realistic eye-contact in 3D video conferencing, increasing immersion for the user [12]. In this research, elements of both [11, 12] will be used to form a novel approach for a scalable avatar interface on 2D display modalities.

3 Our Approach

The LOK8 Avatar System (AS) will allow agents to transition their 3D avatar between display modalities, following their user in the real world environment and give the illusion that the display modalities are windows peering into the 3D avatar's virtual world, this is discussed further in the next section. This means the user's personal agent will be walking virtually side-by-side with them and as a result will allow for seamless interaction between human and agent. The three techniques discussed in Section 3.2 should increase the level of immersion experienced by the user.

3.1 Agent Transitions

In this use-case scenario the LOK8 avatar will accompany the user through a large-scale environment, having many ways to transition from one display modality to another, while still interacting with their user. One of the key questions that this research addresses is how transitions will be initiated, manually by the user gesturing or automatically by the LOK8 AS. As the variation in types of agent transitions can become quite complex, three key transition groups have been identified;

- **Group A:** Wall Mounted Display-to-Wall Mounted Display Transitions.
- **Group B:** Mobile Device-to-Wall Mounted Display Transitions.
- **Group C:** Wall Mounted Display-to-Mobile Device Transitions.

This decomposition helps distinguish any issues relating to particular display modalities from those issues pertaining to the transitioning of the 3D avatars in general, while conducting research on the best ways to carry out these transitions.

3.2 User Immersion

The problem with 2D visual modalities is that when displaying a 3D avatar, the user needs to stand in the optimal viewing position for the 2D display to get the full 3D immersive effect. Any deviations from the optimal viewing position will cause a dramatic loss in immersion for the user. The image becomes skewed on the screen from the user perspective. In order to maintain the immersive 3D experience with the 3D avatar while a user is moving around the TE or viewing the 2D display at a non-optimal viewing angle, three techniques have been devised, using face tracking data detailed in Section 4;

- **Stretching:** Projection of 3D avatar is stretched to compensate for user's viewing angle.
- **Twisting:** Using the *"Mona Lisa effect"* [13] the 3D avatar's head is stationary but the body twists to give the illusion of a head movement towards the user.
- **Boxing:** The 3D avatar is placed in a box which extends back into the 2D display, giving the illusion of standing behind a window pane that moves according to viewing angle of the user.

In the next section a description of the TE will be outlined in detail and how it will be used during evaluations.

4 Experimental System Design

4.1 LOK8 Test Environment

The overall system design will allow for scalability and will accommodate multiple users. Multiple experiments will be carried out within a single TE; much like [14] who's software architecture consisted of a 3D graphics application which was developed using the Unity 3D game engine [15]. The AS will project multiple 3D avatars across multiple 2D display modalities (see Section 3.1) harnessed by the AS at any given time depending on each of the 3D avatars user's needs and location within the TE. The AS will use face tracking to help the 3D avatars engage more realistically with users by

using human-like eye gaze [12]. FaceAPI software [16] will be used to extrapolate six degrees of freedom from the video stream captured on each active web camera. Face recognition software [17] will also be used so that each 3D avatar can follow their user around the TE, vitally important to allow the system to cope with multiple users and distinguish specific users from a crowd. The face tracking data is used directly by the three techniques outlined in Section 3.2 and during the experiments the users will be confronted with 2D displays that have the three techniques switched on and others that don't. This will allowed the users to make side by side comparisons of the 3D immersive effects and they will also be able to make more informed decisions as to whether the techniques actually do increase the sense of immersion felt by users during interaction with the 3D avatars. The proposed evaluation scheme is detailed in the next section.

4.2 Evaluation Process

Historically IVA systems are difficult to evaluate due to the level subjectivity involved, standard evaluation schemes have been put forth in the past but have not been widely adopted, there has been a strong argument [18] for agents to be evaluated under a five category standard [19] and accordingly, the LOK8 AS will be evaluated under these five categories;

- Believability
- Social Interface
- Application Domains
- Agency
- Computational Issues and Production

The proposed evaluation scheme here will adopt the *"paired comparison"* approach [20], used in [21]. The basic idea is instead of explicitly rating visual stimuli, users are asked to select the perceptually better one of two visual stimuli which alleviates the user from making a forced, inaccurate perception decision. Bearing this in mind, any survey questions will be designed to incorporate the *"paired comparison"* approach. Another evaluation scheme that will be adopted is the PARADISE framework [22], predominantly used in the evaluation of dialogue systems, here it could be used to compare two 3D avatar interfaces especially where Wizard-of-Oz (WOZ) experimental set-ups are used.

