An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience

V. Vinayagamoorthy, M. Garau, A. Steed, and M. Slater

Department of Computer Science, University College London, Gower Street, United Kingdom
{V.Vinayagamoorthy | M.Garau | A.Steed | M.Slater@cs.ucl.ac.uk}

Abstract

This paper describes a behavioural model used to simulate realistic eye-gaze behaviour and body animations for avatars representing participants in a shared immersive virtual environment (IVE). The model was used in a study designed to explore the impact of avatar realism on the perceived quality of communication within a negotiation scenario. Our eye-gaze model was based on data and studies carried out on the behaviour of eye-gaze during face-to-face communication. The technical features of the model are reported here. Information about the motivation behind the study, experimental procedures and a full analysis of the results obtained are given in [17].

Keywords: avatars, behaviour modelling, co-presence, DIVE, eye-gaze, embodiment, human-computer interaction, immersive virtual environments, non-verbal behaviour, real-time animation, virtual reality.

ACM CSS: H.5.1 Multimedia Information Systems—*Artificial, augmented, and virtual realities*; I.3.7 Three-Dimensional Graphics and Realism—*Animation*; I.3.7 Three-Dimensional Graphics and Realism—*Virtual Reality;* I.3.7 General—*Human Factors*

1. Introduction

Avatars[1] have long been a feature of 3D chat rooms and computer games. More recently, avatars have been used in immersive multi-participant virtual environments. These environments are *places* of social interaction [18] especially for remotely located users [28, 31] and are used for meetings, therapy, training and simulation, entertainment and multiplayer games [29]. In circumstances that require communication and interaction in a shared Virtual Environment (VE), the effective representation of nonverbal behaviours of the participants is important. This is evident in some studies in which avatars with no associated behaviour were seen as "cold" due to lack of expression [38].

The enhanced non-verbal feedback obtained through a face-to-face communication is not readily available to partic-

ipants in a shared VE. Human behaviour in terms of social interaction and communication is a very intricate phenomenon innate to individuals in the real world, and it is extremely difficult to replicate these behaviours in a virtual environment. Despite the complexities involved, it is not entirely impossible to trigger affective experiences for VE users. Studies have shown that emotional states of participants can be conveyed across a shared VE, with the aid of their tone of voice and a minimal avatar [24].

Unfortunately even with the current state-of-the-art technology available to us, it is computationally expensive to track, manage and replicate the many non-verbal behaviours expressed in a virtual interaction between participants. Perhaps it is sufficient to create a behaviour model using inferences from real human behaviour. The behaviour model could be driven using educated estimations about the intention of the human controller [10] thereby getting the avatar to emulate the user and give the *impression* of an expressive avatar. We have attempted to create such a model focusing our efforts on the non-verbal behaviour of the eyes of an avatar.

---

[1]In this paper, '*avatar*' refers to a virtual human used to represent a participant to others in a shared virtual environment while '*agent*' refers to a virtual human with pre-scripted behaviours.

2. Background: Virtual humans

2.1. Expressive Avatars & Agents

Recent research into building virtual humans as expressive conversational aids in communication [10, 40] and virtual therapy [15, 26], has highlighted the importance of nonverbal behaviour. Embodiment of virtual humans in a collaborative setting has been studied in the context of a small-group interaction in a shared VE [8, 38]. These studies suggest that users in VE respond strongly to humanoid representations of other users in the environment. However social interactions are inhibited by the lack of emotional and gesture content in the virtual humans. In order to create convincing virtual humans, researchers should incorporate sufficiently realistic human-like behaviour, and perception modules [11].

There have been some excellent attempts by various research groups to provide a virtual human controlled by high-level variables including emotions, personality, context of speech, role, beliefs, function etc. Project 'Oz' incorporated concepts of believability, emotions and intents in the Tok [7] architecture. However even though emotions of the agents were derived from events happening in the virtual world, the generation of these events did not depend on the emotion of the agents involved. In addition the agents were very simplistic. Cassell et al. have produced a series of autonomous conversational agents embodied with gesture, behaviour and language models [10, 37, 40]. They have also carried out research into generating non-verbal behaviour based on the text input by users [12, 40]. However input-by-text is not the most instinctive mode of communication in an IVE. Other research has included the development of a framework for emotion-based control of agents [39] and models that focus on the impact of emotion on behaviour [23]. There has also been some work done into building a social-emotional relation between agents and users based on past interactions and anticipating future needs of the user [9].

Elsewhere social robots such as Kismet and Leonardo have been developed to try and engage people in a natural and expressive face-to-face interaction [28]. Kismet perceives a variety of natural social cues from visual and auditory channels, and delivers social signals to a person through gaze direction, facial expression, body posture, and vocal babbles. It has been designed to support several social cues and skills. The control variables were language, social cues and user feedback. Other script-based systems have been produced to enable the generation of single characters [25] or of whole environments populated with intelligent characters [27] [41].

2.2. Non-Verbal Behaviour

The background for building non-verbal behaviour models into virtual humans comes from related psychological studies [1]. There are many non-verbal behaviour traits used in the context of social communication including facial expres-

sion, gesture and posture. Due to the large variety and complexity of possible human behaviours, it is vital to have a well-structured model for virtual humans, in terms of physical appearance and behaviour.

The hypothesis behind our research is that behavioural realism is an important factor in optimising levels of copresence, realism and believability experienced by virtual environment participants. In a small step towards exploring the importance of nonverbal behaviour in VEs, we have chosen to focus on the behaviour of eye-gaze, an important constituent of overall face-to-face non-verbal communications. Amongst other functions, the eyes signal turn taking, direct attention and indicate how interesting the conversation is to other partners.

