TSB Technique: Increasing a User's Sense of Immersion with Intelligent Virtual Agents

[Extended Abstract]

Mark Dunne*
AI Group
DIT, Kevin Street,
Dublin 2, Ireland.
Mark.Dunne@dit.ie

Brian Mac Namee†
AI Group,
DIT, Kevin Street,
Dublin 2, Ireland.
Brian.MacNamee@dit.ie

John Kelleher‡
AI Group,
DIT, Kevin Street,
Dublin 2, Ireland.
John.Kelleher@dit.ie

ABSTRACT

The field of Intelligent Virtual Agent (IVA) research depends heavily on *immersive* techniques when presenting virtual agents to end-users. This sense of immersion relies on the user believing the agent to be real and present in their environment, creating the "*Illusion of Life*" [21]. This poster describes the on-going research into using a combination of three rendering techniques, *Twisting*, *Stretching* and *Boxing* (TSB), to create a fully immersive 3D illusion for an end-user from any viewpoint, as they move freely in front of displays distributed across large populated environments. The novel approach outlined in this poster, uses head tracking, face detection and the TSB technique to increase the user's sense of immersion and subsequently their sense of the agent's presence within the environment. Using only web-cameras, our approach is a hardware-light solution which does not require the end-user to wear any additional apparatus, such as the LED headset used in [10]. The poster primarily discusses our *Approach*, the *Preliminary Experiments & Results*, *Proposed Evaluation Process* and *Future Work*.

Keywords

Immersion, Intelligent Virtual Agents, Avatars

1. INTRODUCTION

Use of virtual agents as visual computer interfaces contributes positively to human-computer interactions [9]. However, badly designed virtual agents run the risk of becoming annoying, for example Microsoft's infamous *Mr. Clippy* [23]. Virtual agents should enrich a user's experience, not frustrate it.

*www.MarkDunne.me/

†www.comp.dit.ie/bmacnamee/

‡www.comp.dit.ie/jkelleher/

AICS 2010 NUI Galway, Ireland

The interaction style between the user and the virtual agent should be intuitive and natural, contributing to the overall immersive experience, where the user is completely engrossed in their environment. This sense of immersion relies on the user believing the agent to be real and present in their environment, creating the "*Illusion of Life*" [21]. A dramatic loss of immersion can be felt by a user if they sense the virtual agent is not engaging with them fully.

The standard approach to displaying virtual agents is to use 3D graphics with animation, visually represented on a 2D display. Maintaining an immersive 3D experience on a 2D display, however, requires the user to remain stationary in front of the display at an optimal viewing angle. This can cause issues, especially as it can restrict the user's movement and prevent a more natural interaction from occurring, which can lead to a loss of immersion.

This paper outlines the combination of three rendering techniques, *Twisting*, *Stretching* and *Boxing* (TSB), to create a 3D illusion for a user from any viewpoint (see Section 3.1). This novel approach of using head tracking and the TSB technique can increase the immersion felt by a user when they are interacting with a virtual agent on a 2D display. This hardware-light solution does not require the user to wear any additional apparatus, such as the LED headset used in [10]. This solution is also scalable to larger environments with multiple 2D displays and multiple users, making it ideal for use in public spaces.

Contribution: This paper outlines three rendering techniques that will help maintain the level of immersion the user experiences. It also outlines how these techniques will be implemented and evaluated within the proposed LOK8[1] Test Environment (TE), see Figure 5.

Overview: In Section 2 background work is discussed in relation to this research, Section 3 outlines our approach, detailing the TSB technique in Section 3.1. Section 4 details the proposed TE (Section 4.1) and the Twisting Experiment (Section 4.2). Section 6.1 outlines the proposed evaluation process and outlines the preliminary results from the Twisting Experiment in Section 4.2. Finally concluding in Section 5 with a brief dicussion about this research and future work.

2. RELATED WORK

There has been much work done to boost the level of realism in IVAs, such as animating realistic humanoid models

[1]LOK8 (Locate) Project: http://lok8.dit.ie/

[1, 2], adding human-like features such as emotions [1] and personality [12, 19], and even incorporating memory into agents [14, 4]. All these contribute to building believable IVAs and as a result create the "*illusion of life*" [21].

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. [6, 17, 2]), 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.

The research outlined in [10] required participants to sit in front of a large display and interact with the agent on the display in an interview type scenario. In order to track the participants head motion, they were required to wear a head set with two infra-red (IR) LEDs attached, the motion data of the LEDs is captured by the IR camera in the Wiimote[2] and is then used to create the "*Mona Lisa effect*" (see Section 3.1 Twisting).

