

Face Reality: Investigating the Uncanny Valley for virtual faces

Rachel McDonnell*
Trinity College Dublin

Martin Breidt†
Max Planck Institute for Biological Cybernetics



Figure 1: Our virtual model rendered in different visual styles: (left) High Quality, (middle) Game Quality, (right) NPR.

1 Introduction

The Uncanny Valley (UV) has become a standard term for the theory that near-photorealistic virtual humans often appear unintentionally eerie or creepy. This UV theory was first hypothesized by robotics professor Masahiro Mori in the 1970’s [Mori 1970] but is still taken seriously today by movie and game developers as it can stop audiences feeling emotionally engaged in their stories or games. It has been speculated that this is due to audiences feeling a lack of empathy towards the characters. With the increase in popularity of interactive drama video games (such as *L.A. Noire* or *Heavy Rain*), delivering realistic conversing virtual characters has now become very important in the real-time domain. Video game rendering techniques have advanced to a very high quality; however, most games still use linear blend skinning due to the speed of computation. This causes a mismatch between the realism of the appearance and animation, which can result in an uncanny character. Many game developers opt for a stylised rendering (such as cel-shading) to avoid the uncanny effect [Thompson 2004]. In this preliminary work, we begin to study the complex interaction between rendering style and perceived trust, in order to provide guidelines for developers for creating plausible virtual characters.

It has been shown that certain psychological responses, including emotional arousal, are commonly generated by deceptive situations [DePaulo et al. 2003]. Therefore, we used deception as a basis for our experiments to investigate the UV theory. We hypothesised that deception ratings would correspond to empathy, and that highly realistic characters would be rated as more deceptive than stylised ones.

2 Experiment Design & Stimuli Creation

Firstly, we recorded a series of truths and lies from an actor which were then applied onto a virtual model and rendered at three different qualities. In a perceptual experiment, participants viewed these sequences and were asked to indicate whether the character

was lying or telling the truth in each trial. We hypothesised that participants would be the least trustful of the most realistically rendered character due to the mismatch between high visual quality and bone-based animation.

Two actors participated in the recording session. One took the role of the “interviewer” whose voice alone was recorded. The second was the “interviewee” whose voice, face, body and eye movements were captured. We included eye-capture in our dataset as Step-toe et al. [2010] found that the addition of eye movement increases participant accuracy in detecting truth and deception when viewing virtual avatars. Both actors were non-professionals but accustomed to the motion capture setup and environment. A series of questions were asked at random by the interviewer. The interviewee was told in advance whether to lie or tell the truth to the question. The answers were not rehearsed to ensure a natural reaction.

A facial scan of the interviewee was taken using a structured light 3D scanner (ABW). This scan was used, along with a series of photographs taken from a range of angles by an artist to create a virtual model of the actor (Figure 2). The virtual model was a typical “next gen” game character, with both facial and body rigs, and high quality (2048 × 2048) diffuse, opacity and normal-map textures.

Motion capture was conducted using a 13 camera Vicon optical system, where 52 markers were placed on the body and 36 markers on the face. A head-mounted eye-tracking device (Eyelink 2) was used to capture the movements of the left eye relative to the head at 250Hz. Two microphones were placed near to the actors and recorded their voices on two separate tracks. The body motion (captured at 120Hz) was mapped onto a skeleton, where joint angles were computed and used to drive the virtual character in Autodesk 3ds Max. The facial motion was directly exported as 3D marker motion, and the facial bones of the character were constrained to these markers to produce the animation. Finally, the eye rotations were computed and applied directly on to the bones driving the eye balls in 3ds Max. Figure 3 shows examples of some of the expressions created by the face and eye motion capture.

The model was rendered in three different styles. Raytracing in mental ray was used for the first “High Quality” style (*HQ*). This included high quality skin and eye shaders with physically accu-

*e-mail:ramcdonn@cs.tcd.ie

†e-mail:martin.breidt@tuebingen.mpg.de

rate reflections and shadows from area lights with indirect illumination (Figure 1, left). HQ took 96 seconds per frame to render on a 2.4GHz AMD Opteron 275 dual CPU system. The second was “Game Quality” (*Game*) where typical game-style Phong shading with point light sources was used which produced no shadows or reflections (Figure 1, middle). Game took only 2 seconds to render per frame on the same CPU. Finally, we used a non-photorealistic rendering style (*NPR*) which created a more cartoon-like look, with outlines and flat shading (Figure 1, right). The NPR render was included as stylisation is often used by game developers to avoid the UV. This style took 3 seconds to render per frame. We ensured that the illumination in all conditions was similar in direction and intensity, to avoid emotional illumination bias. Fifty-four movies were created in total (18 sequences (9 truths, 9 lies) \times 3 rendering styles). The character was viewed from the shoulders up and facing the participant, with a grey gradient background (Figure 1).