Garau et al. [16] investigated the impact of eye gaze on the perceived quality of communication by comparing the effects of a random and an inferred gaze in avatars on participants communicating via a video-tunnel link. The data for their gaze models was based on the changes in gaze direction, timing and the magnitude. More recently, Lee et al. [22] presented a similar study comparing random, static and inferred eye-gaze models. Their inferred eye-gaze model was based on data and trajectory kinematics collected from real dyadic interactions with the aid of an eye tracker. The data took into account timings, direction, magnitude and the velocity of the changes in gaze. We have adapted the model presented in [22] to include the timing data used in [16]. Both studies concluded that an avatar whose gaze behaviour is related to the conversation provided a marked improvement on an avatar that merely exhibits liveliness.

3. The Behaviour Module

Despite a lot of evaluative and probing research into nonverbal behaviour [19], there are some areas within in which little study has been conducted:

- Building a viable behaviour framework that can be adapted over time to create a virtual identity using qualitative high-level variables

- Studying how changes in these variables should affect the non-verbal behaviour of virtual humans. Should it mimic real life and to what extent?

- Exploring optimum levels of visual and behaviour realism sufficient to make a virtual character believable. Are these levels context-dependent? Are there any trade-offs?

The general question posed in this research is how much behavioural realism is necessary in order to make the avatars sufficiently believable to the user. It has already been established that in order to enable participants experience acceptable levels of presence and co-presence in VEs, the avatar has to portray some level of non-verbal behaviour. This has

been found in a series of studies involving interaction between three participants in a shared VE [30, 33, 38]. In [34] there is anecdotal evidence that participants expect more visually realistic avatars to behave in a manner that portrays greater human-like qualities. On the other hand, where the context of the virtual application does not require a highly photo-realistic avatar, it might be suitable to have a less sophisticated behaviour model to animate the avatar.

In both the Garau et al. [16] and Lee et al. [22] studies, participants were shown a visually restrictive, above-shoulders view of the virtual human. In the study [17], that formed the motivation for the work reported in this paper, the results and experiences from [16, 22] were brought into an immersive setting with full-body representation of the participants. A system was designed and used in an experiment [17] in which two people met in a shared VE to discuss an issue. For the purposes of the experiment, each person was represented by a realistic avatar, or by a cartoonish avatar. Independently, each person's eye gaze was either based on our inferred behaviour model (see Section 3.4), or a random model (see Section 3.5). In addition to the impact of simulated eye-gaze behaviour on the perceived quality of communication in the VE, the study was also designed to ascertain if there was a level of correlation between the visual-realism of the avatar and the behavioural realism of the avatar [31].

3.1. Design of the Behaviour Module

A Behaviour Module was designed for avatars used to represent participants in a shared IVE. It was designed to autonomously control important behaviours (including those that are not consciously driven).

The Behaviour Module was designed to adapt the behaviour of the avatar depending on the status of the participant based on real-time speech and tracking data. The participant is not necessarily cognisant that they are controlling the module indirectly. The module was kept independent of the visual representation of the avatar and was designed to incorporate other non-verbal behaviour algorithms in the future.

The Behaviour Module consisted of an eye-gaze model and other simple body animations. There are two sub-models implemented within the gaze model. One of the algorithms, the inferred model, was designed to provide a realistic gaze simulation that mimicked the eye-gaze behaviour of persons in a real dyadic communication. In this module, the internal state of the avatar was used to produce realistic eye-gaze behaviour in run-time. The avatar can be in one of 4 states. It can be in a 'talking' or 'listening' state depending on the current behaviour of the participant. Independently it can have its eyes in a 'primary position' (looking ahead) or not. The Behaviour Module continuously checks the internal states of the avatar, to determine (a) if the avatar is talking or listening and (b) if the avatar's eyes are in their primary position (that is looking directly ahead) or not. The tracked positions of the participant's head and hand were used to drive the body animations.

The design of the current Behaviour Module allows for the future addition of more realistic virtual behaviours and more sophisticated perception modules. The module did not include any other facial behaviour due to lack of collected data or incomplete computational models that mimic real life. We have not even included all the behaviours exhibited by the eyes (blinking, arching of eyebrows etc.) or other dyadic gestures (turn-taking, directing attention).

3.2. Eye Behaviour

There are some terminologies we adapted from past studies [16, 22] to describe gaze behaviour. The eye is in its *primary position* when a person is looking directly ahead. The associated gaze is the '*at*' gaze. When the person is not looking directly ahead, the gaze is called the '*away*' gaze. The rapid motion with which the eye moves from one focused position to another is called a *saccade*. The time it spends at a focused position is called the *inter-saccade interval*. The time it takes to get to the new position is the *saccade duration*. The angle the eye rotates in order to get to the new position is the *saccade magnitude*, the direction it moves in is the *saccade direction*, and the non-uniform velocity it moves with to reach the new position is the *saccade velocity*.

There are different types of saccades in accordance to the position of the eye in the orbit: those that start at the primary position, those that end near the primary position, and those that start and end at a distance from the primary position. The position of the eye in the orbit does not have a significant effect on small naturally occurring saccades [21]. The eye saccades in our module, alternated between those that started from the primary position to those that ended at the primary position.

3.3. Background to the Inferred Eye-Gaze Model

In addition to theoretical information from social psychological studies [1–3], the eye-gaze model was built on information gathered from case studies of people in a two-way conversation [22].

