ABSTRACT

What is realism? Is it to do with the geometry of humans or is it more a measure of their behaviour?

More physical realism demands an increase in behaviour realism in order to produce more co-presence, realism and believability in virtual environment users. It is believed that users of virtual environment expect a more realistic behaviour model consistent with human behaviour in order to buy into the idea that a virtual model of a human is *real*. This not only includes the correct kinematical solution to the motion of the model's body parts but also verbal and non-verbal social communication skills.

This concept builds on the work done by others in the *Virtual Environment and Computer Graphics* group at University College London. In most cases the hypothesis has been tackled by research into a particular behaviour. The behaviour in this research is *posture* and the control variables are various human *emotions* and *moods.*

There is an added factor that has to be taken into consideration. The purpose of this thesis is not only to study the various relationships that may arise between posture and emotions but also to deliver an avatar to incorporate into the project funding this PhD, *Equator*. This gives an added reason to make the avatar completely generalised and manageable by a variety of programs across various platforms.

This report is intended as a first year viva into my PhD. The first three chapters covers work done by others and chapters five and six covers my perceived contribution, project plans and direct/indirect work done.

ACKNOWLEDGEMENTS

This research is being funded by *Equator* and is a requisite part of this PhD. Equator is a six-year Interdisciplinary Research Collaboration (IRC) project supported by the EPSRC. It brings together researchers from eight different institutions (inc. University College London) across a variety of disciplines in order to address the technical, social and design issues involved in the integration of physical and digital interaction.

I would like to take this opportunity to acknowledge everyone, who has been conducive to this research so far, in particular to my supervisors (Prof. Mel Slater, Dr. Anthony Steed), Celine Loscos (for *Pureform*), Pip Bull (for his help in my understanding avatar animation), Andrea Brogni, Marco Gillies, David Swapp and fellow PhD students (Jesper Mortensen, Joel Jordan, Maia Garau). I would also like to thank Benjamin Lok and colleagues at the University of North Carolina, Chapel Hill for their help in the *Internet2* project. This gratitude is extended to Alican Met for giving me the opportunity to take a glimpse into the world of virtual paranoia. My heartfelt appreciation towards my family and friends for the support and encouragement they have shown me throughout various parts of my learning curve.

Finally, I would like to dedicate this work to Vidula who aspires to embark on an academic journey pertaining to the weird and wonderful world of computers.

TABLE OF CONTENTS

LIST OF FIGURES

PRELIMINARY CONCEPTION

Emotional and Neurodevelopmental functions are like a two-way street: emotional problems may weaken the functions and weakened functions can cause emotional turmoil.
[Levine 2002: (42)]

1.1 WHAT GIVES AN AVATAR* LIFE?

Most avatars currently in use in virtual worlds have been designed to look humanoid but do not include a convincing behaviour model within their structure. When users interact with these virtual humanoids, they lack the enrichment that non-verbal communication provides whilst communicating with each other in the real world [Vilhjálmsson *et. al.* 1998: (82)].

The bulk of this research will concentrate around a vital question. What are the behavioural elements needed to make an effective avatar? If so, which non-verbal components command more significance? Finally, can a behaviourally realistic avatar have more *effect*† on users immersed in virtual reality than a detailed photo-realistic avatar with no behaviour?

It is reasonable to believe that if the exterior of an avatar is designed with enough photorealistic details to look human then users will expect it to have a convincing behaviour model. That is, the perceived behaviour of an avatar has to increase in realism as much as its visual consistency and photo-detail. This is because when people accept a virtual humanoid to represent a real person, they expect the avatar to act human as well. Inconsistencies in its behaviour break the sense of presence felt in the virtual environment (VE). This is a hypothesis of this research.

1.2 OBJECTIVES

Due to time limits and good experimental procedure, a single behaviour has to be researched at any one time. The non-verbal behaviour chosen for this research is '*Posture*' and how it is related to the *emotional* state of mind. This includes the consequence of *emotions* on posture and the usage of the knowledge that prolonged association with emotion lead to *moods*.

* Avatar — An electronic/digital representation of a human in virtual environments
† Effect — Measured in terms of levels of Presence, Co-Presence & Believability

1.2.1 DELIVERABLES

At the end of this research it is hoped that some of the following will be available for general usage.

- A set of body poses that attune to the users' perceived emotional state.
- Demonstrate that behaviourally realistic avatars have more "*effect*" on the user than graphically detailed avatars with no innate behaviour.
- Assess if behaviour realism can compensate for diminished photorealism.
- Establish the optimal amount of behaviour animation.
- Create an avatar with variable levels of control for utilisation within the Equator* project.

1.3 HYPOTHESIS

Psychological studies claim that the main channels for communication of the emotional state of mind are through the face, body and the tone of voice in that order. The face has evolved as a social signalling route. Gestures, postures and body movements are a secondary channel for emotional conveyance [Ekman *et. al*. 1967p: (30)]. It has also been suggested and confirmed that the face conveys specific emotions while the body conveys the degree of intensity of the concerned emotion [Graham *et. al*. 1975: (34)]. Although more recently, research suggests that the body can display more specific emotional content than was previously thought [Argyle 1998b: (3)].

The comprehensive hypothesis behind this research in a nutshell is:

> *A virtual humanoid endowed with realistic behaviour models can evoke better levels of presence, co-presence and believability than an avatar with no innate behaviour.*

In this study the behaviour is posture and the control variables are emotions† and moods‡.

> *In addition, it is postulated that an exaggerated behaviour model incorporated into a minimally rendered avatar will evoke a surprisingly convincing state of realism hence compensating for the lack of photo-realism.*

A 2D graph indicating the postulated relationship between photorealism and behaviour realism for a photo-detailed and simplistic avatar is shown below. Note that this is not indicative of a mathematical relationship between these variables. It is merely a pictorial indication of what type of

* Equator: Multi-Disciplinary IRC Project funding this PhD
† Momentary state of mind normally associated with an event
‡ Prolonged state of mind resulting from cumulative effect of emotions

relationship could exist between levels of co-presence/realism and might emerge from research into photorealism and behaviour-realism. Ideally it should be a 3D graph depicting a relationship between photo-realism, behaviour realism and co-presence.

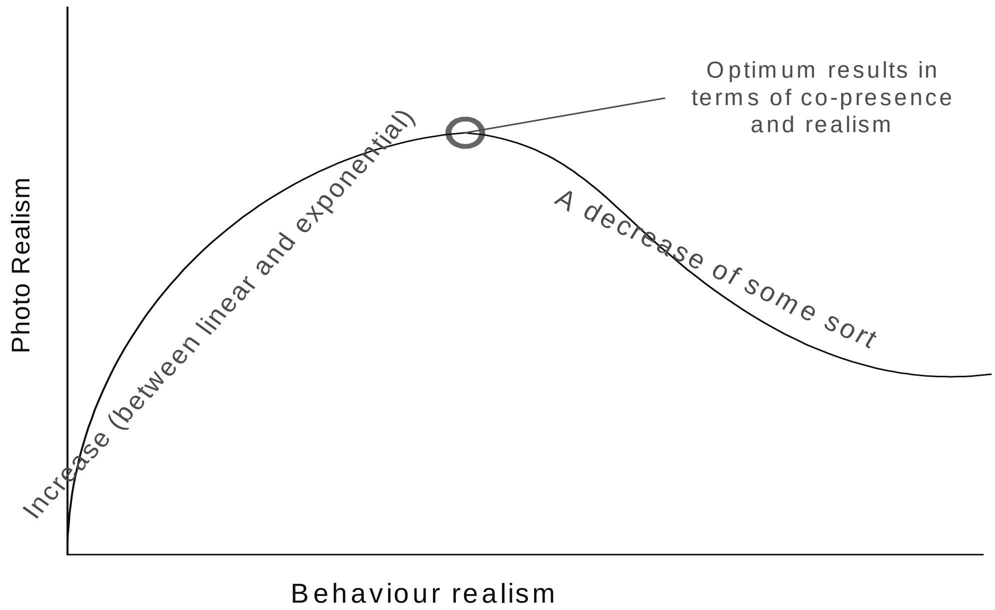


Fig. I Effectiveness for Photorealism vs. Behaviour realism

1.4 MOTIVATION

Emotions affect behaviour at many levels, although reactive behaviours may show less effect than medium to long-term planning ones. Avatars with visual consistency and an innate behaviour model may produce the same (or better) results in virtual environments. If an avatar can successfully portray realistic *emotions*/*moods* (personalities, beliefs, intents…) by displaying appropriate *postures* (or other behaviours) in real-time, then effort put into developing highly behaviour-realistic avatars with acceptable photo-realism and visual consistency could be worthwhile. It could lead to fewer polygons and less processing power/resources used. There are studies that suggest that eye-gaze animations in virtual humanoids in a dyadic conversational situation enhance the experience of virtual environment [Garau *et. al*. 2001q: (33); Colburn *et. al*. 2000: (25); Vertegaal *et. al*. 2002: (80)]. This study has been extended to provide a basis for future experiments concerning this thesis and is discussed in section 6.0. There have even been some systems programmed to generate eye gaze animations [Chopra 1998: (24)], Eyematic [Eyematic 2002: (31)].

2.0 PREAMBLE ON "VIRTUALITY"

> *Face-to-face interaction between people is generally effortless and effective. We exchange glances, take turns speaking and make facial and manual gestures to achieve the goals of the dialogue. Endowing computers with such an interaction style marks the beginning of a new era…*
> *[Thórisson 1996‰: (77); Thórisson et. al. 1999Š: (78)]*

What makes a human being at the bus stop seem real even when we are not directly or deliberately interacting with them are the subtle bodily movements (voluntary or involuntary) going on at any one time. People identify with and respond to cartoon figures in animated movies even though logical process in the mind decode that the characters are not real. In some art pieces, the painter expertly suggests motion in the eyes, breathing and perspiration in a portrait or water movement in a landscape thereby increasing the level of realism. It has been suggested that if characters are consistent (both visually and behaviourally), recognisable, and respond to each other appropriately at all times, believability is maintained. Hence if a character has an acceptable level of realism and a consistent streak of personality traits, users in a VE will respond positively. [Bates *et. al*. 1992a: (13)] [Perlin *et. al*. 1996□: (56)].

Realism arises partly due to understated movements viz. breathing, blinking, random (or other) glances, and facial expressions. When human beings socialise these features (and more animated conversational gestures) increase the levels of presence and effective interaction experienced. Effective face-to-face conversation makes use of the visual channels for interaction where many subtle, intuitive and involuntary cues are read from body posture, eye-gaze, gesture, tone of voice and so on.

> *It is believed that innate behaviour programming in an avatar is necessary to portray realism in it especially if it is photorealistic. With the control variable being the emotional state of the mind (and perhaps the tone of voice), the general 'feel' of the animation should be autonomously controlled to a believable extent.*

An additional issue to consider is that behavioural animations have to appear individual to each avatar. This is especially significant if the avatar is part of a small group – where each one's behaviour is noticeable – in a crowd we don't really see individuals. Finally the amount of movement should depend on the general environmental situation. Is the avatar talking to a group, in a group or to one person? The problem arises when we try to interact with another user and the avatar is more closely scrutinised so general subtle movements are more noticeable e.g. eye gaze [(33)]. This view is applicable to both *Agents* and *Avatars* however the research this thesis is concerned with is mainly with the representation of real users in VR i.e. *avatars*.

