Stereoscopic Avatar Interfaces: A study to determine what effect, if any, 3D technology has at increasing the interpretability of an avatar's gaze into the real world.

Mark Dunne, Brian Mac Namee, and John Kelleher

Applied Intelligence Research Centre
School of Computing, Dublin Institute of Technology
www.comp.dit.ie/aigroup
Mark.Dunne@dit.ie

Abstract. An approach to displaying avatar interfaces *monoscopically* such as the *Turning*, *Stretching* and *Boxing* (*TSB*) technique (a combination of three graphical processes) has been shown to improve the communication efficiency of avatars – by increasing a user's ability to interpret an avatar's gaze direction through the delivery of a sustained 3D illusion of the avatar on a standard 2D display. A reasonable question to ask about this approach is whether or not the improvement in interpretability can be matched or surpassed by using a standard 3D display technology (stereoscopic) with or without the fore-mentioned approach? This paper presents an experiment that pits the *TSB* technique against a standard 2D display using 3D technology in order to answer this question. An analysis of these results shows that, by itself, 3D display technology does not have any influence on the ability of viewers to accurately interpret an avatar's gaze direction. Furthermore, the combination of 3D display technology with the *TSB* technique does not result in any noticeable improvement.

1 Introduction

In an effort to improve human-machine interactions an interface for a user to interact with an agent is required, a 3D virtual avatar is more often than not used. However, the standard approach is to displaying virtual avatar interfaces is to render them onto 2D displays [1–4]. In general 2D displays require the user to remain in a stationary position at the optimal viewing angle of 90° to the display surface, commonly referred to as the *sweet spot* [5]. When viewing angles are more acute than 90°, there is a deterioration in the effectiveness of any 3D illusion due to skewing (or narrowing) of the rendered image from the user's perspective, which results in negative "*foreshortening*" distortions. These distortions damage a user's perceived presence of an avatar and reduce the sense of social richness, realism and immersion felt during interactions with an avatar [6]. Moreover, this distortion limits the viewer's ability to be able to read and

interpret the avatar's gaze towards points of interest (e.g. objects, places and people) in the viewer's real-world vicinity. The *Turning*, *Stretching* and *Boxing* (*TSB*) rendering technique [7] has been developed to address these issues and has been shown to be effective [7]. With the recent prevalence of 3D display technology, we must ask if it alone can combat these unwanted distortions, and so could eliminate the need for techniques like *TSB*.

Contribution. This paper presents the results of an experiment, that pits the *TSB* technique against 3D display technology. During this experiment participants had to guess which marker on the floor between them and the 2D display an avatar was looking at. The avatar only uses its gaze to direct attention towards specific markers on the floor. The results indicate that 3D technology by itself does not have any influence on the ability of participants to interpret which floor marker the avatar is looking at, meaning that a technique such as the *TSB* is still required to enable viewers to interpret the avatar's gaze.

1.1 The Turning, Stretching & Boxing (*TSB*) Technique

The TSB technique[1] is a combination of three image altering processes, that deliver a constant 3D illusion of a 3D avatar on a 2D display from a user's perspective. The maintained 3D illusion is similar to the user looking through a '*real-life*' window at the avatar. This means that when a user moves in front of the display, the 3D scene continuously updates to match the user's perspective. As a result of this the avatar is able to accurately reference points of interest (e.g., objects, places and people) in the user's real-world vicinity with just its gaze.

The *TSB* technique is dependent on constant awareness of the user's situation based on on an accurate measure of the user's head position, obtained using a Microsoft Kinect in conjunction with the Kinect for Windows SDK (more specifically using the the joint labelled JointType.Head from the Kinect skeletal tracking data). Based on the recorded position of the viewer's head relative to the display surface, the **Turning**, **Stretching** and **Boxing** processes are applied as follows:

- **Turning.** Simply a process that enables the avatar to direct its gaze towards the user to make realistic eye contact. It is achieved by having the avatar's head turn to direct its gaze down the virtual camera lens in the 3D scene in order to make and maintain eye contact with the user as they move. The virtual camera's position in the 3D scene is updated to match the user's head position in the real-world. So, when the avatar's gaze is directed towards the virtual camera it seems to be directed towards the user no matter where they move in front of the 2D display. This creates a well known illusion coined as the "*Mona Lisa* Effect" by Kipp et al. [2] because Leonardo da Vinci's painting of *Mona Lisa* is a very well know example of this illusion. The illusion, extensively studied by Koenderink et al. [8], becomes apparent

[1] *TSB Technique In Action*: http://youtu.be/OWDMGoDH640

as the gaze of the avatar seems to follow the viewer as they move in front of the 2D display, regardless of their viewing angle. Kipp et al. [2] harness the illusion in their research to increase the realism of their avatar's gaze behaviours. The ability to direct gaze towards users is a crucial form of referencing for social agents [9, 10] as it can be an indicator of the willingness of one social entity to engage in social interactions with another [11]. The addition of the next two processes (i.e., '*stretching*' and '*boxing*') means that the user is not restricted to the *sweet spot* and can move freely in front of the 2D display.

– **Stretching.** This process involves stretching the image of the avatar on the 2D display according to the user's viewing angle. This counteracts any negitive foreshortening of the avatar's 3D image by skewing or narrowing due to acute viewing angles that would otherwise break the 3D illusion and is similar in effect to the Responsive Workbench [12]. The *stretching* process does not effect the 3D models or any relating animations in the 3D scene. Rather, it merely adjusts the view-port of the 3D scene to reflect the user's viewing angle. This means all 3D models and animations remain intact.
– **Boxing.** This process places the 3D avatar in a virtual box that appears to adjoin the real-life room in which the user is present. The effect is to give the appearance of a window into another room, similar to Rational Craft's Winscape project[2]. The user's view of the room changes according to their viewing angle, acting just like a real-world window.

These three processes have been used in other research before [2, 12] and the combination of the three in the *TSB* technique has been shown to increase a user's ability to interpret the gaze direction of an avatar [7]. The *TSB* technique has also been shown to increase the perceived '*corporeal*' presence [13] of an avatar on a 2D display [7]. Although the *TSB* technique has proven benefits, it is unclear whether or not these same benefits could be attained, or even surpassed, by simply using a standard 3D display technology such as Anaglyph 3D (see Section 1.2) or if a combination of the *TSB* technique with a 3D display technology would be beneficial. These are the questions that this work addresses.

1.2 Anaglyph 3D

Anaglyph 3D is a stereoscopic effect achieved by requiring that the viewer wear special glasses (see Fig. 1) that place different coloured filters over each of the viewer's eyes, and that the image is rendered so that slightly different versions of the image are produced for the left and right eyes (see Fig. 2). The different coloured filters in the glasses, and matching colours in the rendering, mean that only the version of the image tailored for each of the viewer's eyes is visible to that eye and so a stereoscopic illusion is created.

Anaglyph 3D was chosen for these experiments as it can be easily used with a standard 2D display or projector with the addition relatively low cost glasses

[2] *Rational Craft's Winscape Project* website: http://www.rationalcraft.com/Winscape.html

Fig. 1. Anaglyph 3D glasses with filters in front of each eye: red filter covering the left eye and a cyan filter covering the right eye.

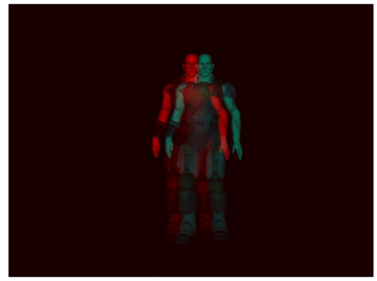

Fig. 2. 3D model rendered with the anagylph 3D stereoscopic effect.

(see Fig. 1). It is easily switched on for 3D applications when an *NVIDIA GeForce GPU* is installed along with the *NVIDIA 3D Vision Discover*[3] firmware. Anaglyph 3D is not a prefect solution as it can cause discomfort [14] and crossfeedback or ghosting [15] in some users. However, it still remains a complete and low cost solution to achieve stereoscopic 3D.