The main assumptions behind the inferred eye-gaze model were that for behaviour in a dyad:

- The mean saccade magnitude of the eye while a person is listening is less than that of a person speaking [1–3]
- People tend to look at their conversational partner more *while listening* than *while speaking* [1–3]
- The inter-saccadic interval between the focus positions is shorter when a person is speaking than when a person is listening [1–3]

Table 1: *Mean inter-saccadic intervals*

	Looking *At* (s)	Looking **Away** (s)
While **listening** to conversational partner	2.5	1.6
While **speaking** to conversational partner	1.8	2.1

Table 2: *Occurrence of saccade directions*

Direction: anti-clockwise from right	0°	45°	90°	135°
Percentage of occurrence (%)	15.5	6.5	17.7	7.4
Direction: anti-clockwise from right	180°	225°	270°	315°
Percentage of occurrence (%)	16.8	7.89	20.4	7.8

The magnitude of the saccades and the intervals between each saccade varies with the state of the participant. The inter-saccadic intervals are dependent on two factors:

- Whether the person is speaking or listening and
- Whether the avatar's eye is in the primary position or not

3.4. Computing the Inferred Eye-Gaze Behaviour

The eye begins at the primary position at the start of the conversation. In the earlier non-immersive studies [16, 22] the primary position of the avatars' eyes and head was defined as directly ahead. In our model, with the advantage of tracking in the VE, participants were left in full control of the avatar's head, even though the primary position of the eyes are still "*directly ahead*." The eye behaviour was kept independent of the head movements unlike [16, 22]. The avatars were in mutual gaze, only if each person *chose* to look at the avatar of their partner and vice versa.

During the conversation, each avatar's Behaviour Module checks the status-monitoring variables to determine if the avatar should be in 'talking' mode or in 'listening' mode. It also checks if the avatar should be in "*looking at*" or "*looking away*" mode.

Once the status of the avatar's behaviour is determined, the appropriate inter-saccadic interval is calculated on the basis of an exponential probability distribution, about the mean times in Table 1. The eye is paused for the duration of the interval. At the end of the inter-saccadic interval, the magnitude and the gaze direction of the next saccade, if not to the primary position, is calculated.

Table 1 shows the mean inter-saccadic intervals and as can be observed the data is in line with the theory that people tend to look away more often and for longer periods of time while speaking in a dyad [1–3, 16]. Unlike [22], we implemented our timing functions in terms of absolute times (precision to microseconds) rather than frame rates. This was to ensure that the timers would not be influenced by the changing refresh rates of different equipment.

According to the statistical analysis of the data on saccades in [22], the frequency of occurrence of a particular saccade magnitude can be fitted to an exponential function (Equation 1). The saccade *magnitude* of an '*away*' gaze is calculated as shown in (Equation 1).

$$Magnitude = -6.9 * \ln(Rand/15.7) \qquad (1)$$

The variable '*Rand*' is a random number from a uniform distribution between 0 and 15. The maximum magnitude of a *saccade* in the distribution is dependent on whether the person is talking or listening. While a person is talking in a dyad the maximum saccade magnitude is 27.5 degrees. It reduces to 22.7 degrees when the person is listening. The module ensures the magnitude calculated does not exceed the maximum magnitude value corresponding to the state of the participant. The model is consistent with the theory that 90% of all natural saccades are less than 15° [6]. The accompanying gaze *direction* of an '*away*' gaze is chosen out of eight distinct saccade directions evenly spaced at 45 degrees (Table 2). Provided the next saccade is not to the primary position of the eye, its direction is determined using a uniform random number generator and the list of probabilities (Table 2) that were made available in [22]. The eye-gaze direction is independent of head rotation.

The duration of each saccade is dependent on the magnitude of the saccade [22]. The relationship between saccade duration and magnitude is given in (Equation 2).

$$
\begin{aligned}
Duration &= \Delta + \delta * Magnitude \\
&\text{where} \\
\Delta &= 25 \text{ milliseconds} \\
\delta &= 2.4 \text{ milliseconds/degree}
\end{aligned}
\qquad (2)
$$

'Δ' is the intercept of catch up time of the eye. Experimental evidence suggests the value of Δ is between $20 - 30$ milliseconds and we have chosen a constant value of 25 milliseconds. The variable 'δ' is the incremental slope.

The final variable in the inferred eye-gaze model is the velocity of the eyes during a saccade. In reality, the eyes do not maintain a continuously constant velocity. They accelerate from a position to a maximum velocity and then decelerate to a stop at a new position. A higher magnitude results in a higher average saccade velocity. Lee et al. [22] constructed a polynomial function to calculate the instantaneous velocity of the eye during a saccade. However, the function did not yield the expected numerical results as given in the collected data. We replaced the function with an alternative model.
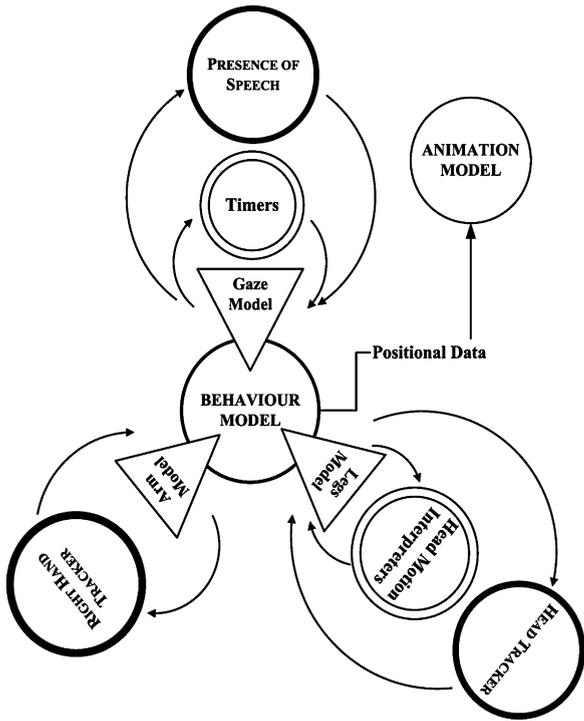
Figure 1: *The Data Flow in the Behaviour Module. It consists of an inferred eye-gaze model and other models to animate the legs and arms of the avatar.*