Figure 2: (left) Photograph of actor; (middle) Virtual character; (right) Underlying geometry of virtual model.

3 Experiment

Stimuli were displayed on a 24” LCD monitor at a distance of 60 cm and participants used headphones to listen to the audio. Movies were displayed at 800×600 at 30 frames/sec, using lossless compression. Twenty-three volunteers took part in this experiment (16 male, 7 female). All were naïve to the purpose of the experiment and from different educational backgrounds. University ethical approval was granted for the experiment, and participants received a book voucher to compensate for their time.

Participants viewed 108 movies (18 sequences (9 truths, 9 lies) \times 3 render styles \times 2 repetitions) in random order. They were asked after every movie to indicate whether they thought that the character had just told a *lie* or the *truth*, using a right or left mouse-click. In order to avoid any bias towards the right or the left, we randomly assigned the mouse to the answer for each participant (and indicated the button order on the screen). We also asked participants for a confidence rating, to indicate how confident they were in their answer, on a scale from 1–6, where 1 indicated that they have very little confidence, and 6 indicated that they were very confident. No feedback was given to participants to indicate if they were correct.

On completing the experiment, participants filled a questionnaire for each render style to indicate on a scale of 1–5: the overall quality of the virtual character, and the quality of the eye and facial animation. They also indicated how friendly and trustworthy they found the character to be.

4 Results

Firstly, we counted for every participant the number of times that they selected “lie” throughout the experiment, for each render style. After conducting an ANalysis Of VAriance (ANOVA), we found a significant difference between the three styles on this data ($F_{2,44} = 5.53, p < 0.008$). Post-hoc analysis using Newman-



Figure 3: Facial and eye motion applied to High Quality character.

Keuls tests showed that this difference was due to the fact that sequences rendered in HQ were rated significantly more often as ‘lie’ than sequences rendered in NPR ($p < 0.006$). We then calculated the percentage of correct responses for each render style, for each participant. Overall, participants were not very accurate at the task of differentiating truths from lies, and after conducting an ANOVA we found that accuracy was not affected by render style (51% correct for HQ, 51% for Game, and 50% for NPR).

Finally, individual ANOVA tests were conducted on the qualitative data, and we found no difference between ratings for friendliness or trustworthiness between the 3 renders. However, participants rated HQ as having the highest quality overall, Game was next, with NPR being rated the lowest on overall quality ($p < 0.007$ in all cases). In addition, HQ was rated as having higher quality facial ($p < 0.007$) and eye ($p < 0.004$) animation than NPR.

5 Discussion

In this experiment, we aimed to determine if changing the rendering style of a conversing virtual human alone can change how trustworthy they are perceived to be. We found some evidence to support this, since participants judged HQ to be lying more often than NPR during the task. However, all styles were judged as equally trustworthy in the qualitative ratings, which implies that a subconscious feeling of un-trustworthiness was felt by participants towards HQ. It may be argued that this was due to subtle cues being easier to detect in HQ than NPR. Contrary to this, we found that task accuracy and confidence levels were the same regardless of render style. However, accuracy was low in general so further investigation will be necessary in order to determine if deception is the correct paradigm to explore the UV.

Acknowledgements: This work was funded by a TCD postdoctoral Innovation Bursary, the Science Foundation Ireland Metropolis Project, DFG grant Perceptual Graphics PAK 38 CU 149/1-2, and EU Project “Tango” (ICT-2009-C 249858).

References

- DEPAULO, B. M., LINDSAY, J. J., MALONE, B. E., MUHLENBRUCK, L., CHARLTON, K., AND COOPER, H. 2003. Cues to deception. *Psychological Bulletin* 129, 74–118.
- MORI, M. 1970. The uncanny valley. *Energy* 7, 4, 33–35.
- STEPTOE, W., STEED, A., ROVIRA, A., AND RAE, J. 2010. Lie tracking: social presence, truth and deception in avatar-mediated telecommunication. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, 1039–1048.
- THOMPSON, C. 2004. The undead zone: Why realistic graphics make humans look creepy. *Slate*.