2 3D vs TSB Experiment

The purpose of the *3D vs TSB* experiment is to compare whether it is easier to interpret an avatar's gaze using (1) the *TSB* technique, (2) anaglyph 3D rendering, or (3) a combination of both *TSB* and anaglyph 3D rendering.

Procedure. Each experimental trial began with the participant standing in one of two fixed starting positions (see the green and red square starting positions in Fig. 3). The participants would start on the opposite square after every two trials, e.g., red, red, green, green, red, red, green, green,... and so on. A trial consisted of the avatar directing its gaze at the participant and then redirecting it to one of the 7 floor markers when the participant was ready. The participant had to determine which marker the avatar was looking at and, when ready, would indicate their choice by speaking aloud the number of that floor marker. When deciding on which marker the avatar was looking at, participants were allowed to move around freely in front of the display screen (whilst staying within the confines of the Kinect's field of view (57°) in a range of just under 4 metres referred to as the "*Player Free Move Area*" in Fig. 3 and is clearly marker by the white tape on the floor in Fig. 4), so as to best interpret the avatar's gaze. Once the participant's choice was recorded the participant would return to the relevant starting position for the next trial.

[3] *NVIDIA 3D Vision Discover*. Anaglyph 3D, although not a perfect solution, is a complete and low cost way to achieve stereoscopic 3D. Website: http://www.nvidia.co.uk/object/3D_Vision_Discover_Main_uk.html

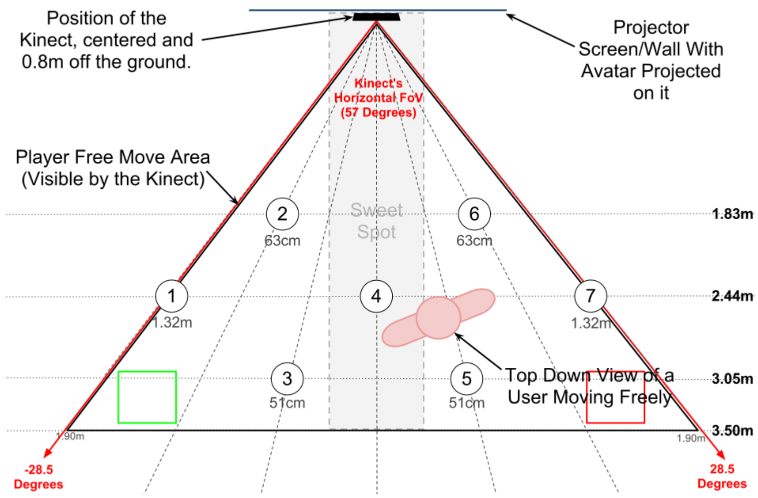

Fig. 3. A floor plan for the *3D vs TSB* experiment. Visible are the 7 floor markers, the *sweet spot*, the two starting positions (green and red squares) and the field of view of the Kinect sensor (57°).

Each participant performed 28 trials: the avatar would look at each of the 7 floor markers (see Fig. 3) once for each of the 4 experimental conditions which were as follows:

– **2D Control**: The avatar appears as it would on a regular 2D display, i.e., the rendering does not update to reflect a participant's viewing position.
– **3D Control**: The same as **2D Control** but with Anaglyph 3D switched on.
– **2D TSB**: The *TSB* technique is switched on, therefore the image of the 3D avatar is continuously updated to reflect the participant's perspective.
– **3D TSB**: The same as **2D TSB** above but with Anaglyph 3D switched on.

Learning effects were controlled for by varying the sequence of the 4 trial conditions across participants. The sequence in which the floor markers were gazed at by the avatar were also varied between participants. This design resulted in each participant carrying out one trial per floor marker per condition.

Participants. There were 7 participants in total (2 females, 5 males) with ages ranging from 25 to 58 years.