Figure 2: *Instantaneous saccade velocities; our heuristic model with the reproduced data markers.*

We broke the path of each saccade into six equally spaced frames and calculated instantaneous velocities based on a heuristic model (Equation 3).

$$Y = 14 * \exp\left\{-\frac{\pi}{4} * (X - 3)^2\right\}$$

where $\qquad$ (3)

Y = Instantaneous Velocity at frame X

X = [1, . . . , 6]

The six frames in the saccade are set based on the saccade duration calculated in (Equation 2), the eye is then moved to the corresponding intermediate positions at the end of each frame depending on the elapsed time and the instantaneous velocity (Equation 3).

Figure 2 depicts the instantaneous saccade velocities as predicted by our model, which closely mimics the data in [22]. Since the intermediate shifts in eye gaze is normalized to the total magnitude of the eye as calculated, the eye follows the curve (Figure 2) and moves to the corresponding saccade position calculated originally. When the eyes reach the end of the saccade, the whole process of finding a new saccade magnitude, direction, duration and an inter-saccadic interval begins again as depicted in Figure 3. Figure 4 depicts the
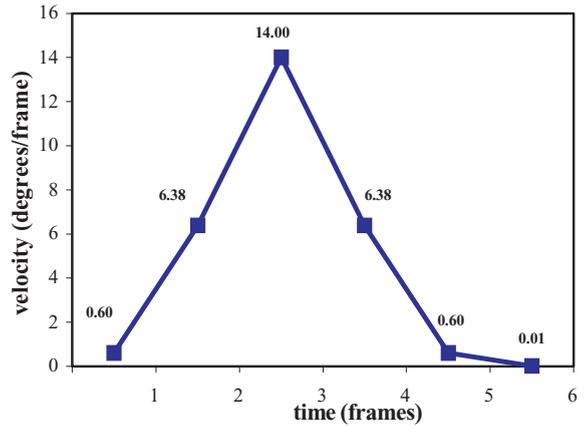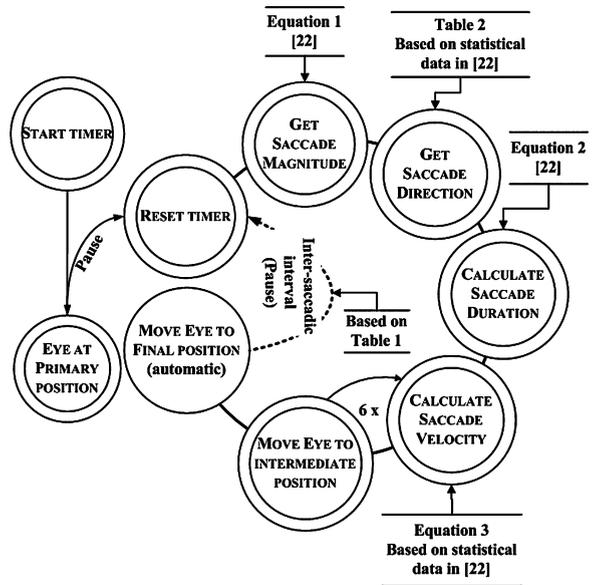

Figure 3: *Inferred Eye Gaze Model.*

inferred eye gaze model in action while the avatars were in listening mode.

3.5. Computing the Random Eye-Gaze Behaviour

We implemented a random eye-gaze model to use in the control conditions of studies. The model follows the same principles as the inferred eye-gaze model except there are no saccade velocities involved. The magnitude of the saccade is obtained using a uniformly distributed random number generator. In order to avoid having a model that is too unrealistic, we limited the saccade magnitude to a maximum of 15 degrees [6]. The direction of the eye was implemented as a
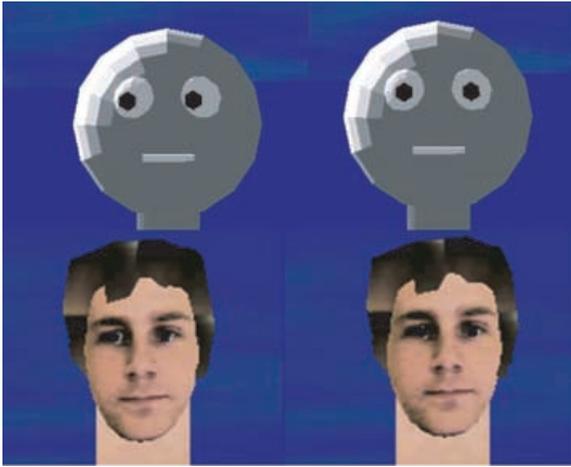
Figure 4: *Top: Cartoonish avatar 'looking away' and 'looking at'. Bottom: Realistic avatar 'looking away' and 'looking at'.*

random direction uniformly distributed around a 360° circle centred at the eye as opposed to the eight distinct directions in the inferred eye-gaze model. The inter-saccadic interval was a constant 2 seconds independent of whether the avatar was talking or listening, looking at or looking away from the conversational partner. The inter-saccadic interval (2 seconds) was the mean of all the inter-saccadic averages from the inferred eye-gaze model.