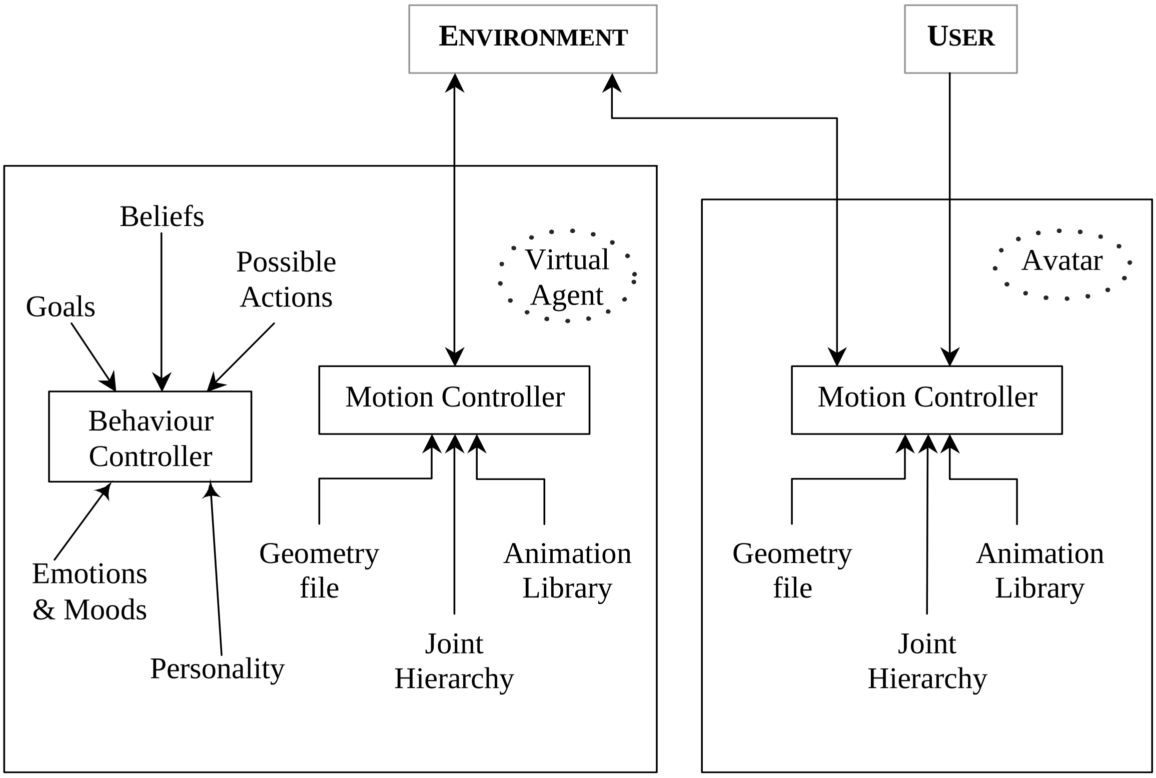


Fig. II Conventional differences between Avatars and Agents

Depicted above are the conventional differences between avatars and agents. The idea now is to incorporate a behaviour-modelling engine for the model of an avatar as well. There are hitches that will arise due to the extensive number of behaviour patterns to consider in modelling. Even if the exact behaviours needed to achieve such realism in an avatar are accurately deduced, the odds of accomplishing self-directed realistic behaviours in an avatar are somewhat trouble-ridden. If realistic behaviour is animated in run-time by programming trigger* mechanisms that detect changes in the immersed user and transfer then to the avatar†:

- Can it be assumed that introducing controlled autonomous movement to an avatar will bring a convincing illusion of "life" to it?

* Bumping into another avatar, recognising another user etc.
† Head and one hand position, tone of voice decoding, pulse measurements, general state of mind

- What kind of behaviour takes precedence in effecting the apparent realism in the avatar i.e. which behavioural characteristic is more important and why?
- Finally, how much behaviour realism is considered optimum before the animations become ineffective?

2.1 VIRTUAL ENVIRONMENTS

A virtual environment system consists of a computer system for generating VEs, human operators and a multimodal human interface for interacting with the synthesized virtual world. Virtual environments provide people with the power to create a world with complete control, bound only by the computer resources available and the creative limits of the designer. The use of virtual reality has been used extensively for training, creative design, communication, games, cinematic entertainment, social events, and being studied for psychotherapy [Slater *et. al*. 1999:(66);Slater*et.al*.È(65);Pertaub*et.al*.È(57);Pertaub*et.al*.È(58);Basdogan*et.al*.È(11)] .

Users of a virtual world share the same space, specialised to the needs of a particular activity, while maintaining awareness across it. It is a multi-user medium. For instance, there are many virtual chat communities, where people meet in an environment to socialise. Until recently these have been text-based, however some environments are going graphical viz. *active world, alpha world, the Palace, EverQuest, WorldsChat, Ultima Online* etc.

When users connect to such a system, they choose to be represented by an avatar that becomes their virtual *self* in the world. Users can then explore the environment, often from a first person perspective. The avatars of all other users, currently logged onto the system, can be seen and interacted with either by text or voice. Although these systems have now become graphically enhanced, communication is still mostly based on text or digitised speech streams. The graphics simply provide scenery and indicate the presence of a user. Some worlds have incorporated voice based communication but the avatars representing the user appear emotionless and without personality aside from some basic pre-programmed random movement.

The same extent of warmth got from a face-to-face conversation in the real world is not achieved*. This is changing with the advent of endowing avatars with non-verbal communication (NVC) skills [Masip 2002x: (46); Patil 1999~: (54); Mole 2002{: (50); SPARC 1998†: (70); (30); Gratch *et. al.*

* Personal Opinion

2001s: (36); Bates 1994f: (12); Grassia 2000r: (35); Salem *et. al*. 2000: (64); (46); (54); (50); (70); (30)].

2.2 VIRTUAL HUMANOIDS

The original meaning of the word *avatar* (in Sanskrit*) is '*the incarnation of a Hindu deity*' in the form of Vishnu. It is perhaps not strange then that it has become common to use this word to mean the representation of a user in virtual reality (VR). Avatars are a three dimensional representation of a user in cyberspace. They are also called Virtual Humans, Virtual Characters, Virtual Actors, Synthespians, E-People, and Autonomous Actors etc.

2.2.1 AVATARS & AGENTS

There are conflicts as to the exact meaning of an avatar, for instance on the issue of whether to include representations of humans not controlled by real humans known as agents. So how does an avatar different from a virtual agent? Various characterisation of an avatar include:

- Virtual characters that look physically realistic to a human but are *fully* controlled by the user,
- Abstract models that do not look like a human but are controlled by the user as well,
- Programmed bodies with a live video and audio fed face controlled by the user,

Avatars aid the user to experience cyberspace by visually interacting with other users or other pre-programmed agents. On the other hand, autonomous agents are generally not controlled by end-users in a virtual environment.

2.2.2 OTHER USES

Apart from populating collaborative virtual environments, various uses of avatars are as computer representatives of people usable as substitutes for 'the real thing' in ergonomic evaluations of computer-based designs for vehicles, work areas, machine tools, assembly lines viz. Jack [EDS 1996o: (28)]. This provides an alternative design technique prior to the actual construction of the spaces. Another usage is for aesthetic purposes in advertising, presentations etc.

2.3 POSSIBLE AREAS FOR EXPLORATION WITHIN AVATARS

There are many areas in the behaviour of avatars that can be delved into. Some of which have been more studied than others. Some of which are more important in research-commercial terms than others. Some are more related to graphics and others are more interesting in a psychological sense.

* An ancient language of the Vedas and Hinduism; an official language of India now used only in religion

Facial Animation

- Facial Motion Capture & Facial Action Coding System (FACS)
- Morphing faces & Muscle modelling
- Basic Emotional Faces, Lip Syncing, Speech Synthesis, and Eye Gaze…

Social/Facial Communication (empathy)

- Emotion vs. Reaction
- Conversation vs. Non-conversational
- Outside vs. Inside
- Cognitive Modelling

Body Behaviour

- Methods of Control
- Emotion Modes
- Ethological Models
- Motion Transformation
- Gesture
- Online Social Behaviour
- Task oriented, Role Oriented, Online Training/Task Relevance
- Gait
- **Posture**

As have already been stated, the non-verbal behaviour chosen for research in this thesis is *Posture*.

3.0 PSYCHOLOGICAL SURVEY

> *Learning is acquired by reading books, but the much more necessary learning, the knowledge of the world, is only to be acquired by reading men, and studying all the various editions of them.*
> *Lord Chesterfield, Letters to his Son*

A lot of research into computer graphics and human computer interaction (HCI) has been undertaken in conjunction with psychological and clinical studies. This applies more fervently to the area of this research since the study of non-verbal communication in humans is more established in psychological studies than it is in graphics.

3.1 NON-VERBAL COMMUNICATION (NVC)

Non-verbal communication is also referred to as *body language*. The study of body language is a relatively new science uncovered in the last few years. Both its theoretical study and its scientific basis have been labelled *kinesics*. Some works state that non-verbal communication is not the same as non-verbal behaviour in terms of the first being voluntary and the later is not. However in this study, both are used interchangeably as in the virtual world they are both programmed into an avatar with no consciousness.

Through communication on a relational level, the meaning of a message or a relationship to another person can be clarified. This can take place with or without words and is termed *meta-communication*. Non-verbal communication is a form of *meta-communication*. Of course non-verbal cues may contradict the message sent via verbal communication for instance in the case of deception or sarcasm. A person's body language is important in social situations and is sub-consciously read under almost every type of human interaction. The pressure of body language can especially be felt in emotional situations usually leading to body language prevailing over words. For instance bodily cues are the most reliable of non-verbal signals to detect deception. People generally have less conscious control over non-verbal signals [(30)].

There are many models partitioning the different types of behaviour. The most widely suggests that up to 93% of communication is non-verbal (55% consists of body language, 38% is expressed through tone of voice and only 7% through words) [Mehrabian 1971: (48)].

Channels for non-verbal communication include [Badler *et. al*. 2000c: (5); Allbeck *et. al*. 2001a: (1)]:

- Facial expressions (smiles, nodes)
- Gestures (hand and arm movements)
- Body movements
- Posture
- Visual Orientations (eye contact)
- Physical contacts (handshakes, patting)
- Spatial behaviour (proximity, distance, positions)
- Appearance (including clothes)
- Non-verbal vocalisations
- Smell

These non-verbal signals can be categorised by how far people are consciously aware of sending and receiving communication signals [(3)]. Most non-verbal communications happen with the sender being mostly unaware of sending the signal and receiver being mostly aware of the communication. However in the case of posture the sender is *fully* unaware of the conveyance of a non-verbal signal whereas the receiver is fully aware of the signal and at times trained in the interpretation of the signal [(3)]. The meaning of these signals to a sender can be found in the emotional state or message they intend to send.

3.2 POSTURE

As already stated in a preceding section, the body is thought to convey the degree of intensity of the emotions expressed by the face [(3)]. Without facial expressions, a forward lean of the body and head may be intended to convey a number of various emotions:

- Portray interest in a presentation that might accompany this report,
- Portray pity towards an audience member in the presentation for having to listen to it,
- Even perhaps personal anger for having wasted time in attending the presentation, or
- Frustration while finding it difficult to comprehend the content of the presentation…

So is posture alone sufficient to decode the right message being sent by the owner or is posture only useful if accompanying a basic facial expression. The range of stable human postures is about a thousand according to the anthropologist [Hewes 1957: (37)] who studied the postures used in different human cultures [(3)]. Some postures are discussed in [Rosenberg *et. al*. 1965: (63); (3)].

3.2.1 TYPES OF POSTURE

The complexity of variations between postures is not know at this stage in this research hence it has been decided to concentrate on standing poses initially and move to other stances if time permits. For the ease of comprehension and programming, the number of postures is divided into a few groups.

- Standing
- Sitting
- Squatting
- Kneeling
- Lying

Each of these then has further variations corresponding to the different positions of the limbs and lean of the body.

- Lean *Forwards, Backwards, Sideways*
- Arms *Open, Closed, On hips*
- Head *Lowered, Raised, Tilted sideways*
- Legs *Stretched, Open, Crossed*

All non-verbal messages (emotion or definition), being sent using a body pose will result from a combination of the aforementioned variables in varying intensities. In order to perceive the different effectual postures used by human beings in everyday life, it is necessary to decipher the messages that are sent using body posture as a major means of non-verbal communication. It is also worthwhile to postulate that a realistic avatar would require less exaggeration on the part of acting out the animation than a simplistic avatar.