3. OUR APPROACH

From the perspective of creating an immersive 3D user experience, one problem with 2D displays is that when a 3D avatar is been displayed, the user needs to stand in the optimal viewing position (perpendicular to the 2D display, see 'User' in Figure 1) in front of the 2D display in order 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's perspective, a process called *lateral foreshortening*. It is that loss of immersion this research is addressing and in order to maintain the immersive 3D experience with the 3D avatar on 2D displays, the TSB technique (see Section 3.1) has been devised, using face tracking data detailed in Section 4.1;

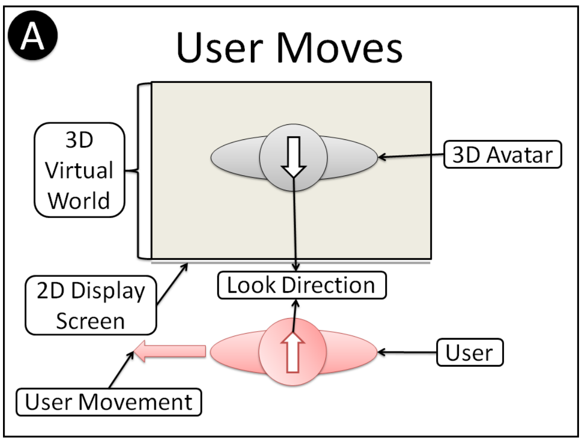

Figure 1: A. User moves in front of the 2D display to an non-optimal viewing angle.

3.1 TSB Technique

In order to maintain a 3D immersive experience on a 2D display, while a user is moving (see Figure 1) or viewing the 2D display at a non-optimal angle, a combination of the three following techniques has been devised, using face tracking data (see Section 4.1);

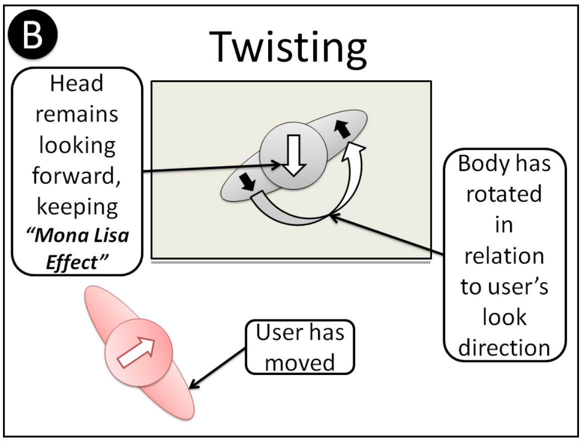

Figure 2: B. Twisting

***Twisting*:** The virtual agent's body is twisted at the opposite angle to the user's viewing angle, in order to create the illusion of the virtual agent's head following the user. This makes the most of the "*Mona Lisa effect*" [10], where the virtual agent looks directly forward, seemingly maintaining eye contact regardless of the user's viewing angle. (see Figure 2).

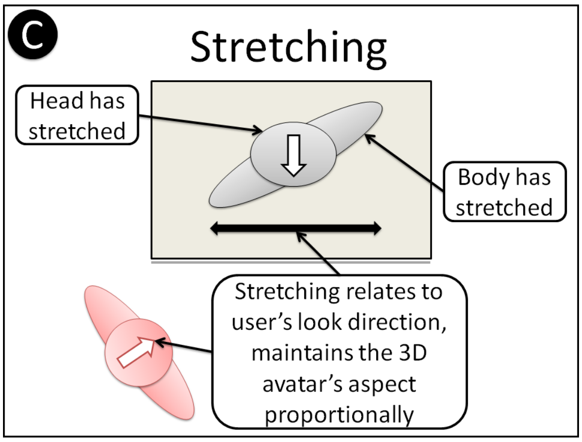

Figure 3: C. Stretching

***Stretching*:** This technique involves stretching the image of the virtual agent on the 2D display according to the user's viewing angle. This halts any skewing or narrowing of the virtual agent's 3D image, which would otherwise break the 3D illusion. Similar in effect to [8]. (see Figure 3).

***Boxing*:** The virtual agent is placed in a virtual box (room), that the user views through a virtual window, for example [24]. The user's view of the room will change according to their viewing angle acting just like a real-world window. (see Figure 4).

[2]The Wii Remote unofficially nicknamed "*Wiimote*", is the primary controller for Nintendo's Wii console.