3.6. Accompanying Body Animations

Some simple motion algorithms were included into the Behaviour Module with a view to preserving visual consistency between the movement of the eyes and the rest of the body. Animations were created for the avatars' head, right arm and legs. The complexity of the body animation in the avatars was limited since the only body parts tracked in our IVE systems are the head and the right hand. The body animations of the avatars in this study *semi-mirrored* the motions of the participants in the shared virtual space depending on the data information obtained from the limited number of tracking sensors used on the participant (head and right hand) [5]. However, there was an advantage in our VEs to counterpart the lack of tracking. The participant using our systems did not see both their real body and their animated virtual self at the same time (see Section 5.2). This meant that when animating the avatar we could place more importance on visual consistency over accuracy in mirroring the body of the participant.

3.6.1. Head Control

Direction of travel in the virtual environments was based on the direction of movement of the users' head tracker in the

X-Z plane. Initially, the avatar's head and body simply imitated the rotation and translation of the participants' head tracker in the X-Z plane. However, it was noticed during pilot studies that the head trackers in our virtual reality systems were not placed directly above the centre of the participants' head. This gave the impression that a translation was associated with every nod and shake gesture imparted by the participants. Early versions of the Behaviour Module deduced that the body of the avatar was being moved along with its head during these conversational gestures. It animated the avatar's body by rotating it from side to side during a headshake gesture e.g. when the user gestured 'no'. Whereas if the user nodded their head, e.g. to gesture 'yes', the Behaviour Module deduced that the whole avatar was moving forward and backward rapidly.

Simple gestures such as head nods and shakes play an important factor in social interactions. Therefore we implemented a simple rule in the Behaviour Module that defined a boundary circle around the position of the avatar's head. The boundary circle had a radius equivalent to seven centimetres in the X-Z plane. When the head tracker is moved a distance greater than the radius of the boundary circle (>7 cm), the whole body of the avatar rotated to the direction of travel of the participant and moved the corresponding distances moved by the head tracker, hence representing the position of the participant to their partner in the shared VE. However, when the motion detected by the users' head tracker was smaller than the boundary radius (<7 cm), the Behaviour Module prevented the body of the avatar from translating along the ground but allowed the avatars' head to correctly emulate the rotations of the participants' head.

This allows the Behaviour Module to accurately represent the gestures of the avatar's head using only the rotational data collected from the participants' head tracker. Participants could take full advantage of any head gestures they would have normally used in the real world in the virtual world. In addition to the head animations, the right arm and legs of the avatars were animated to follow the movements of the VE user. Some of the resultant body poses of the avatars are depicted in Figure 5.

3.6.2. Arm and Leg movements

In order for the Behaviour Module to accurately animate the avatars limbs, the correct height of the avatar and the length of the avatars' limbs were deduced in the first virtual frame and initialised. This allowed the Behaviour Module to deduce if the participant was bending their right arm in latter frames or crouching. It also allows for deductions of simple situations such as the participant taking off their trackers or positional data from one of the trackers not being received correctly. The right arm was animated based on a deduction of joint locations based solely on the positions of the hand and head.

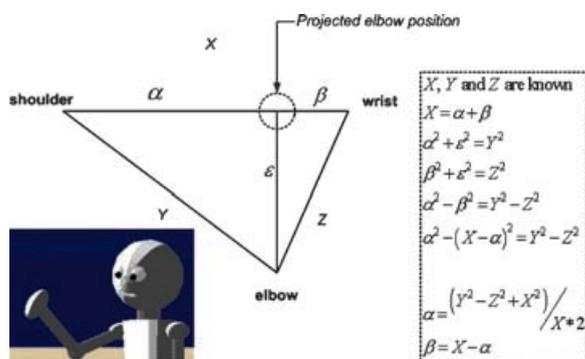**Figure 5:** *Some body poses displayed by the avatars.*

Figure 6: *Depiction of the arm in a single plane and on the cartoonish avatar.*

The position of the avatar's shoulder was always known due to the hierarchical nature of the avatar. In accordance to the hierarchical structure of the H-Anim [42] complaint avatars, the shoulders of the avatar were children joints of the body. Since the body followed the X-Z position of the head tracker, it was possible to deduce the current position of the right shoulder using an object location functionality in DIVE [35]. The position of the right wrist was always obtainable due to the data collected from the right hand tracker embedded in the navigational joystick in our systems.

The vector position of the right elbow of the avatar was then deduced using the 3D vector co-ordinates of the right shoulder and the right wrist of the avatar. The 3D position of the right elbow of the avatar projected down on the axis between the right shoulder and the right wrist was deduced using the cosine rule as depicted in Figure 6. In addition, some geometrical deductions were made to ensure the tracked right wrist of the avatar was not detached from the body of the avatar e.g. when the participant dropped the joystick. The deduced vector co-ordinates of the *projected elbow* was then cast onto a direction orthogonal to that of the axis defined by the right shoulder and the tracked wrist. Once the right elbow position of the avatar was deduced, the geometries of the right upper and lower arm were aligned between the appropriate joints to re-produce a possible 3D posture of the avatars' right arm in a virtual frame.