In terms of research to expand on the knowledge specific to postures exhibited due to certain human emotions, there is much more to be done. Since this type of research is not wide spread in the computer graphics community, it can only be obtained form psychological archives. In terms of postural research this is the next step.

3.3 EMOTIONS AND MOODS

Expressing emotions is a form of body communication. The *emotion* of a person is the cumulative result of recent events and knowledge. Emotions are fluctuated hence more related to events but a prolonged, conscious state lead to a *mood* or an *emotional state*. Since this research is related to the expression of emotions and not the generation/detection of emotions, it will not be dealing with the detection of emotions in goal-directed events or agent systems. The avatar developed will be used to express different types of emotional states using body postural animation. There is some work into the subject area of using emotions as variables in animation: [Amaya *et. al*. 1996: (2)] [Kshirsagar *et. al*. 2002u: (40)].

3.3.1 MODELS OF EMOTIONAL STATES

One of the emotion models, called the OCC* model, categorises emotions into 22 groups [Ortony *et. al.* 1988}: (53); Niu *et. al*. 2002: (52)]. The OCC model was used famously in the OZ project discussed under section 4. The model is categorised according to the positive and negative reactions to events, actions and objects. Eleven of these expressions are positive and the remaining emotions are negative.

There are six different basic facial expressions defined in [Ekman 1982: (29)]. These are joy, surprise, disgust, fear, anger, and sadness. As can be observed, there are some overlaps between these expressions and the emotions modelled in the OCC model. However the concepts of surprise and disgust are not covered in the OCC model. All the emotions in the OCC model can be matched up to one of the Ekman basic expression.

Joy	*Happy-for, Gloating, Joy, Pride, Admiration, Love, Hope, Satisfaction, Relief, Gratification, and Gratitude*
Sadness	*Resentment, Pity, Distress, Shame, Disappointment, and Remorse*
Anger	*Anger, Reproach, and Hate*
Fear	*Fear, and Fear-confirmed*
Surprise	
Disgust	

The computation necessary to incorporate all 22 expressions from the OCC model plus surprise and disgust (i.e. 24 emotions) is complex especially since it is not known at this stage if the expression of emotions have to be thoroughly meticulous in order to be effective. Yet in real life they are important for making conversational exchanges rich with expressions. So is it presumptuous to assume that all the compound emotions described in the OCC model can be re-enacted using the

* A. Ortony, G. Clore and A. Collins

basic emotions in different proportions/intensities? Such work has been researched into already to some extent [Velásquez 1997‹: (79)].

Each of these emotions leads to either a good or a bad mood. Prolonged *joy* leads to a good mood whereas prolonged *anger*, *fear*, and *sadness* lead to a bad mood. Disgust and surprise may lead to either mood or remain in a neutral mood. There is a third expression, which may be categorised, has an emotional state: *interest*. The OCC model or Ekman has not covered *interest*. It is an emotion, which is ambiguous as well, in concern to what type of mood it leads to in an extended situation.

Hence in any case, a third mood (the neutral) may be relevant in the ongoing research to encompass the control group. In a neutral mood the person will tend to hold back on displaying emotions readily and if expressed, it is done so with less intensity. Whereas in a bad mood they might react with more intensity to the emotion of *anger* than they would if they were in a good mood prior to the input of the stated emotion.

3.3.2 EMOTIONS VS. PERSONALITY

Emotion is an area, which has been and is still being constantly studied, categorised and re-categorised. There is a dual hypothesis that emotion is a brief, focused change in personality or that personality is a permanent and global emotion [Baldes 2000d: (8)].

Just as non-verbal behaviour is affected by emotions and emotions have an influence on non-verbal behaviour; personality and non-verbal behaviour are interdependent as well. Personality, emotions and non-verbal behaviour are a three-way exchange. Models of personality exist as well such as the five-factor model [McCrae *et. al*. 1992y: (47)].

The five-factor model consists of five variables in the dimensions of personality space. These are extraversion, agreeableness, conscientiousness, neuroticism and openness. Extraversion denotes how talkative and outgoing a person is. Agreeableness relates to how friendly the individual is. Conscientiousness is how methodological a person is. Neuroticism is how insecure and emotionally distressed a person is. Openness relates to how imaginative and creative the individual is.

Nevertheless dealing with a computational model incorporating all three factors into a virtual human will take a much longer time and introduce more external factors. Work is being done in giving personality and emotional behaviour to facial expressions of virtual humans [(40)] but this shall not be incorporated in this research.

3.3.3 EMOTIONS VS. TONE OF VOICE

The only reason tone of voice has been included in this review is because it is a potentially good way to input an inference of what the user's emotion is into the avatar. Some attempts have been made into classifying the different elements in speech/voice to match the emotional state of the owner [Davitz 1964: (27); Collier 1985: (26)].

The different components of vocal expression are loudness, pitch, timbre, rate, enunciation, rhythm, and inflection. The matched feelings are affection, sadness, boredom, satisfaction, impatience, cheerfulness, joy, and anger. As can be observed from the tables some of the cells are blank due to lack of sufficient coherent information [(8)].

Joy

Loudness	***Pitch***	***Timbre***	***Rate***	***Enunciation***	***Rhythm***	***Inflection***
Loud	High	Mod. Blaring	Fast	…	Regular	Upward

Sadness

Loudness	***Pitch***	***Timbre***	***Rate***	***Enunciation***	***Rhythm***	***Inflection***
Soft	Low	Resonant	Soft	Slurred	Irregular pauses	Downward

Anger

Loudness	***Pitch***	***Timbre***	***Rate***	***Enunciation***	***Rhythm***	***Inflection***
Loud	High	Blaring	Fast	Clipped	Irregular	Irregular Up & Down

Since most behaviourally animated avatars have been either in a text-based conversational interface or in a gesture-recognition-based environment, the emotional expressions have been chosen based on language decoding or gesture recognition.

There have been recently developed agent systems, which recognise the context of speech input but these exhaust a lot of processor resources. Immersive systems demand that the use of keyboard input at run-time is abandoned so recognising and decoding the varying elements in the voice of the owner may be a worthwhile route of exploration.

There has been some work into the recognition or generation of emotional expression in speech [Cahn 1990: (17); Pelachaud *et. al.* 1994: (55); Picard 1998: (59)].

4.0 EXISTING VIRTUAL HUMANS & THEIR CREATORS

> *The computer can't tell you the emotional story. It can give you the exact mathematical design, but what's missing is the eyebrows.*
> *Frank Zappa*

Conventionally, research into avatars has not been a study in itself. Most avatars in existence have been created as a by-product to serve as either an aesthetic or a user-presence indicator within a wider project. A few prominent research groups have begun to study the generation of suitable virtual humanoids. Even so most of these avatars fail to generate behaviour in run time that can be ported across platforms and research groups. In relatively recent years, various attempts have been made to generate realistic scripting systems, avatars, or autonomous agents [(40)] [(5)].

4.1 FOREWORD

Computer Animation is generally categorised under the field of computer graphics. Typical applications of computer-generated animation are found in the fields of entertainment (motion pictures and cartoons), advertising, scientific and engineering studies, training and education. In particular, *articulated figure animation* has become popular in recent years because of the desire to use human beings as synthetic actors in three-dimensional virtual environments viz. the feature length movies Antz, Toy Story, Shrek, Final Fantasy, Steven Spielberg’s *Jurassic Park* and *The Lost World*, George Lucas’s *Star Wars* the increasing number of commercials etc.

The computer animation of articulated figures is an extremely painstaking process. Various algorithms have been developed to aid figure animation and much research has been done into quick low cost rendering of the physical attributes of the figures and characters. In recent times, research has been diverted to study the possibility of extending animation techniques to render animation based on the emotions, beliefs, personality, mood, goals etc. of the character.

This research has been built on previous and currently running projects. Even though this particular study has not been done before, results and deficiencies from other systems have provided valuable preliminary ideas. In the following there are a number of work and groups detailed. More information, good points and critiques have been made in the review for systems of particular interest viz. BodyChat [Vilhjálmsson 1996Œ: (81); (82)], Rea [Cassell *et. al*. 1999a: (19); Cassell *et. al*. 1999c: (22); Cassell *et. al*. 2000l: (21)], Oz [(13)], SimHuman [Vosinakis *et. al*. 2001Ž: (83)], InViWo [Richard *et. al*. 2001,: (62)], Cathexis [(79)] and Improv [(56)].

4.2 UNITED STATES OF AMERICA

4.2.1 MASSACHUSETTS INSTITUTE TECHNOLOGY (MIT)

There are numerous groups at MIT researching into agents. The groups include:

- Robotic life
- Synthetic characters
- Gesture and narrative language
- Artificial intelligence
- Software agents

Their two main recent agents are *Rea*[*] and *BodyChat*. There are other predecessors to these agents viz. Max, BEAT, Gandalf etc. Creating fully autonomous agents capable of natural multi-modal interaction at MIT entails integrating speech, gesture and facial expression. Applying knowledge from discourse analysis and studies of social cognition has been used to develop more systems at MIT viz. BEAT [Cassell *et. al*. 2001n: (23)], BodyChat [(81); (82)], Gandalf [(77)], ALIVE [Maes *et. al*. 1995: (45); Maes 2002: (44); (78)] and more recently Rea [(19)],[(22); (21)] discussed in subsequent sections.

4.2.1.1 BodyChat

BodyChat is a system that allows users to communicate via text while their avatars automatically animate attention, salutations, turn taking, back channel feedback and facial expression as well as simple body functions such as the blinking of the eyes [(82)]. By treating the avatar as a communicative agent, a method is proposed to automate the animation of important communicative behaviour, deriving from work in context analysis and discourse theory. BodyChat consists of a Client program and a Server program. Each client renders a single user's view into the Distributed Virtual Environment. All the users connected to the same server see each other's avatars as a 3D model representing the upper body of a cartoon-like humanoid character. The cursor keys are used to navigate the avatars. The mouse is used to give command parameters to the avatar and interaction is carried out using text messages through a two-way chat window.

[*] Real Estate Agent

In BodyChat, the avatar is treated as an autonomous agent in a world inhabited by other avatars. The autonomy is limited to a range of communicative expressions of the face and head, leaving the user in direct control of navigation and speech. Appropriate behaviour is based on the current situation and user input.

The system demonstrates the automation of communicative behaviours in avatars. The avatar reacts to an event by selecting the appropriate *Conversational Phenomenon* that describes the new state, initiating the execution of associated *Communicative Behaviours*. Each Conversational Phenomenon viz. salutation is associated with a set of Communicative Behaviours revealing the state to other users viz. the looking, head tossing, waving and smiling. A prospective communicative behaviour triggers an animation engine that manipulates the avatar geometry in order to change the visual appearance.

Fig. III *The avatars used*

BEAT: Behaviour Expression Animation Toolkit

This is a system built under the same underlying theory as *BodyChat* by using typed language analysis to generate appropriate gestures. One difference is that in the *BEAT* system it is possible to animate the agent with extra tag language. Essentially it is a whole language based system for animating agents from desktops.

The system can be plugged into their larger existent system. More on the system can be got from the paper published on BEAT [(23)]. More systems were built on a similar foundation viz. *Gandalf* [(77); (78)].

Good key points

There has been an effort to keep separate engines for different tasks. There are no menus involved so the users do not have to keep thinking about an interface to express themselves. The animations are autonomous.

One way of automating the avatar behaviour is to use trackers to map certain key parts of the user's body or face onto the graphical representation. The avatar imitates the user’s motion. However, in non-immersive settings, it shares a similar problem as in video conferencing. The user's space is different from that of the avatar. This flaw becomes apparent when multiple users try to interact. The gaze pattern and orientation information gathered from a user looking at a monitor doesn't map properly onto an avatar standing in a group of other avatars. Turns out that tracking may be more appropriate for VR applications where head mounted display (HMD) are employed.