3 Results

Table 1 shows the average % of accuracy achieved by all 7 participants for each of the 7 floor markers across all 4 experimental conditions. A *single factor ANOVA* statistical test was carried out to test for a significant difference between the 4 experimental conditions, for each of the 7 floor markers. The results at the p <

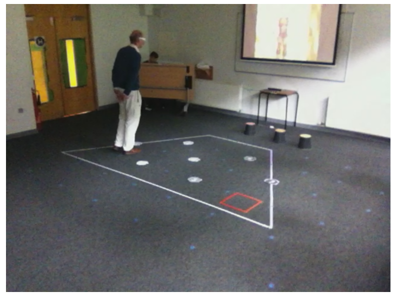

Fig. 4. A participant performing the *3D vs TSB* experiment wearing Anaglyph 3D glasses (the participant is standing behind floor marker **1**).

0.05 level [F(3, 24) = 6.97, p = 0.0016] show that a significant difference exists. This difference was further investigated through 5 posthoc testing – in this case 5 *paired two sample for means Student t-Tests* carried out on the results from Table 1 and are as follows:

1. **3D Control** (M=0.41, SD=0.28) and **2D TSB** (M=0.76, SD=0.14): t(6)=-3.61, p=0.01. A statistically significant difference between the 3D and *TSB* results exists, with the latter out-performing the 3D condition. The test was done to compare 3D technology on a 2D display with the *TSB* technique on the same 2D display but without 3D technology turned on.
2. **3D Control** (M=0.41, SD=0.28) and **2D Control** (M=0.38, SD=0.21): t(6)=0.22, p=0.83. The *Control* condition shows no statistically significant difference with 3D technology turned on and off.
3. **3D TSB** (M=0.73, SD=0.15) and **2D TSB** (M=0.76, SD=0.14): t(6)=-0.57, p=0.59. When the *TSB* technique is used there is no statistically significant difference with 3D technology turned on and off.
4. **3D TSB** (M=0.73, SD=0.15) and **3D Control** (M=0.41, SD=0.28): t(6)=-2.64, p=0.04. When the 3D technology is turned on with both the *TSB* technique and the *Control*, a statistically significant difference exists between them.
5. **2D TSB** (M=0.76, SD=0.14) and **2D Control** (M=0.38, SD=0.21): t(6)=-4.61, p=3.6×10^{-3}. Likewise when the 3D technology is turned off, there is a statistically significant difference between the *TSB* technique and the *Control*.

The results of the 1^{st} t-test (**3D Control** vs **2D TSB**) indicate that the accuracy of gaze interpretations made by participants when the *TSB* technique is used on a standard 2D display is significantly better than when 3D display technology is used without the *TSB* technique. However, the results of the 2^{nd} t-test show no significant difference between the results for **3D Control** and **2D Control** conditions. This pattern is repeated in the 3^{rd} t-test when the comparison of the **3D TSB** and **2D TSB** conditions show no significant difference. This outcome indicates that 3D technology alone does not eliminate the need

Table 1. The average accuracy (% of correct interpretations of the avatar's gaze) achieved by the 7 participants for each of the 7 floor markers (see Fig. 3) over the 4 experimental conditions. The overall average accuracy for each condition is also shown.

Floor Marker	**2D Control**	**2D TSB**	**3D Control**	**3D TSB**
1	57.14%	75%	85.71%	57.14%
2	25%	57.14%	28.57%	57.14%
3	42.86%	62.50%	14.29%	57.14%
4	59.82%	79.46%	28.57%	85.71%
5	12.50%	71.43%	14.29%	85.71%
6	57.14%	100%	42.86%	85.71%
7	12.50%	85.71%	71.43%	85.71%
Average	**38.14%**	**75.89%**	**40.82%**	**73.47%**

for a technique like *TSB* when the avatar needs to be able to direct its attention to points of interest in the viewer's real-world vicinity. Furthermore, the addition of 3D display technology does not improve the effect achieved using the *TSB* technique, which clearly enables viewers to reliably interpret avatar gaze (indicated by the results of the 4^{th} and 5^{th} t-tests).

4 Discussion & Future Work

In terms of displaying avatars to users in 2D displays, 3D technology does not eliminate image distortion from occurring when users are viewing the avatar from non-optimal viewing angles. It at most only enhances the way that the viewer perceives the image of the avatar – the avatar appears stereoscopically rather than *monoscopically*. In fact it is important to note that 3D technology still requires the viewer to be viewing the 2D display from within the *sweet spot* for the best results. In order to allow for a better means of communication between avatar and viewer on a 2D display for this specific experimental set-up, a technique like the *TSB* technique is still needed.