There was a special consideration when the arm of the participant was longer than that of the corresponding avatar. In this case, the directional data from the tracker is used to reproduce the correct posture of the right arm but the translation of the right hand tracker is not conveyed to the right wrist of the avatar. During pilot studies, this methodology for manipulating the right arm was found to be adequate.

As mentioned, the Behaviour Module infers information about the dimensions of the avatar as part of the initialisation process. The legs of the avatar were animated using deductions based on the height of the participant. Once a reduction in height of the participant is detected, the Behaviour Module uses the translation data from the users' head tracker to lower the position of the body of the avatar (including the head) in all axes. The current position of the hips of the avatar during the virtual frame is obtained. Once the 3D coordinates of the hips and feet were determined, the positions of the knees were bent in the direction of the Z-axis based on total leg length (Figure 5). The feet of the avatar however, are only updated in the X-Z plane i.e. the Y co-ordinates of the feet are reset to null. The upper and lower leg geometries were aligned in between the appropriate joints of the avatar. All body animations were updated for every frame.

4. Implementation of the Behaviour Module

The scenarios for the study were implemented on a derivative of DIVE 3.3x [14, 35] which had been ported to support immersive systems [32]. DIVE is made modular by the use of plugins, which provide a set of symbols and functions recognised by DIVE. These symbols are looked up by the DIVE plugin interface and invoked when needed. A plugin was constructed in C to animate the avatar appropriately when it detected interactions or events in run-time.

When speech is detected via the microphone attached to the participants, appropriate event messages were sent to DIVE. The Behaviour Module scanned for these messages by the frame to determine whether the avatar should be in "*speaking*" or "*listening*" mode. The messages were detected and processed by the plugin to provide appropriate eye animation in the avatar. Virtual humans in a DIVE world can be either an avatar or an agent. In the case of an avatar, DIVE reads the participants' input devices and maps the physical actions undertaken by them to logical actions in the system. In this study the head and the right hand were tracked. The Behaviour Module used these tracked positions to animate the body of the avatar.

Figure 7: *Left - Participant in the CAVE^{TM}-like ReaCTor system. Right - Participant in the Head-Mounted Display.*

5. Experimental Design and Materials

5.1. Specifications from the Experimental Design

The Behaviour Module was used to animate avatars in a study designed to assess the impact of visual and behavioural realism in avatars on the perceived quality of communication in a shared IVE. The study required two participants to undertake a ten-minute conversation task in a shared IVE.

Three different avatars were used: one cartoonish gender-neutral avatar, and two realistic gender-specific avatars. The avatars used were adapted to comply with an H-Anim based hierarchical structure in order to ensure similar results from the Behaviour Module. The same module was used in all the avatars to ensure identical functionality and avoid confounding results. In terms of behaviour, the only variation was between the inferred-gaze model and the random-gaze model. A fixed-gaze model was not implemented as there is evidence [4] that continuous gaze can result in negative evaluation of a conversation partner. The main purpose of the experiment was to investigate if our inferred eye-gaze model was noticeably better than a random (but not obviously incorrect) model. Refer to [17] for details on the task, methodology and justification behind the study.

5.2. Equipment

Since there were two participants (Figure 7) involved in each trial of the study, two sets of equipment were used. One participant operated in a ReaCTor system, which is similar to the CAVE^{TM} described by Cruz-Niera [13], while the other used a head-mounted display (HMD). The Trimension ReaCTor, consists of three 3 m × 2.2 m walls and a 3 m × 3 m floor. The participants wore CrystalEyes stereo glasses,

Figure 8: *The avatars (left to right): cartoonish and genderless; realistic male; realistic female.*

which were tracked by an Intersense IS900 system accurate to within 2mm with an end-to-end latency of 50ms. The HMD used was a Virtual Research V8. The tracking system has two Polhemus Fastraks, one for the HMD and another for a 5 button 3D mouse. Further details on the equipment used can be found in [17].

The participant using the ReaCTor did not see their virtual self while the participant using the HMD did not see their real body. In addition, the rate at which the trackers updated the networks was kept constant at 10Hz for both the ReaCTor and the HMD. This was a further precaution to ensure results obtained were purely due to the implementation of the eye-gaze behaviour in the avatars and not a difference in the quality of display systems or fluidity of the body animations.

Both participants had wireless microphones attached to their clothing. Existing software was used to distinguish between speech and background noise in the audio data from the microphones. The software was calibrated by setting a threshold value for each microphone at the beginning of the day. Audio input higher than the threshold was indicative of the presence of speech. When the machines detected the presence of speech from either microphone, event messages were sent to DIVE for appropriate action to be taken in the Behaviour Module of the corresponding users' avatar.

5.3. Physical Design

Three H-Anim [42] complaint avatars were designed for use in this study, one of which was visually very cartoonish. The cartoonish avatar was built based on mathematical proportions used by artists [20]. The dimensions of all the body parts were described, as a multiple of the size of it's head e.g. the height of the avatar was eight times that of it's head. The whole avatar could be scaled to a required height by manipulating a variable.

An additional design feature taken into account in the realistic avatars (Figure 8) was the subtleness of all features on the face in order to avoid converging attention to any one feature. However, the eyes of the cartoonish avatar were emphasised in order to encourage the participant's attention to the eyes and less on the lack of other facial expressions. It was vital to avoid other accompanying facial expressions in

order to ensure any resulting effects were purely due to the eye-behaviour.