Critique

Discourse analysis refers mainly to the linguistic analysis of naturally occurring connected speech or written discourse. It involves the study of the organisation of language above the sentence or above the clause, and therefore to study larger linguistic units, such as conversational exchanges or written texts. Discourse analysis is also concerned with language use in social contexts, and in particular with interaction or dialogue between speakers.

Discourse analysis is ambiguous. The analysis does not take into account of cultural differences in gestures. More importantly, despite its successful usage in desktop applications, typing messages is not acceptable in immersive settings especially if the user is already wearing a Head Mounted Display to view the VR.

4.2.1.2 Rea: Real Estate Agent

This system is slightly different from the other systems in that it responds to speech and not typed input. *Rea* is an embodied conversational agent that is capable of both multimodal input understanding and output generation in a limited application domain. She supports social and task oriented dialogue with the context of a virtual real estate agent.

Fig. IV Rea in virtual flesh

Rea is viewed on a large project screen. The system is able to sense user passively through the use of two cameras mounted on top of the screen which track the user's head and hand positions in space. The user speaks to *Rea* with the aid of a microphone. A single SGI computer runs the graphics and conversation engine for *Rea*. Several computers manage the speech recognition, speech generation and image processing.

Rea is able to use this input and conjure up a response to a user's statement, gesture, and questions. She allows interruptions in her responses and also is able to take turn again where she left off. Also Rea is able to initiate conversation repair if she misunderstands the user. She is able to respond to the user using speech, hand gestures, body postures, eye gaze and facial displays.

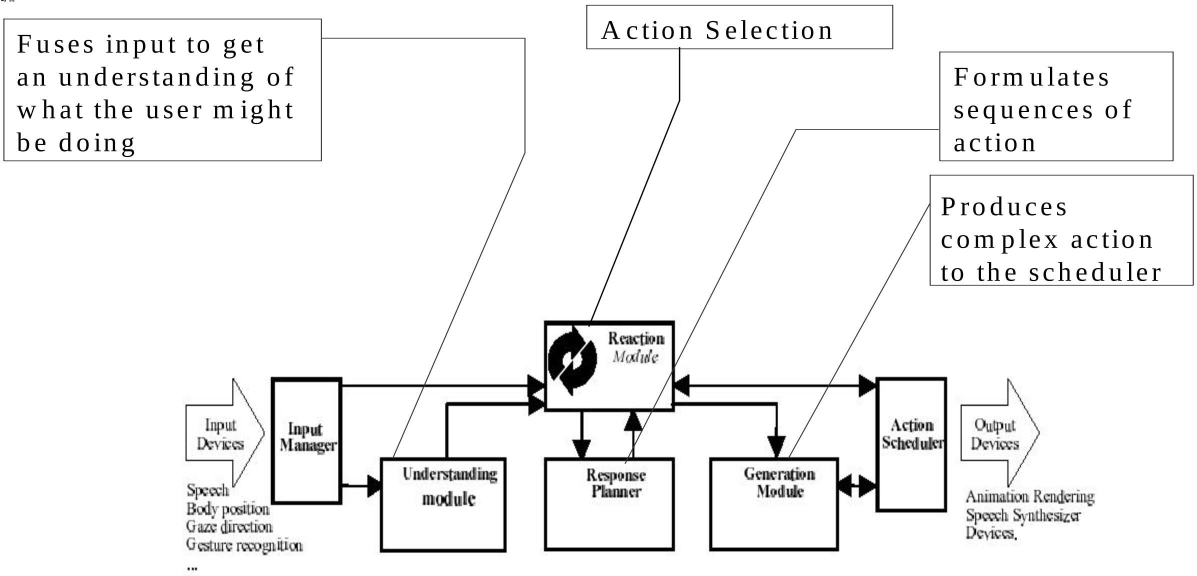


Fig. V   Rea software architecture

The figure above shows how this system was implemented. More on *Rea* is available in the many papers that came out in this project [(19); (22); (21); Cassell *et. al.* 1999b: (20); Bickmore *et. al.* 2000: (16)]. This system is well made to suit the purpose.

Good key points

The way the user's position is tracked passively is useful since one of the main barriers stopping a user from getting fully immersed is the weight and cumbersome designs of Head Mounted Displays, trackers and joysticks.

4.2.1.3 Cathexis

Cathexis[*] is an older model from MIT and hailed from the *Artificial Intelligence Laboratory* in 1997. *Cathexis* is a distributed computational model offering an approach to model the dynamic nature of different affective phenomena (emotions, moods and temperaments). It is a distributed model for the generation of emotions and their influence in the behaviour of autonomous agents. the model is tested on a test-bed in *Simón the toddler* in C++. More on the system Cathexis is available from the paper [(79)].

Good key points

- The model is backed with psychological research.
- It resulted in a computational model, which was tested on a test bed independent of the building of the model.

[*] Concentration of emotional energy on an object or idea

- There is the presence of an emotion decay function dependent on time.
- There is an intensity assigned to the emotional state.
- More than one emotion may be active and none are in control of the system.
- There are basic emotions and then a blend of mixed emotions.
- The emotions influence the behaviour of the agent.

Critique

The emotions used are not enough in number. Only the basic emotions [(30)] are used. The groups at MIT have used some physiological measures to control the emotions of some agents. However these physiological measures are not dependable.

4.2.2 CARNEGIE MELLON UNIVERSITY (CMU)

The school of computer science in conjunction with the Drama and English departments, at CMU produced the *Oz* project, which studied the construction of artistically effective simulated worlds. It resulted in an architecture called *Tok*, which supports reactivity, goals, emotions, and social behaviour. The world includes moderately competent, emotional agents.

4.2.2.1 Oz Project: the edge of intention

The *Oz* world contains four primary components:

- A simulated physical environment
- A set of automated agents populating the world (*Wolf, Shrimp, Bear*)
- A user interface allowing users to participate in the world (*Woggles*)
- A planner for the long-term structure of the user's experience

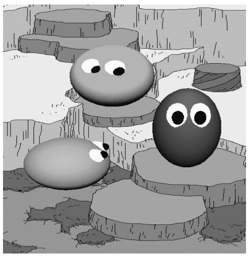


Fig. VI The Woggles

One of the keys to an artistically engaging experience in the *Oz* world is for the user to be able to *suspend disbelief*. This is because the agents inhabiting the world are not even close to humanoids. They have a broad set of tightly integrated capabilities even if the capabilities are not sophisticated. The capabilities include goal-directed reactive behaviour, emotional state and behaviour, social knowledge and behaviour, and some natural language abilities.

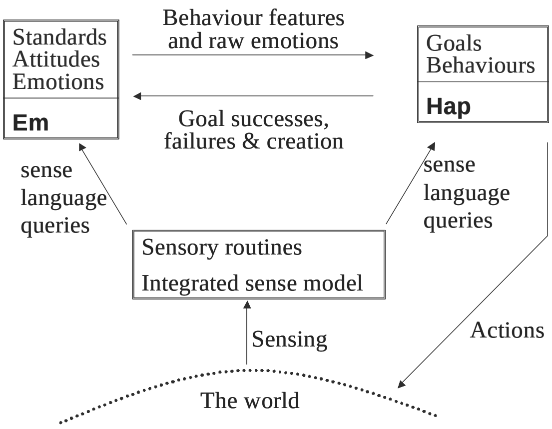


Fig. VII The Tok Architecture

Primary capabilities of the agents are perception, reactivity, goal-directed behaviour, emotion, social behaviour, natural language analysis, and natural language generation. The agent architecture *Tok* assigns these tasks to several communicating components.

• Perception	Sensory routines and the integrated sense model (ISM)
• Reactivity and goal-directed behaviour	Hap
• Emotion and social relationship	Em
• Language analysis	Glump
• Language generation	Glinda

Many papers are available on the Oz project [Loyall *et. al.* 1993w: (43); Bates *et. al.* 1992b: (14); (13); (12); (43); (14); (13)].

Good Key Points

Even though the agents are not as sophisticated as needed for this thesis, the architecture is one of the strongest seen so far. Every sentient in the architecture is independent and was developed as a separate puzzle in the big picture. It is advantageous to have an avatar with a similar architecture in order to able to tailor the avatar to various needs.

- A sensory model: Agents sense the world via sense data objects, which propagate from the item sensed through the world to the agents.
- All sensory data collected are recorded in the sensory routines, an attempt is made to merge them into the *Integrated Sense Model (ISM)* maintain the agents best guess about the physical structure of the world. The *ISM* is regularly updated.
- *Hap* continuously chooses the agents next action based on perception, current goals, emotional state, behavioural features and other internal states.
- All active goals and plans are stored in an *Active plan tree (APT)* representing the *Hap* current execution state. The *APT* expands and contracts as goals and plans succeed and fail.
- Each goal has a priority status and an importance status issued by *Em*.
- *Hap* does not do any explicit planning.
- *Em* models emotion and social aspects based on Ortony et. al. [(53)]
- Emotions are generated by *Em* dependent on whether goals are succeeded or not. This information comes from *Hap*. Not all the goals generate emotional reaction. Most of the goals are of zero importance and hence does not produce a reaction.
- Emotions acted are hope, fear, pride, shame, reproach, admiration, anger, gratitude, remorse, gratification, love and hate.
- Emotions fade with time and this is modelled as decay in emotional state.

Critique

The main issue in this architecture other than lack of sophisticated behaviour models is the lack of communication of the emotions from the *Em* entity to the *Hap* entity, which means there is no dependency in that direction.

4.2.3 NEW YORK UNIVERSITY (NYU)

The group at NYU dealing with virtual worlds is based at the *Media Research Laboratory*. Ken Perlin et. al. published a paper based on their system, *Improv* [(56)].

4.2.3.1 Improv

Improv was developed to create real-time behaviour-based animated actors. It provides tools to generate actors that respond to users and to each other in real-time, with personalities and moods consistent with the author's goals and intentions. There are two sub-systems involved:

- The *Animation* engine enables authors to create layered, continuous, non-repetitive motions and smooth transitions between them.
- The *Behaviour* engine enables authors to create sophisticated rules on how actors communicate, change and make decisions by giving the actor a set of possible behaviours to choose from.

The paper on Improv [(56)] discusses the problem of building believable characters that respond to users and to each other in real-time with consistent personalities, properly changing moods and without mechanical repetition, while always maintaining an author's goals and intentions. Actors follow *scripts*[*] used to determine the appropriate animated *actions* to perform in time. There is a behavioural architecture to support author-directed multi-actor coordination as well as run-time control of actor behaviour for the creation of *avatars*.

Good key points

There are many good traits in the system, which can be imported into an avatar plugin in the future:

- An actor can do many things at once and these simultaneous actions can interact in different ways.
- Authors make distinctions between different actions, which can in the real world co-exist gracefully or not, using a set of *rules*.
- Motion is a layered affair with weights assigned to actions. So if the actor is walking and decides to wave an arm. The weight on the walking activity reduces and the weight on the waving activity increases allowing a smooth transition.

[*] Author defined rules governing behaviour

- Actions are put into groups.

Actions in the same group compete with each other. Actions in a group further below obscure those in groups further back. Example: An actor is not allowed to wave with his left hand and touch his ear with the same hand simultaneously.

- Transitions between some actions in a group are awkward resulting in an un-naturalistic animation.

For instance the hands passing through the body when an actor is retrieving the hands from a folding position behind the body into a position in front of the body. To prevent this the author declares a "*buffering* action" in this case perhaps the action of keeping the hands at the side frame. This forces the hands to be taken from the back to the sides and then to the front.

- The behaviours are triggered by scripts, which contain several items to choose from. Generally more than one script is run at any one time in parallel.
- Probability shaping tools are provided to guide the actor through the author's choices.
- The scripts are also organised into groups. Choosing one script immediately halts the running of any other script in the same group.
- There are weights attached to behavioural choices as well.
- The actors can have Personalities and moods
- There is an API for "laypersons"
- Perlin Noise gives the animation a *jittery* effect hence reducing the actors resembling robots.