5 Conclusions & Future Work

This paper highlights the issues relating to agent transitions and the loss of user immersion with 3D avatars on 2D displays. The integration of three techniques (stretching, twisting and boxing) into an IVA interface is proposed to increase user immersion. The future work involves evaluating these techniques in a scalable multi-modal and multi-user environment.

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] J. Gratch, J. Rickel, E. Andre, J. Cassell, E. Petajan, and N. Badler. Creating interactive virtual humans: some assembly required. *IEEE Intelligent Systems*, 17(4):54–63, July 2002.

[2] Matthieu Courgeon and Jean-claude Martin. Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 201–214, Amsterdam, The Netherlands, 2009. Springer Verlag.

[3] Patrick Gebhard, Marc Schroder, Marcela Charfuelan, Christoph Endres, Michael Kipp, Sathish Pammi, and Martin Rumpler. IDEAS4Games: Building Expressive Virtual Characters for Computer Games. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents*, volume vol. Springer, 2008.

[4] M.Y. Lim and R. Aylett. Feel the Difference: A Guide with Attitude! *Lecture Notes in Computer Science*, 4722:317, 2007.

[5] Michael Rushforth, Sudeep Gandhe, Ron Artstein, Antonio Roque, S. Ali, Nicolle Whitman, and David Traum. Varying personality in spoken dialogue with a virtual human. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, page 542, Amsterdam, The Netherlands, 2009. Springer Verlag.

[6] M.Y. Lim, R. Aylett, W.C. Ho, S. Enz, and P. Vargas. A Socially-Aware Memory for Companion Agents. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, page 26, Amsterdam, The Netherlands, 2009. Springer Verlag.

[7] W. Ho and K. Dautenhahn. Towards a Narrative Mind: The Creation of Coherent Life Stories for Believable Virtual Agents. In *Lecture Notes in Computer Science*, volume 5208, page 5972. Springer, 2008.

[8] F. Thomas and O. Johnston. *The illusion of life: Disney animation*. Disney Editions, rev sub edition, 1995.

[9] Dusan Jan, Antonio Roque, Anton Leuski, Jacki Morie, and David Traum. A Virtual Tour Guide for Virtual Worlds. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 372–378, Amsterdam, The Netherlands, 2009. Springer Verlag.

[10] Christopher Mumme, Niels Pinkwart, and Frank Loll. Design and Implementation of a Virtual Salesclerk. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, page 379, Amsterdam, The Netherlands, 2009. Springer Verlag.

[11] B.R. Duffy, G.M.P. O'Hare, a.N. Martin, J.F. Bradley, and B. Schon. Agent chameleons: agent minds and bodies. pages 118–125. Ieee, May 2003.

[12] Andrew Jones, Magnus Lang, Graham Fyffe, Xueming Yu, Jay Busch, Ian McDowall, Mark Bolas, and Paul Debevec. Achieving eye contact in a one-to-many 3D video teleconferencing system. volume 28. ACM, 2009.

[13] Michael Kipp and Patrickl Gebhard. IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *Lecture Notes in Computer Science*, pages 1–8. Springer, 2008.

[14] M.Y. Lim and R. Aylett. Feel the Difference: A Guide with Attitude! *Lecture Notes in Computer Science*, 4722:317, 2007.

[15] Unity 3D. http://unity3d.com/unity/, April 2010.

[16] faceAPI, Seeing Machines. http://www.seeingmachines.com/product/faceapi/, April 2010.

[17] Real Time Face Tracking and Recognition (RTFTR), Open Source Project. http://rtftr.sourceforge.net/, April 2010.

[18] Brian Mac Namee and Mark Dunne. Widening the Evaluation Net. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 525–526, Amsterdam, The Netherlands, 2009. Springer Verlag.

[19] Katherine Isbister and Patrick Doyle. Design and evaluation of embodied conversational agents: A proposed taxonomy. In *The First International Joint Conference on Autonomous Agents & Multi-Agent Systems*. Citeseer, 2002.

[20] Patrick Ledda, Alan Chalmers, Tom Troscianko, and Helge Seetzen. Evaluation of tone mapping operators using a High Dynamic Range display. *ACM Transactions on Graphics*, 24(3):640, July 2005.

[21] Xiaohan Ma and Zhigang Deng. Natural Eye Motion Synthesis by Modeling Gaze-Head Coupling. *2009 IEEE Virtual Reality Conference*, pages 143–150, March 2009.

[22] Melita Hajdinjak and France Mihelič. The PARADISE Evaluation Framework: Issues and Findings. *Computational Linguistics*, 32(2):263–272, June 2006.