6. Conclusions and Future Work

A complete report on the results of the study can be found in [17]. Briefly, the conclusion was that in the case of the cartoonish avatar, the realistic (inferred) eye-gaze model did not improve the perceived quality of communication whereas for the realistic avatar the inferred eye-gaze behaviour increased effectiveness. It was concluded that it is likely that there would be a requirement for consistency between the visual realism of the avatar and the behavioural realism that it exhibits with respect to eye-gaze, which supports the anecdotal results reported in [34]. It was also noted during the experiment that the conversational partners stood facing each other and maintained personal space.

The Behaviour Module designed for the experiment was able to cope with the varying behaviour needs of each of the four experimental conditions, which demanded either a highly-realistic inferred eye gaze behaviour model or a non-realistic random eye-gaze model on two visually different avatars. There is however some outstanding issues that have to be dealt with in future work.

The study was not extended to encompass the effects on participants' perceived quality of communication when a cartoonish avatar, embodied with an *exaggerated* behaviour model [36], represents the conversational partner. In a sense the non-realistic random eye-gaze model was more dynamic and visually stimulating than the subtler eye movements of the realistic behaviour model. Perhaps participants in a shared virtual environment need to be made to believe that a cartoonish character is "*alive*" in order to feel a high sense of presence when interacting with the avatar. This theory will be tested in our future work.

A drawback in our study was the low frame rate. However, due to the task being based on holding a conversation about an engaging matter [16, 17] much like a virtual teleconferencing application, the effects of having a low frame rate (10 Hz) in a run-time animation was minimised since most participants stood face-to-face maintaining personal space.

The experiment was conducted in a purpose-built low-detailed virtual environment. This was important since in real life, the more features in an environment the more visually distracted the participants are likely to be, hence diverging from the collected data in [22]. Obviously interesting features, especially those that are moving, will change the behaviour of the persons' gaze in real life. Therefore the eye-gaze model of the avatar will have to be modified to reflect these changes.

It cannot be assumed that the results found in this study will hold when applied to other forms of human behaviour within an immersive virtual reality setting. The behaviour model is still lacking in other non-verbal behavioural factors such as posture, gait, gesture and other facial expressions (smiling, lip synching). This was apparent from some of the comments from the participants. However, it does open an exploratory question into other non-verbal behaviours. It is highly probable that there is correlation between visual realism and behaviour realism with respect to most other behaviours. In which case, we could potentially model all non-verbal behaviour in a similar manner thereby providing an inexpensive way to build affective virtual humans.

In the future, we aim to make the behaviour model completely independent of the rendering platform and allow for further high-level abstract control variables such as the possible emotions, personality, role, identity etc. of the participant. This would involve the collection of more data based on real-life scenarios and experiments to explore optimum levels of visual-behaviour realism with respect to behaviours in the context of other applications. By embodying an avatar with behaviour, emotional or personality skills, we provide the participant with a *virtual character* in the full sense of the word.

References

1. M. Argyle. *Bodily Communication*, Second Edition, Methuen & Co Ltd, 1998.

2. M. Argyle and M. Cook. *Gaze and Mutual Gaze*, Cambridge University Press, 1976.

3. M. Argyle and R. Ingham. "Mutual Gaze and Proximity," *Semiotica*, 6, pp. 32–49, 1972.

4. M. Argyle, L. Lefebvre and M. Cook. "The Meaning of Five Patterns of Gaze," *European Journal of Social Psychology*, 4(2), pp. 125–136, 1974.

5. I. N. Badler, M. J. Hollick and J. P. Granieri. "Real-Time Control of a Virtual Human Using Minimal Sensors,"

Presence: Teleoperators and Virtual Environments, 2(1), pp. 82–86, 1993.

6. T. A. Bahill, D. Adler and L. Stark. "Most naturally occurring human saccades have magnitudes of 15 degrees or less," *Investigative Ophthalmology*, 14, pp. 468–469, 1975.

7. J. Bates, B. A. Loyall and S. W. Reilly. "An Architecture for Action, Emotion and Social Behaviour," CMU-CS-92-144, May 1992.

8. S. Benford, J. Bowers, L. E. Fahlen, C. Greenhalgh and D. Snowdon. "User Embodiment in Collaborative Virtual Environments," *SIGCHI*, Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 242–249, 1995.

9. W. T. Bickmore and J. Cassell. "Relational Agents: A Model and Implementation of Building User Trust," *SIGCHI*, pp. 396–403, 2001.

10. J. Cassell, W. T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. H. Vilhjálmsson and H. Yan. "Embodiment in Conversational Interfaces: Rea," *SIGCHI*, pp. 520–527, 1999.

11. J. Cassell, W. T. Bickmore, L. Campbell, K. Chang, H. H. Vilhjálmsson and H. Yan. "Requirements for an Architecture for Embodied Conversational Characters," *Computer Animation and Simulation*, pp. 109–120, 1999.

12. J. Cassell, H. H. Vilhjálmsson and W. T. Bickmore. "BEAT: The Behaviour Expression Animation Toolkit," *SIGGRAPH*, pp. 477–486, 2001.

13. C. Cruz-Neira, D. J. Sandin and T. A. DeFanti. "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," *Proceedings of Computer Graphics (SIGGRAPH)*, pp. 135–142, 1993.

14. E. Frécon, G. Smith, A. Steed, M. Stenius and O. Stahl. "An Overview of the COVEN Platform," *Presence: Teleoperators and Virtual Environments*, 10, pp. 109–127, 2001.