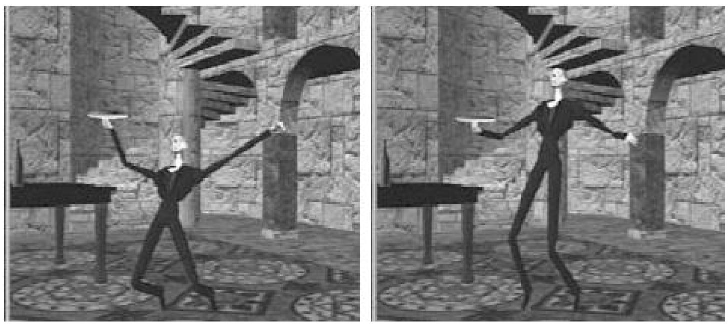


Fig. VIII Two results from the same script on a different networked computer.

Critique

There is too much choreography involved in the system and not enough changeable optional dependent variables. Even though the system could give out a good agent. An avatar should be more independent of scripts.

4.2.4 UNIVERSITY OF PENNSYLVANIA (UPENN)

Professor Norman Badler has been engaged for a long time in human body modelling and simulation. Much of his work at the University of Pennsylvania has centred on the Jack software, widely regarded as the world's most advanced and versatile commercially available human modelling system. This is now a commercial product maintained by Eds [Badler *et. al.* 1999: (6)].

4.2.4.1 Jack and Jill

These are bio-mechanically correct models in which the human factor can be tested. They are used widely in the automobile industry to check if humans can access all the parts of increasingly compact engines. It is used mainly in the design phase. The system uses inverse kinematic solutions to animate the agents. It contains kinematic and dynamic models of humans based on biomechanical data and features balance, reaching, grasping, walking, and running. Jack's capabilities include complex articulated motion, with balance-aware motion modification; collision avoidance; gesture and facial expressions; goal-based tasking; natural language processing, and many other features. It is used for a wide range of applications, including industrial ergonomic testing, military simulation and training, and human factors research [(28)].

Fig. IX Jack and Jill

Good key points

The use of inverse kinematics is very advantages especially in the design of avatars, which are partially tracked.

4.2.4.2 Embodiment of agents

Overall the Jack and Jill agents were not made to inhabit virtual worlds so there was a lack of embodiment. Currently the work at UPenn has shifted to include the emotional and personified factor into its agents.

They have brought up some issues to consider in building a framework of consistent behaviour models [(5); (1)]. People are more likely to notice the unexpected than the more natural or expected. When an agent’s behaviour conforms to nominal expectations, they are termed *believable*. This means it is accepted as *real*. One only has to look at cartoons that communicate presumed thoughts and feelings to see that reality in expression is the stronger determiner of believability [(5)].

Mechanical speech, and mechanical walk can distract the eye from viewing subtle facial expressions leading to the hypothesis of this research that behaviour realism is expected to accompany a photorealistic avatar. The paper [(5)] concentrates on non-verbal communication channels. Another paper that deals with the framework of the embodied agent is [(1)].

More recently a paper was published on the behaviour of eye-gaze in a dyadic situations. This was used extensively in the eye-gaze experiments, which are part of this thesis. They are detailed in sections 5.0 and 6.0. More on the eye-gaze model is available in the paper [Lee *et. al.* 2002v: (41)].

4.3 EUROPE

4.3.1 MIRALAB, UNIVERSITY OF GENEVA & VIRTUAL REALITY LABS, EPFL

MIRAlab and EPFL have done a lot of work together hence are discussed under the same section. They have done extensive work over a long period of time over a wide variety of areas surrounding avatars and agents. They have done a lot of work in the real time animation of virtual humans [Thalmann *et. al.* 1998ˆ: (76)].

In MIRAlab, the most recent and relevant research area is in empowering their virtual humans with personalities and identities. They aim to create virtual humans that can interact using emotions, natural language and gestures. The paper [(40)] discusses a system, which allows the design of emotional virtual humans with personality. They use the five-factor model [(47)] of personality from psychological studies. They discuss an approach for modelling personality, moods and emotions. They discuss the idea of *personification*, the attribution of personal qualities and representations of ideas in the humanoid.

There are four types of personification: physical, expressional, logical and emotional. Physical personification refers to the humanoid's appearance. Expressional personification refers to the design of emotional gestures and facial expressions. Logical personification tells the humanoid on how to analyse and interpret input. Emotional personification controls the way a humanoid's emotions evolve over time. The humanoid also has the concept of moods and emotional decay over time. Another concept researched in this paper is that of the Bayesian Belief Networks (BBN). BBN is used to model domains with uncertainty with the aid of a conditional probability table. The emotional tags have an associated probability value.

In EPFL, recently they have developed an *Intelligent Virtual Agent (IVA)* which is to be plugged as the brains into virtual humans. IVAs are based on the existent *Beliefs, Desires and Intentions (BDI)* architecture [Rao *et. al.* 1991: (61)]. To the existent architecture, EPFL have added a short and long term memory of beliefs, internal states of basic emotions, reliability on trust, and the inclusion of emotions in the plan's structure and in the evaluation plans. Apart from work in behaviour, there is also some work on networked collaborative environments [Capin *et. al.* 1998j: (18)] and virtual human society [(18); (76); Thalmann *et. al.* 2000: (75)] in which there is the concept of social sciences and social non-verbal communication.

Other by-products of their research have been in the production of various languages used to script out avatars and worlds viz. *Avatar Markup Language (AML)*, Virtual Music Markup Language, and Body Animation Markup Language.

Good key points

- The use of a set architecture, which can be expanded as needs increase.
- There is the use of a behaviour engine, which selects the actions that has to be carried out by the agent.
- There is an events interface, which passes the user's events to a module, which decides what to do with them.
- The division of personification into four distinct modules means that each module can be independently developed and activated separately.
- The use of uncertainty means that the personification can be made unique in order to exclude repetitive animations.
- There are not two types of humanoids but three… the fully tracked avatar, the autonomous agents and the user-guided virtual human (part scripted and part tracked).
- The avatars are H-Anim standard.

Critique

In relation to the amount of research done, there isn't much evidence of a lot of facts backed by psychological evidence or case study.

4.3.2 UNIVERSITY COLLEGE LONDON (UCL)

Much work has been done in social paranoia [Met 2001z: (49)], virtual acting [(65)], collaborative environments [Mortensen *et. al.* 2002|: (51)], crowds, and non-verbal behaviour [(33)] and haptics [Jordan *et. al.* 2002t: (39); (11); (39)]. Other works done at UCL are published in various papers [Slater *et. al.* 2000c: (68); Slater *et. al.* 2002: (69); Slater *et. al.* 2000b: (67); (66); (57); (58)].

4.3.3 UNIVERSITY OF SURREY (UNIS)

The university of surrey has relatively recently started doing research under 3D graphics. They are involved in modelling virtual humans and are involved with UCL in modelling virtual clothing algorithms. However their work does not extend to behaviour even though there are aspects of inverse kinematics and animation involved in their research [Sun *et. al*. 1999: (73); Hilton *et. al.* 1999: (38)].

Good key points

The avatars are very realistic but there is too much effort placed in creating photo-realistic agents.

4.3.4 ENST/INRIA/LIP6

The *InViWo (Intuitive Virtual Worlds)* project, based at France, aimed at producing high-level intuitive tools to describe virtual worlds populated with *intelligent* creatures and avatars. The project also came up with a language for the high-level description agent behaviours namely, *Marvin*. *Marvin* is essentially a language based on the primitive actions an agent can perform, reactive rules, temporal control structures and constraints. The architecture behind the agent is message based, decentralised (no one is in charge) and homogeneous (any object is an agent). The messages can be in form of unicast, multicast or broadcasted communication.

An *InViWo* world is essentially a set of agents with no specific structure. There are two special agents namely:

- An environmental agent manages the database containing public information sent by agents (current shape and position).
- A perception manager handles a list of agents to be called back when specific events occur to optimise sensor filtering.

An agent in the world is made of a set of attributes representing the characteristic of the agent (shape, position, or mood) and its knowledge (personal map of the world). It also contains sensors for perception, effectors to perform primitive actions, and a decision process.

An avatar within the world includes navigation and basic perceptive capabilities. There is also a camera to show the user around the world. A user can control several avatars in the same world. Several users in the same world can control an avatar. So there are multi-users and multi-user avatars. More on the system is available in the paper published [(62)].

Critique

The concept of an avatar being used by many users is not completely clear. The use of many avatars allows for different viewpoints within the environment.

4.3.5 UNIVERSITY OF PIRAEUS

This university is based in Greece and the laboratory, which created the system concerned, is the *Knowledge Engineering Laboratory* in the department of informatics. The system is called *SimHuman* [(83)].

SimHuman is a platform that allows the user to define 3D scenes with virtual agents and user-controlled avatars to serve as a basis for virtual environments and simulation systems. The system is not based on fixed scenes and models and has a physically based modelling engine. All objects in the scene are described in terms of 3D polygons. Agents are not based on a fixed model instead they are loaded dynamically from a *geometry file* containing details of their appearance. In addition a *hierarchy file* containing skeletal information including joint hierarchy, position and limits of joints and how they are to be transformed is loaded. The basis of all animation sequences is *keyframing*. A smooth transition between the poses is generated. Various poses are stored in the *animation library* and used to produce more complex animation sequences viz. walking.

A physically based modelling engine is used to increase the believability of the animation. The physical simulation is conducted in discrete time-steps. Each object has a mass, position and velocity and in each time-step the later two are recalculated following the laws of kinematics and collision. In addition, collision detection is used to detect a whole object or part of it within another since the object are supposed to be rigid bodies. It also determines the new position and velocity of the objects that collided.

The *SimHuman* platform generates a dynamic environment and hence the need for the agent to be able to execute dynamic actions brought up in run-time in order to interact with the environment is paramount.

Path planning and obstacle avoidance are included in the system. The agent has a sensing mechanism based on the principles of ray casting for this. This gives the agent an idea of the size, position and type of object in its field of view. The agent keep its own internal memory map of the scene that it keeps updated after every look operation. The *SimHuman* system uses a *planner* to control behaviour.

Procedures such as walking to a target, avoiding obstacles and grasping objects are not calculated at the high-level declarative part of the 'mind' but are part of the agent's functionality as standard actions. This way the planning time needed to execute those functions does not slow down the system's performance.

Good key points

A classic problem is the breakage of a joint in an articulated figure when a segment is rotated, due to the faces of the adjacent segments not adjoining anymore. One fix is to insert spheres in the middle but this distorts the model at times. In *SimHuman* the ends of the adjacent segments always connected and the common vertices are not transformed when a segment is rotated.

Bounding primitives around the human body segments, in the shape of cylinders, are used to perform collision detection. This reduces computational time and makes it possible to run in real-time systems.

The use of an alternative approach to inverse kinematic seems to work well in this system. Instead of a generic inverse kinematic solver, the system uses a different approach in which the best rotation for each joint to achieve the target is tested at each step. If *jittering* occurs in one of the segments (i.e. the angle is increased in one step and then decreased in the next step), the speed is reduced (i.e. the step is reduced) and on the other hand if the angle change is always heading to the target, the speed of the rotation is increased (i.e. the step is increased).

4.4 OTHERS

There is work being done on avatars in the Pacific, South Africa and South America but none that is relevant to this research. There is work being done under psychology on non-verbal communication in Spain [(46)].

There is also a system developed in the University of Bielefeld, Germany called *INSIGHT* [Strippgen 1998: (72)]. This is a 3D simulation environment, a virtual laboratory, which has been designed to complement experiments with mobile robots in complex environments. It supports the design and debugging of control concepts for behaviour-based autonomous agents and allows the study of effects of the systematic variation of agent environment parameters.

There has been more work on emotions and personality, which are not discussed [(35)] [(36)]. There is also work being done on the gesture control of avatars [Barrientos 2000: (9); Barrientos *et. al*. 2001: (10)]. The literature review continues not only into the psychological branches of non-verbal communication but also into the graphical side. Work into this area is widespread.