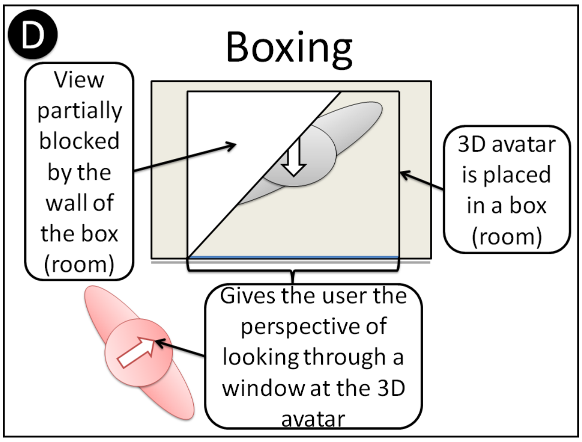

Figure 4: D. Boxing

4. PRELIMINARY EXPERIMENTS

4.1 LOK8 Test Environment

The proposed TE will allow for scalability and will accommodate multiple users. Multiple experiments will be carried out within the TE (see Figure 5); much like [13] who's software architecture consisted of a 3D graphics application which was developed using the Unity 3D game engine [22]. The Avatar System (AS) will project multiple 3D avatars across multiple 2D displays at any given time depending on each of the users needs and location within the TE.

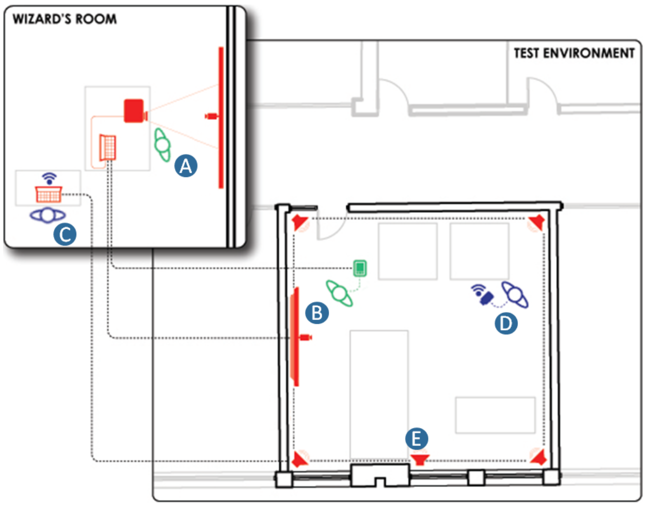

Figure 5: LOK8 Test Environment: *A.* User interacting with an avatar on projectors with camera attached; *B.* User with iPhone, avatar is on LCD TV with camera attached; *C.* Controller runs the experiment and monitors users; *D.* Experiment observer videos the interactions and interviews users; *E.* Auditory equipment.

The AS will use face tracking to help the 3D avatars engage more realistically with users by using human-like eye gaze [7]. FaceAPI software [20] will be used to extrapolate six degrees of freedom from the video stream captured on each active web camera, which will be placed over each 2D display in the TE. Face recognition software [18] will also be used so that each 3D avatar can follow their user around the TE, this is important as it helps the AS cope with multiple users and distinguish specific users from a crowd. The face tracking data is used directly by the TSB technique outlined in Section 3.1 and during the experiments the participants will be confronted with 2D displays that have the TSB technique switched on and others that have one or none of the techniques switched on (See Section 4.2 Twisting only experiment). This will allow the users to make side by side comparisons ("*Paired Comparison*" details Section 6.1.2) 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 Section 6.1.

4.2 Twisting Experiment

This preliminary experiment was carried out on 8 participants ranging in age from 20 years old to 59 years old. There was 7 males and 1 female participants. During the experiment each participant sat on a chair with wheels (see Figure 6), so that they could freely move in front of the displays while remaining at eye level with the displays as they were on a regular height desk. The participants were giving a brief introduction to the setup and told the rules of the game they were about to play.

The game was based on the "*Simon Says*" game typically played by children, where players must follow all instructions issued by the game master, in this case the 3D agent, as along as the game master speaks the words "*Simon Says*" before each instruction. If the game master does not say "*Simon Says*" before an instruction and a player carries out the instruction they loose the game. To beat the game in this preliminary experiment, the player had to follow five correct instruction in a row.

The instructions in this experiment were asking the participant to touch a letter (*A*, *B*, *C* or *D*) on the desk in front of them, for example "*Simon says, touch the letter 'A'!*" (see Figure 6). The participant would then touch the correct letter, all the time the agent would be projected onto two displays; the *Left* 2D display showing the agent with the *Twisting* technique outlined in Section 3.1 on and the *Right* 2D display showing the agent without any techniques switched on ("*Paired Comparison*" evaluation, see Section 6.1.2).

The participants were all asked to complete a survey after they played the game, this survey was made up of three questions which were as follows (see Section 4.2.1 for the results);

- **Question 1: Did you enjoy the "Simon Says" game** (*Answer: Yes/No*)**?** This question establishes the participants level of enjoyment, this helps gauge the rest of the participants responses and may require a follow up question to explain a negative response.

- **Question 2: Which screen did you feel displayed a more engaging agent** (*Answer: Left/Right*)**?** This question is where we are trying to prove the hypothesis, the expectation is for participants to favour the *Left* 2D display as it has the *Twisting* technique switched on, which should increase their sense of immersion. This

also helps gauge the *Believability* of the agent (see Section 6.1.1).