15. D. Freeman, M. Slater, P. Bebbington, P. A. Garety, E. Kuipers, D. Fowler, A. Met, C. M. Read, J. Jordan and V. Vinayagamoorthy. "Can Virtual Reality be Used to Investigate Persecutory Ideation?," *Journal of Nervous and Mental Disease*, 191(8), pp. 509–514, August 2003.

16. M. Garau, M. Slater, S. Bee and M. A. Sasse. "The impact of eye gaze on communication using humanoid avatars," *SIGCHI*, Conference on Human Factors and Computing Systems, pp. 309–316, March 2001.

17. M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed and M. A. Sasse. "The Impact of Avatar Realism and Eye Gaze Control on the Perceived Quality of Communication in a Shared Immersive Virtual Environment," *SIGCHI*, April 2003.

18. S. Harrison and P. Dourish. "Re-place-ing space: the roles of place and space in collaborative systems," *Proceedings of the conference on Computer Supported Cooperative Work*, pp. 67–76, 1996.

19. S. Kshirsagar and N. Magnenat-Thalmann. "A Multilayer Personality Model," *Smart Graphics*, pp. 107–115, June 2002.

20. R. Larmann. "*The Figure Drawing Laboratory, University of Evansville*," Last accessed: 30 Oct. 2003, `http://www2.evansville.edu/drawinglab`.

21. S. P. Lee. "Facial Animation System with Realistic Eye Movement based on a Cognitive Model for Virtual Agents," Thesis/Dissertation, 2002.

22. S. P. Lee, I. N. Badler and J. Badler. "Eyes Alive," *SIGGRAPH*, pp. 637–644, July 2002.

23. S. Marsella, J. W. Lewis and C. LaBore. "Interactive Pedagogical Drama," *Proceedings of Autonomous Agents,* 4th International Conference, pp. 301–308, 2000.

24. J. Mortensen, V. Vinayagamoorthy, M. Slater, A. Steed, M. C. Whitton and B. Lok. "Collaboration in Tele-Immersive Environments," *Eight Eurographics Workshop on Virtual Environments*, pp. 93–101, May 2002.

25. K. Perlin and A. Goldberg. "Improv: A System for Scripting Interactive Actors in Virtual Worlds," *SIGGRAPH*, pp. 205–216, 1996.

26. D.-P. Pertaub, M. Slater and C. Barker. "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence: Teleoperators and Virtual Environments*, 11(1), pp. 68–78, February 2002.

27. N. Richard, P. Codognet and A. Grumbach. "The InViWo toolkit: Describing autonomous virtual agents and avatars," *Intelligent Virtual Agents*, pp. 195–209, 2001.

28. Robotic Life, "*Media Lab, Massachusetts Institute of Technology*," Last accessed: 30 Oct. 2003, `http://robotic.media.mit.edu/`.

29. R. Schroeder. *The Social Life of Avatars; Presence and Interaction in Shared Virtual Environments*, Springer-Verlag New York, Inc., New York, 2002.

30. M. Slater, A. Sadagic, M. Usoh and R. Schroeder. "Small Group Behaviour in a Virtual and Real Environment: A Comparative Study," *Presence: Teleoperators and Virtual Environments*, 9(1), pp. 37–51, February 2000.

31. M. Slater and A. Steed. Chapter 9: Meeting people virtually: experiments in shared virtual environments, In *The social life of avatars: presence and interaction in shared virtual environments*, Springer-Verlag New York, Inc., New York, pp. 146–171, 2002.

32. A. Steed, J. Mortensen and E. Frécon. "Spelunking: Experiences using the DIVE System on CAVE-like Platform," *Immersive Projection Technologies and Virtual Environments*, (B. Frohlicj, J. Deisinger and H.-J. Bullinger Eds. Wien: Springer-Verlag), pp. 153–164, 2001.

33. A. Steed, M. Slater, A. Sadagic and J. Tromp. "Leadership and collaboration in virtual environments," *IEEE Virtual Reality*, pp. 58–63, March 1999.

34. S. Strippgen. "Insight: A Virtual Laboratory For Design, Test And Evaluation Of Autonomous Agents," *International Conference on Autonomous Agents*, 1998.

35. Swedish Institute of Computer Science, "*The DIVE home page*," Last accessed: 3 Nov. 2003, `http://www.sics.se/dive`.

36. F. Thomas and O. Johnson. *Disney Animation—The illusion of life*, Abbeville Press, New York, 1981.

37. R. K. Thórisson and J. Cassell. "The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents," *Applied Artificial Intelligence*, pp. 519–538, 1999.

38. J. Tromp, A. Bullock, A. Steed, A. Sadagic, M. Slater and E. Frécon. "Small Group Behaviour Experiments in the COVEN Project," *IEEE Computer Graphics and Applications*, 18, pp. 53–63, 1998.

39. J. D. Velásquez. "Modelling Emotions and Other Motivations in Synthetic Agents," *American Association for Artificial Intelligence*, pp. 10–15, 1997.

40. H. H. Vilhjálmsson and J. Cassell. "BodyChat: Autonomous Communicative Behaviours in Avatars," *Proceedings of Autonomous Agents,* 2nd International Conference on Autonomous Agents, pp. 269–276, May 1998.

41. S. Vosinakis and T. Panayiotopoulos. "SimHuman: A platform for real-time virtual agents with planning capabilities," *Intelligent Virtual Agents*, pp. 210–223, 2001.

42. Web3D Consortium, "*Humanoid Animation Working Group of the WEB3D Consortium*," Last accessed: 30 Oct. 2003, `http://www.hanim.org/`.