The results from the *3D vs TSB* experiment when using 3D technology showed no significant effect on a participant's ability to accurately interpret an avatar's gaze. More specifically, when the avatar's gaze is being directed at a point of interest in the viewer's real-world vicinity the *TSB* technique preforms just as well with or without the 3D technology switched on. For this specific experimental set-up there seems to have been no obvious benefit to use 3D technology.

That being said, there is a very obvious visual benefit to using 3D technology; the avatar appears stereoscopically and this should have an effect on the user's perceived presence for the avatar. The stereoscopic image creates a more realistic looking avatar and considering *realism* is an important part of presence [6],

3D technology has to be still considered useful in the bigger picture. However, making users wear uncomfortable 3D glasses in order to interact casually with an avatar could potentially negate any positive effects it brings in the first instance.

Considering the insignificant difference between the results of the *TSB* conditions (*2D TSB* and *3D TSB*), with and without 3D technology being switched on. A user based study with a clear emphasis on evaluating the level of perceived presence for an avatar for both *TSB* conditions is a good next step. Keeping in mind the many difficulties in evaluating presence due to its very subjective nature and the inadequacies of using surveys to gather qualitative data outlined by Holz [13] in this regard, a very robust experimental design will be required.

References

1. Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y.: MACK: Media lab Autonomous Conversational Kiosk. In: Imagina. (2002)
2. Kipp, M., Gebhard, P.: IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In: IVA, Tokyo, Japan, Springer (2008) 191–199
3. Wiendl, V., Dorfműller-Ulhaas, K., Schulz, N., André, E.: Integrating a Virtual Agent into the Real World: The Virtual Anatomy Assistant Ritchie. In: IVA. Volume 4722. (2007) 211–224
4. Kruppa, M., Spassova, L., Schmitz, M.: The Virtual Room Inhabitant - Intuitive Interaction with Intelligent Environments. In Zhang, S., Jarvis, R., eds.: Australian Conference on Artificial Intelligence. Volume 3809. (2005) 225–234
5. Raskar, R., Brown, M.S., Yang, R., Chen, W.C., Welch, G., Towles, H., Scales, B., Fuchs, H.: Multi-Projector Displays Using Camera-Based Registration. In: IEEE Visualization, IEEE (1999) 161–168
6. Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. Journal of Computer- Mediated Communication **3**(2) (1997)
7. Dunne, M., Mac Namee, B., Kelleher, J.: The Turning, Stretching and Boxing Technique: A Step in the Right Direction. In: IVA. Volume 7502., Santa Cruz, Ca, USA, Springer (2012) 363–370
8. Koenderink, J.J., Van Doom, A.J., Kappers, A.M.L., Todd, J.T.: Pointing out of the picture. Perception **33**(5) (November 2004) 513–530
9. Lance, B., Marsella, S., Koizumi, D.: Towards expressive gaze manner in embodied virtual agents. In: AAMAS Workshop on Empathic Agents. (2004) 194–201
10. Lee, J., Marsella, S., Traum, D.R., Gratch, J., Lance, B.: The rickel gaze model: A window on the mind of a virtual human. In: IVA. (2007) 296–303
11. Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., Poggi, I.: A model of attention and interest using gaze behavior. In: IVA. (2005) 229–240
12. Agrawala, M., Beers, A.C., McDowall, I., Fŕohlich, B., Bolas, M., Hanrahan, P.: The two-user Responsive Workbench: support for collaboration through individual views of a shared space. In: SIGGRAPH. (1997) 327–332
13. Holz, T., Campbell, A.G., Ó'Hare, G.M.P., Stafford, J.W., Martin, A., Dragone, M.: MiRA – Mixed Reality Agents. International Journal of Human-Computer Studies **69**(4) (2011) 251–268
14. He, S., Zhang, T., Doyen, D.: Visual discomfort prediction for stereo contents. Volume 7863., SPIE (2011) 78631X
15. Woods, A.J., Rourke, T.: Ghosting in Anaglyphic Stereoscopic Images (February 27 2004)