5.0 VISION, STRATEGY, TIME-SCALE >> HYPOTHESIS

That people's brains are programmed in such a way that the corners of the mouth are pulled up with joy, that the eyebrows are pulled up and that a corner of the mouth goes up according to the feeling that is fed to the brain.
Julius Fast

5.1 ROUGH GUIDE TO THE VISION

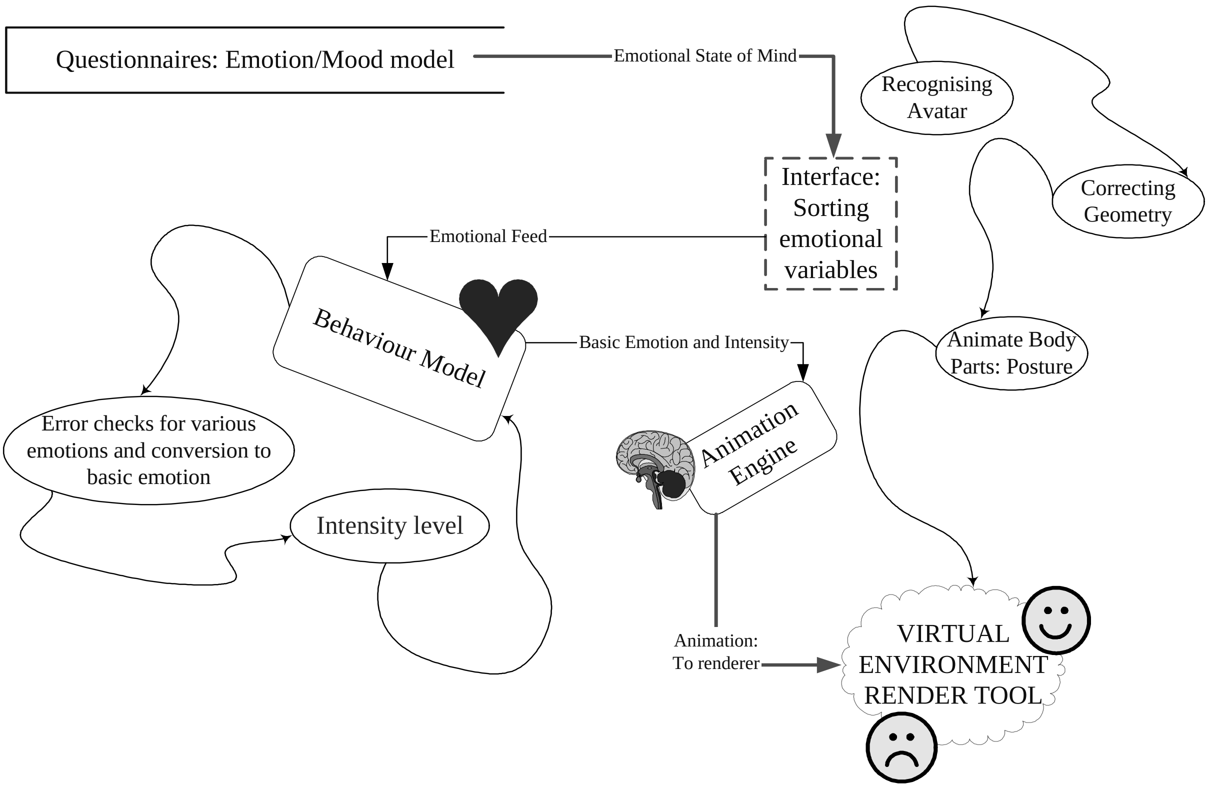


Fig. X Proposed Bender Architecture

The preceding figure (I) portrays a potential outline of how the workings of the avatar resulting from this research might look. The resultant avatar should be able to take a number of input variables corresponding to the supposed emotional state of mind of the avatar and feed it to the behaviour model. The behaviour model then converts the compound emotion (OCC model) to a simpler set of emotions (Section 3.3.1) and conveys it on to the animation engine.

The animation layer will then produce the appropriate standing, sitting, squatting, kneeling or lying poses appropriate to the given emotional mood with the aid of either a script (autonomous agent) or the position of the user (avatar).

5.2 STRATEGY: EXPERIMENTAL PLAN

Prior to the avatar envisaged (Section 5.1) being realised, three vital pieces of research have to be acquired and the plan to get these pieces of information are listed as follows.

- Sub experiment A:
  - Is behavioural realism the missing link between bizarre robotic humanoids and convincing realistic humanoids?
  - How effective is behaviour animation (eye-gaze in dyads) on avatars in comparison to avatars with no behaviour?
- Sub experiment B:
  - Which emotions are portrayed with the aid of body posture?
  - What are the poses corresponding to the emotions?
  - How intense should the emotions be in order to be conveyed?
- Sub Experiment C:
  - Does the system emulate these emotions across a virtual environment?
  - Are simple accompanying facial animations necessary to convey the right message?
  - How effective are the poses on a simple avatar in contrast to a photo-realistic avatar?

Some of this information necessary may exist for human beings in the real world but the same cannot be taken literally for simplistic avatars in a virtual world. This is probably due to the lack of photo-realism in the avatar and the virtual world. The following sub-series of experiments is to examine human postural behaviour, the influence of emotions on posture and concluding in a hypothesis (Section 1.3) that states:

> *Postural behavioural realism in avatars is a viable component in evoking levels of presence, co-presence and believability challenging those evoked purely by physical photorealism alone, the control variables being emotions* and moods†. The hypothesis is that increased behavioural realism leads to increased co-presence (other things being equal).*

* Momentary state of mind normally associated with an event

† Prolonged state of mind resulting from cumulative effect of emotions

This subject matter surrounding this thesis cannot be proved mathematically in front of a computer screen i.e. experimental evidence from *unsuspecting* subjects are necessary to validate this theory. Performance levels have to be measured with the aid of experiments. In the following sections are three possible experiments that could be undertaken to lead to the proof of the hypothesis, one of which has already been carried out.

5.3 SUB EXPERIMENT A: EMOTION & NON VERBAL BEHAVIOUR

The conception of this hypothesis was based on the presumption that behaviourally animated simplistic avatars evoke competitive levels of presence and co-presence in users of virtual environment as much as photo-realistically detailed avatars.

Due to postural and emotional models being too complicated (and lack of research knowledge at this stage), initial testing was reserved for more specific behaviour experiments, and were designed to testify a couple of preliminary sub hypothesises.

- Emotional states of mind can be conveyed across virtual environments
- Behavioural animation (eye-gaze) makes a positive impact in dyadic conversational situations

The reasons for the testing of these hypothesises were to validate (or invalidate) the following presumptions.

- Emotional state of mind can be conveyed across a Virtual Environment
- Behaviour modelling in an avatar increases realism than an avatar with no behaviour.
- Simplistic avatars can cause and effect user's level of presence
- Photo realism can be compensated for with behaviour realism

5.3.1 PROPOSAL: CONVEYANCE OF EMOTION [INTERNET-2]

The first part of the experiment was carried out in conjunction with the University of North Carolina, Chapel Hill (UNC-CH) as a by-product of a wider project to test Internet2. A collaborative virtual environment was built and two subjects were asked to carry out a virtual stretcher through a maze together.

The subject at University College London (UCL) was an unwitting ordinary subject while the one at UNC-CH was in actual fact a collaborator. The collaborator was told to role-play in alternative running of the subjects to act in exaggeration the emotional state of happy or sad. More on the experiment and how it was undertaken is detailed in chapter 6.

The hypothesis was:

> *Emotions can be detected across networks even with the aid of a simple avatar and an appropriate tone-of-voice.*

5.3.2 PROPOSAL: CONVEYANCE OF NON VERBAL BEHAVIOUR (NVB) [EYE-GAZE]

The later part of this sub hypothesis was carried out in conjunction with a final year PhD student, Garau, at UCL. It is an extension of an earlier published work, which was done on a desktop/video scenario [(33)]. Similar work has been carried out at the University of Pennsylvania [(41)] and results from the study have been included in this experiment.

The work postulates that the inclusion of eye gaze animations in dyadic situations enhances the user's sense of presence. However the earlier experiments failed to separate the effect of head tracking information from the eye animations.

The experiment proposed here is designed to infer the effect (if any) of eye-gaze animations on users independent of head movements. The experiments are also designed to ascertain if the level of photo-realism of an avatar matter. The design of the experiment is detailed in the ensuing section.

The hypothesis was:

> *Inferred eye gaze modelling in a photo-realistic avatar increases the users' sense of co-presence when compared to a similar avatar with random eye-gaze modelling (other things being equal).*

5.3.2.1 Design of Experiment (NVB)

Scenario: Two subjects were to be in a shared immersive virtual environment. The two subjects were asked to play a role in a fictional dyadic conversation to be undertaken for ten minutes. The scenario was similar to the role-playing activity use in the forerunner experiment [(33)].

Role-Playing: The aim of the scenario is to encourage two strangers in VR to have an involved chat for enough time to notice any oddities about the avatars but not too long as to make them bored.

The two roles invented were that of a mayor (Mr./Mrs. Carlis) and a baker (Mr./Mrs. Brent). The mayor's son, Mark, hears news from the baker's daughter, Carol, that she is pregnant. He promptly goes incognito. Understandably worried, Carol goes to her parents and spills out the whole story of how she loves Mark, the pregnancy and how she intends on keeping the baby. Mark has not offered any signs of marriage or help with any financial arrangements. The baker phones up the mayor and asks for a meeting after a briefly bring him/her up to date. The mayor in turn phones Mark who accepts that he has been involved with Carol but doesn't think that she has been exclusive in terms of relationship with him.

The experiment is described in further detail under section 6.0. They meet for the first time in the VR.

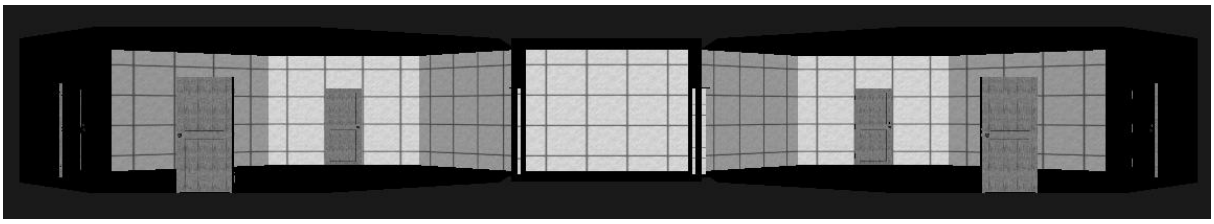


Fig. XI The VR environment

Methodology: Every effort was made to ensure that the collaborating subjects did not meet each other in the physical world before and during the experiment. This was to ensure that all interactions in the virtual world are initiated due to the experience in the virtual world alone. Also the subjects were gender matched to ensure the interactions were not affected by a difference in gender or viewpoint.

The levels of presence, co-presence and believability were established from questionnaires (pre and post experiment) and interviews.

Cases: There were four sets of experimental cases. These were:

- Photo-realistic avatar with inferred eye gaze animation (Jack/Nancy)
- Photo-realistic avatar with random eye gaze animation (Jack/Nancy)
- Simplistic Avatar with inferred eye gaze animation (Bender)
- Simplistic Avatar with random eye gaze animation (Bender)

The first two cases were to discern if accompanying photo-realism largely enhances the levels of presence experienced with a behaviourally realistic avatar. The third case was used in comparison with the first to ascertain if inferred behaviour was more effective than random behaviour. The fourth was compared with the second in the same way.

The stated two comparisons, in theory, provided information on if behavioural animation is important to improve the experience of users in the virtual environment. It also indicated if photo-realism is a necessity or a bonus in virtual environment if behavioural animation is utilised as well. The structure and design of the avatars (photo-realistic & simplistic) and the C plugin to animate the avatar is detailed in section 6.0.

5.3.3 EXPECTED RESULTS

The expected results were that the experience of the users would be enhanced if the more realistic avatars were accompanied with realistic inferred behaviour. In relation to the postulation through the whole thesis, it could be expected that a simplistic avatar with inferred eye gaze animation would perform better than a realistic avatar with random eye gaze animation. However the later postulation might not come through in the experiment since the simplistic avatar might need an exaggerated behaviour model to compensate for the lack of photo-realism.