- **Question 3: In relation to computer games do you understand the meaning of the word immersion** (*Answer: Yes/No*)**?** This last question is to gauge the participants understanding of the experiment, to see if they actually know what is being tested. As prior knowledge of the area or experiment may taint the results and force a specific outcome.

Figure 6: The participant gets a verbal instruction from the 3D agent being displayed on the two displays, "*Simon says, touch A!*" The participant then moves his hands towards '*A*' while looking at the displays; Left display has the Twisting technique switched on; Right display is the control.

4.2.1 Twisting Experiment Preliminary Evaluation

Table 1: Results of Survey (see Section 4.2)

Question	Yes or Left	No or Right
1	*7*	*1*
2	*8*	*0*
3	*4*	*4*

In Table 1 we see the results of the survey carried out on the 8 participants during the preliminary experiment (see Section 4.2). The results for **Question 1** state that 87.5% of participants enjoyed the game, an encouraging result and an argument to continue to use games as a means of experimentation. The results for **Question 2** state that 100% of the participants preferred the sense of immersion they experience while interacting with the agent on the *Left* display, this goes some way to prove the hypothesis and encourages further research in this area. The results for **Question 3** state that 50/50 split of participants had some knowledge of the area of research, either from playing games or just being knowledgeable in that area. Also indicates that a diverse group of people were selected to participate in this preliminary experiment.

5. CONCLUSIONS

This paper outlines the combination of three rendering techniques, *Twisting*, *Stretching* and *Boxing* (TSB), to create a 3D illusion for a user from any viewpoint (see Section 3.1). The results of a preliminary experiment (see Section 4.2.1) indicate that participants did actually feel a higher level of immersion when just one of the techniques, *Twisting*, was turned on. This result encourages further research into the TSB technique.

The future work will involve evaluating fully the TSB technique (see Section 3.1) in the LOK8 TE (see Section 4.1) using all the proposed evaluation methods outlined in Section 6.1. This will help establish if the TSB technique can maintain a high level of immersion for a user, while interacting with a 3D avatar on a 2D display.

6. FUTURE WORK

The future work will involve evaluating fully the TSB technique (see Section 3.1) in the LOK8 TE (see Section 4.1) using all the proposed evaluation methods outlined in Section 6.1. This will help establish if the TSB technique can maintain a high level of immersion for a user, while interacting with a 3D avatar on a 2D display.

6.1 Proposed Evaluation Process

6.1.1 5 Category Evaluation

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 [16] for agents to be evaluated under a five category standard [5] and accordingly, the LOK8 AS will be evaluated under these five categories;

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

6.1.2 Paired Comparison

The proposed evaluation scheme here will adopt the "*paired comparison*" approach [11], used in [15]. 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.

6.1.3 PARADISE Framework

Another evaluation scheme that will be adopted is the PARADISE framework [3], 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.

Acknowledgements

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

7. REFERENCES

[1] M. Courgeon and J.-c. 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.

[2] P. Gebhard, M. Schroder, M. Charfuelan, C. Endres, M. Kipp, S. Pammi, and M. Rumpler. IDEAS4Games: Building Expressive Virtual Characters for Computer Games. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents*, volume vol. Springer, 2008.

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

[4] 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, pages 59–72. Springer, 2008.

[5] K. Isbister and P. 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.

[6] D. Jan, A. Roque, A. Leuski, J. Morie, and D. 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.

[7] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, and P. Debevec. Achieving eye contact in a one-to-many 3D video teleconferencing system. volume 28. ACM, 2009.

[8] Julian Beever's Pavement Drawings. http://users.skynet.be/J.Beever, June 2010.

[9] G. Kempter, K. Weidmann, and P. Roux. What are the benefits of analogous communication in human computer interaction? In *C. Stephanidis (ed). Universal access in hci: inclusive design in the information society*, pages 1427–1431, 2003.

[10] M. Kipp and P. Gebhard. IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *IVA'08 Proceedings of the 8th International Conference on Intelligent Virtual Agents*, pages 191–199. Springer, 2008.

[11] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a High Dynamic Range display. *ACM Transactions on Graphics*, 24(3):640, July 2005.

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

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

[14] M. Lim, R. Aylett, W. 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.

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

[16] B. Mac Namee and M. 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.

[17] C. Mumme, N. Pinkwart, and F. 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.

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

[19] M. Rushforth, S. Gandhe, R. Artstein, A. Roque, S. Ali, N. Whitman, and D. 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.

[20] Seeing Machines FaceAPI. http://www.seeingmachines.com/product/faceapi/, June 2010.

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

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

[23] B. Whitworth. Polite computing. In *Behaviour & Information Technology*, volume 24, pages 353–363. Taylor & Francis, 2005.

[24] Winscape: Virtual Windows. http://www.rationalcraft.com/Winscape.html, June 2010.