Results are indicative of the photo-realistic avatar demanding a more realistic behaviour model. Whereas in the simplistic avatar a behaviour model does not make any difference in the levels of co-presence experienced in the VR. More on the experiment is detailed in section 6.0.

5.3.4 TIME-SCALE

The Internet2 trials have been carried out and published [(51)]. The second experiment has been concluded and analysed as well. The findings are to be submitted to CHI 2003. The total time for the undertaking amongst other things have been approximately a year.

5.4 SUB EXPERIMENT B: POSTURE VS. EMOTIONS

The second stage of this research diverges quite a bit from the first stage. As can be observed, the purposes of the first stage was to ascertain that behavioural animation of any sort causes the user in VR to experience relatively increased levels of presence and aids in the recognition the possible emotional state of any other collaborating user in the virtual world.

This stage is to be used solely to determine the different emotions recognisable from the avatar posture alone and the corresponding intensities necessary to reconstitute the non-verbal postural behaviour in VR so as to recreate the emotions intended.

The main assumption made here is that:

> *Some (if not all) emotions are portrayed with the aid of postural emphasis, some emotions with more zeal than others are.*

It is also a personal belief that:

> *Human beings use postural stamps more habitually to portray negative and neutral emotions than positive ones. There might be a cultural variant on this postulation.*

Hence subjects would probably find it easier to recognise negative/neutral emotions more readily than positive emotions using postural information alone. The introduction of facial information should improve the rate of success in recognising positive information.

5.4.1 PROPOSAL: POSTURE RECOGNITION

As detailed in section 3.2, the possible variations in body parts to create the stable postures used by humans in everyday life are as follows.

- Lean *Forwards, Backwards, Sideways*
- Arms *Open, Closed, On hips*
- Head *Lowered, Raised, Tilted sideways*
- Legs *Stretched, Open, Crossed*

A system enabling the variation of these individual body parts poses in varying intensities could be shown to subjects. The subjects could be asked to recognise and label set postures or asked to vary the body parts to depict what they see as the portrayal of a set of emotions.

The sets of posture could be extended from standing poses to sitting, kneeling, squatting or lying postures. For the sake of simplicity the kneeling and lying poses will be left out of the experiment and system since it is unlikely the subjects will do either. Plus even if the subjects do carry out these postures, a version of squatting can be used instead.

After all the emphasis throughout this study is placed on the visual consistency of the avatar not the accurate portrayal of the user's pose. The avatar is solely as a visual aid of fellow users in the system not the owner; in fact the avatar would probably not even be visible to the owner of the user.

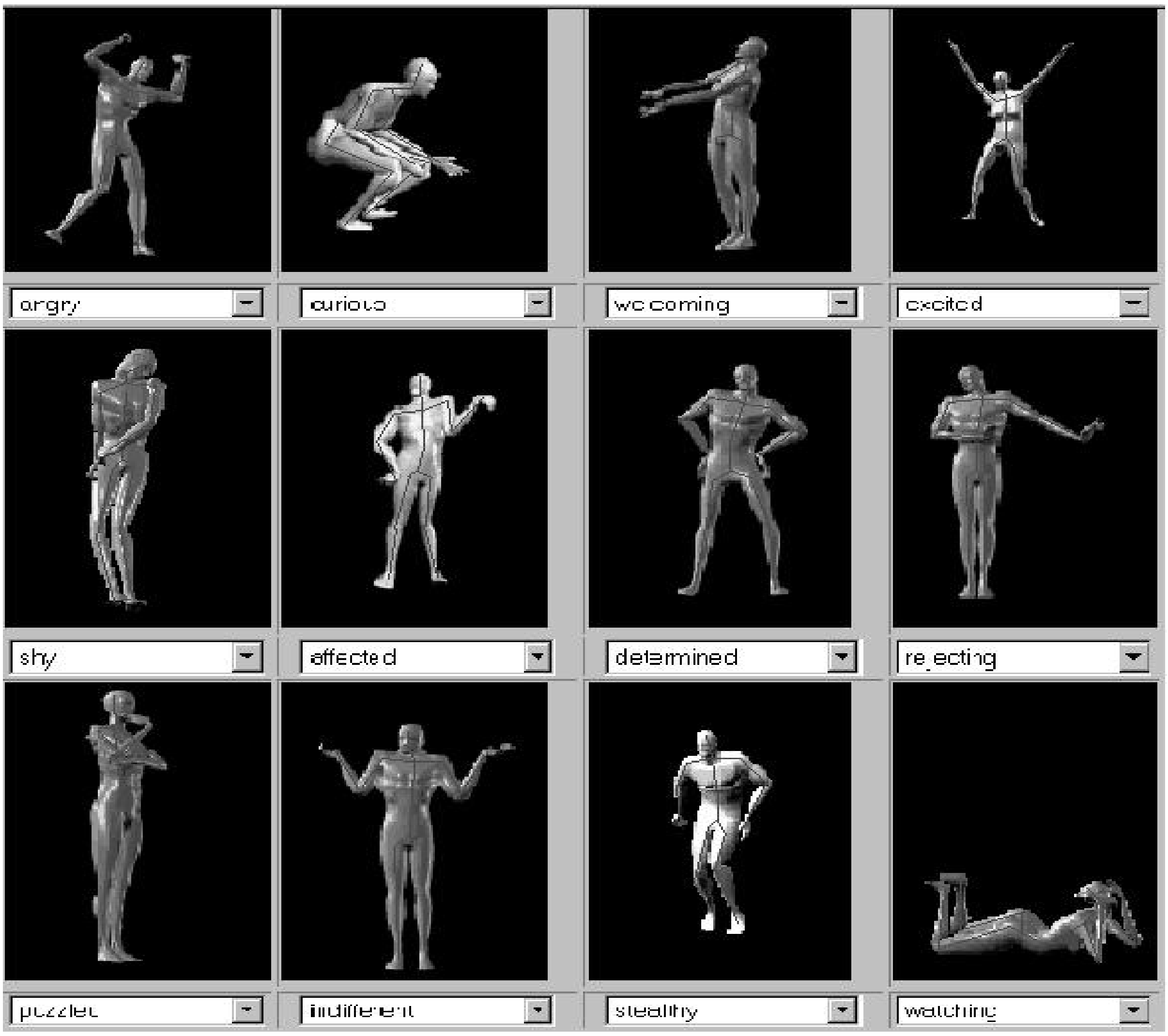


Fig. XII Some poses portraying emotional state

5.4.2 EXPECTED RESULTS

The results are expected to substantiate the assumptions and sub hypothesises stated in this section. More importantly are the resulting poses that are got from this stage. These poses will be fed into the next stage of the study to be automatically rendered to produce a set of poses for a given emotional state.

5.4.3 TIME-SCALE

This stage is expected to take just over a year however this is a guideline. This sort of study has been done as a part of the creation of a system [Polichroniadis 2000: (60)].

5.5 SUB EXPERIMENT C: FEELINGS & POSES

The third and final stage of this research is a direct resultant of the second stage. The poses determined in the second stage are recreated by the system to produce animations with accordance to the emotional needs of the avatar.

This stage is also an opportunity to retest the postures and also to ascertain if accompanying facial expressions are necessary for effective conveyance of the emotion. The testing of the main hypothesis will be carried out in this stage using four test cases similar to those mentioned under stage 1 (Section 5.3.2.1). Additional cases might include testing for difference in performances of the subjects if the posture is accompanied with an appropriate basic facial expression.

The assumption here will ***probably*** mean that:

> *Facial expressions will enhance the ability of users in VR to recognise postures and what they indicate in terms of the emotional state of mind of the avatar.*

5.5.1 EXPECTED RESULTS

The results are expected to substantiate the assumptions and sub hypothesises stated throughout this report.

5.5.2 TIME-SCALE

This stage is expected to take just over a year however this is a guideline and this is the most important stage so caution is more important than in other stages.

6.0 WORK SO FAR

The idea that men are created free and equal is both true and misleading: men are created different; they lose their social freedom and their individual autonomy in seeking to become like each other.
David Riesman, The Lonely Crowd

This section provides an insight into the work that has been completed over the last year both technical and experimental. They are listed in the chronological order in which they were undertaken. The experiments detailed in sections 6.1 and 6.2 aided in the comprehension of DIVE and virtual reality in addition to experimental procedures and team skills. DIVE is dealt with under section 6.3.

6.1 INTERNET 2 PROJECT [SUB EXPERIMENT A, PART I]

The aim of the Internet2 project was to investigate the extent to which people in physically remote locations could collaborate together within a shared virtual environment in order to carry out a joint task. As part of this goal the research was also designed to examine the extent to which the DIVE* system [Swedish Institute of Computer Science 2002‡: (74)], could be used within the context provided by Internet2.

Although DIVE had been developed over several years in collaboration with partners in other projects, and used successfully before for remote collaboration, in this project there was the need to significantly extend the notion of collaboration from 'seeing' and talking only to include manipulation of shared objects as well. This had never been tried in virtual reality across a network before.

There were two parts to the proposals. First, given the time delays of the network, to what extent was it possible for people to carry out a relatively complex joint manipulation task together? In particular, they would carry an object (like a stretcher) from one place to another in virtual environment over several virtual meters. This task was carried out in an experiment, which included collaboration with UNC-CH.

* Distributed Interactive Virtual Environment

The second involved trying to repeat a version of an earlier haptics feedback experiment in which two people manipulate small objects together (blocks on a table top) with mutual haptics feedback between the two participants.

6.1.1 THE EXPERIMENT: COLLABORATIVE NAVIGATION TASK

An experiment was carried out in order to assess the extent to which users can collaborate together in carrying out a joint manipulation task. The experiment was a study designed to observe and record the results, as input to subsequent research. The results were published [(51)].

At UCL the subject was in a system similar to a CAVE™* called a ReaCTor. At UNC-CH the participant used a head-tracked head-mounted display.

6.1.1.1 Experiment Design

The task had two levels of difficulty. In real life when people carry a large object together around corners or up stairs, it is rarely a simple feat. There are the difficulties associated with weight and with getting the object around sharp bends.

A similar task in a VE has several major added difficulties. The first issue was that the virtual depiction of the people (avatars) was simplistic in comparison to reality. Previous studies [(67); (67); (67)] have shown how non-verbal feedback such as body language and facial expression are crucial to successfully carrying out virtual tasks together (even non-collaborative ones). In this experiment the avatars were block-like structures, with only two moveable parts: head and a pointer indicating the position of the person's tracked hand.

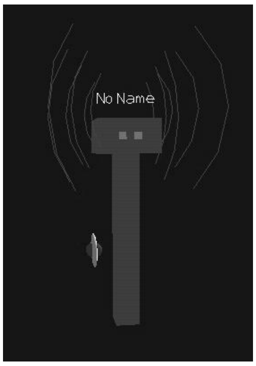


Fig. XIII The Internet2 avatar

The second predicament is that in real-life physical manipulation, there are natural constraints imposed. When two people are jointly holding an object such as a stretcher with a weight on it, they are constrained to move in conjunction to each other. The absence of real weight of the stretcher is in fact a disadvantage.

Thirdly, all the real-life feedback associated with locomotion through a physical space is missing. Within a typical VE the art of navigation becomes a skill associated with manipulating a joystick or

* CAVE is a registered trademark of the University of Illinois Board of Trustees

other device. Since the carrying of the object also involves the use of such a device, the task is doubly complex. It requires training.

Overall the task of manipulating an object together in a virtual environment, where the object is large (i.e. body size) is actually far more difficult than carrying out the same task in real life. This was therefore an exceedingly difficult test of the extent to which collaboration between people in remote locations was possible. Each subject has to be trained in a virtual world on how to navigate the environment and how to pick objects.

The phrase "*Patience is a virtue*" takes on a whole new meaning during the training of subjects within a short space of time.

6.1.1.2 Scenario

On arrival for the experiment the UCL subjects were asked to sign a consent form, which informed them of possible negative effects from using the system such as simulator sickness. They were told that they could withdraw from the experiment at any time without giving a reason, and that they agreed that they would not be driving or operating complex machinery for at least 3 hours after the conclusion of the experiment. This is a standard procedure used in almost all the VR studies undertaken at UCL. Their permission to be audiotaped and video taped was also obtained in the consent form.

They were then given a sheet outlining the task. This sheet informed them that the task was to meet with the other person at UNC, and negotiate lifting a stretcher together. They were to take the stretcher together along a blue path that led into the building, and then put it down on a red coloured area inside the building.

There were two experimenters at UCL. One took the subject through a training task. This was to make sure that each of the subjects was familiar with how to navigate the VE, how to pick things up and hold them, and then carry them as they navigate. This was done in a different VE to the main experimental task.

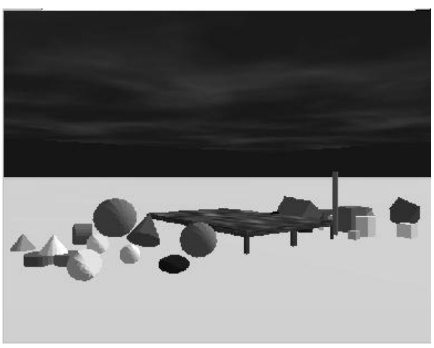


Fig. XIV The training world

When the training was completed, the second experimenter started up the link to UNC, and then explained the task again. The person at UNC was not actually an experimental subject, but rather a confederate. At UNC the experiment was conducted by three people, one of who virtually met each of the 17 subjects recruited at UCL. Altogether there were 19 people for whom results are available (the UNC participants although repeating the experiment only answered the questionnaire the first time). There were 5 women and 14 men, ranging in age from 19 through 34.

They were informed them that the task was to meet with the other person at UNC, and negotiate lifting a stretcher together. They were to take the stretcher together along a blue path that led into the building, and then put it down on a red coloured area inside the building. The subjects were told to negotiate with the other person. It was emphasised that the blue path was only a guide and that more important than following it, was to follow the direction indicated. Once all this was explained, the UNC-CH avatar entered the VE and the experiment started.

The collaboration was stopped after 5 to 8 minutes, depending on the stage to which the process had reached. Although the avatars were very basic, simple movements can be viewed: e.g. when the UNC person nodded his head, bent down or held his head low, the UCL subject would see this in real-time and vice versa. In spite of such a basic avatar, to what extent could “mood” be conveyed across such a narrow bandwidth?

The UNC confederate was instructed to behave as either a “very happy” or “very depressed” collaborator. As part of the data gathered the UCL subject was asked the extent to which they recognised the mood of the other person. Of course such mood guessing would be based on voice, but also on the disposition of the avatar body: drooping head indicating depression.

6.1.1.3 Results

The data was analysed by Prof. Mel Slater. The UCL subjects were able to assess the mood state of the UNC person (recall that the UNC person was acting as happy or as depressed). The mean mood score (i.e., estimated degree of happiness) amongst those who experienced the depressed acting was: 2.5 ± 0.8, and for those who experienced the happy state: 3.9 ± 1.4.

The difference is significant at the 5% level. The relatively high variance amongst the “happy” responses probably reflects that it seemed to be much harder for the UNC subjects to convincingly act in a “happy” manner than in a depressed manner.

Some of the comments written in the online questionnaire administered to the subjects in order to determine the amount of co-presence and believability they felt during the period of the collaboration are posted below:

- I wasn't sure whether 'laws of physics' applied, i.e.: whether walking 'through' the stretcher would prevent the other person from completing the task.
- I wasn't sure whether the other person was always listening. I was expecting to be told what to do by the other person, but he was not forthcoming.
- There was a slight transmission delay that dulled the realistic sense of the experience and the "stretcher" which did not particularly look all that realistic, which made it hard to know what to manipulate in order to get the task done.
- The person I was working with didn't realise if I was in difficulty - when I asked to put down the object - in order to reassess the situation, an "I'm okay" answer was received indicating a communication problem.
- In the beginning it felt more like a that I was in a virtual world but once we started the task and communication between us harmonized then it felt as we were doing the task in a real world. I think in a situation like this one the only thing that makes u realize that you are not in a real world is the lack of communication

This evidence is highly promising considering the inadequate software support, the lag in networks across the Atlantic, bottlenecks between the different network speeds, the simple avatars, and the scheduling problems. It seems to suggest emotional states of minds can be passed with less than perfect expressions of it. More quantitative results as reported in the paper [(51)] are appended under section 8.

6.1.2 PROBLEMS ENCOUNTERED

In the first experiment the only problem that was encountered in terms of performance of the system was that navigation messages were lost across the networks. So when the avatar in one incarnation of the world stopped, it would send a multicast of ‘stop’ events to other incarnations of the world to other users of the same world. If the ‘*stop*’ message were lost, the representations of the avatar that just stopped would continue moving in copies of the world hence giving an impression of the avatars shooting away from each other into infinity. The messaging problem was solved by resending lost messages across the networks to corresponding incarnations of the DIVE world.

The haptics scenario was not carried out within the time limit and this was due to a coupling between the rendering loop and the network-messaging loop. DIVE renders at a speed of 60 Hertz (Hz) whilst the haptics device sent messages at the much faster rate of 1000 Hz. This caused a jittering effect in the world. This problem has now been solved by a colleague and is to be published at a future date [(39)].

6.2 SOCIAL PARANOIA EXPERIMENTS

> *It could be that we have already noticed so many times that body language gives more of a hold-on than words, which makes us automatically doubt the words when they do not correspond to the non-verbal signals.*
> *F. R. Oomkes*

Can Virtual Reality induce paranoid-like ideas in people? How do people respond to virtual people in a virtual setting? In the real world, people with persecutory ideas have a pervasive tendency to distort experience by misconstruing the neutral or friendly actions of others as hostile or contemptuous. Will this also happen when faced with simulated people, despite knowing that the avatars are computer generated?

In order to study these questions a series of pilot experiments were conducted in September 2001 to aid the completion of the masters' thesis of Alican Met [(49)] and to aid the author of this report in the comprehension of the possible thought-impregnations that could happen in users of virtual environment when avatars are programmed with behaviour. The purpose of the experiments were to asses the extend to which persecutory ideas was induced by the virtual reality environment and the virtual people.

6.2.1 THE EXPERIMENT

For this research a library room was created using VRML*. The five avatars that were used in this project were also used for the fear of public speaking study [(65)] and they were seated around one of the tables. Several books, papers and sheets are distributed among the tables and the walls are covered with bookshelves to makes the room look more realistic and not too clean and empty. This in itself is a feat. Trying to make a virtual environment look cluttered is not easy! The avatar animations and the experiment scenario were pre-designed by Alican in collaboration with Dr. Daniel Freeman from the Department of Psychology at the Institute of Psychology. Professor Mel Slater undertook the analysis of the data collected.

6.2.1.1 Scenario

The scenario can have three different variables. The avatars can be static, or they can be animated but without sound, or they can be animated with sound. The measurement of the persecutory ideation is done by questionnaires and interview. The subjects were given a set of questionnaires where some of them were screening questionnaires to prevent individuals with a high level of paranoid thoughts or delusional beliefs to participate in the experiment. More information on the procedure and the experiment can be obtained from the thesis [(49)].

6.2.1.2 Results

This project confirmed that VR could induce persecutory ideas in people that have persecutory ideas in real life. The "realness" of the environment has made it possible to trigger the same set of thoughts as a real experience triggers in the individuals. The high mean score of presence is a clear indication that the virtual environment was convincing to the subjects. Although the number of subjects was too small, it was predicted that a high degree of presence would increase the level of paranoid thoughts.

The analysis of the questionnaires showed a clear indication that people with tendencies towards paranoid thought had triggered the same set of thoughts in the virtual environment as a real experience. For both males and females the more negative their prior self-evaluation the higher the 'paranoid' score after the virtual environment experience. Similarly the higher their presence, the greater the paranoid score. However, for males there is a strong linear relationship between their prior paranoid screening score and the post-experience score. For females this relationship although maybe in the same direction is not significant. Once again females scored higher on the paranoia scale. Some of the comments made include:

- "The two people to the left, I didn't like them very much – well, I don't know, maybe because when I entered the room I felt being watched and then they started talking about me. The other people were more neutral and more inviting except the guy with the beard."
- "It was probably more real to me than I expected it to be. At some point, I was trying to navigate around a table and almost found myself saying sorry to the person sitting there. I felt that they were getting annoyed with me for doing that…"
- "I was conscious that some of the people were staring at me at some point and there certainly was made a couple of comments, eh don't' know if it was ordinary conversation or about me… When I walked around in the room I heard a couple of comments and a "oh my god" but I don't know if I am being paranoid but they were looking at me and I was conscious of that…"
- "When I was looking at the bookshelves, I did wonder if they were looking at me and when I turned back on one or two occasions they were and on one or two occasions they weren't."
- "I was really weird, because they were all definitely in on something and they were all trying to make me nervous. It was clear that they were trying to mock me, they kept on looking at me and when I looked back, and they were … The guy with the suit was really weird because he kept smiling at me and it was quite sinister."

* Virtual Reality Modelling Language

- "There were three people on the right – one with a suit, I think he was a business man working on his laptop…"

With reference to the last comment above, there was no laptop. The avatars were not programmed to portray any particular type of personalities either. Since there was no clear tendency towards the results expected it was decided to re-run the experiment with a larger group of subject.

A larger study with 24 participants (gender balanced) was conducted with the help of Joel Jordan and one of Daniel Freeman's employees, Cristina Reed. The results found that positive views about the virtual characters were common. However a number of participants had ideas of reference and ideas of persecution about the virtual characters.

Individuals who had persecutory thoughts about the virtual characters tended to have higher levels of trait paranoia, and had significantly higher levels of inter-personal sensitivity and anxiety. Some of the comments provided in the second study include:

- 'Friendly people just being friendly and offering a smile'
- 'People were nicer than real people!'
- 'Part of a game (flirting but being shy)'
- 'It was nice when they smiled, made me feel welcome.'
- 'They looked friendly – that was my overall impression'
- 'They were very ignorant and unfriendly'
- 'Sometimes appeared hostile, sometimes rude'
- 'It was their space: you're the stranger.'
- 'They were telling me to go away'
- 'One person was very shy and another had hated me'
- 'The two women looked more threatening'
- 'Some were intimidating'

6.2.1.3 Conclusion

The study provides evidence that individuals can attribute mental states to virtual reality characters. Importantly for the study of clinical phenomena, individuals can have thoughts of a persecutory nature about virtual characters. Further, the findings indicate that feelings of vulnerability may directly contribute to the development of persecutory ideation. Virtual reality may prove to be a valuable methodology for developing the understanding of persecutory ideation. According to the Institute of Psychology the results were consistent with other studies.

6.3 BENDER

As stated in the abstract the overall concept of this thesis is that an increase in behaviour realism demands an increase in behaviour realism in order to produce convincing levels of co-presence, believability and realism. Generally it is believed that users of virtual environments expect a behaviour model reminiscent of human behaviour. They are less forgiving to visual inconsistencies when the virtual representation of the user is closer to a humanoid avatar. It might be the case that if the avatar is physically less sophisticated an exaggerated behaviour model might be necessary to compensate for the lack of photo-realism and produce satisfactory levels of realism.

It was decided to build a simplistic humanoid form from scratch in order to have a simple humanoid representation for subjects in future experiments, which are designed to compare realistic and simplistic avatars. This also had the added bonus of providing a way of learning about the building tool, DIVE, and study about a bone hierarchy standard, *H-Anim*. Both are detailed in subsequent sub-sections.

6.3.1 DISTRIBUTED INTERACTIVE VIRTUAL ENVIRONMENT

The Distributed Interactive Virtual Environment (DIVE) [(74)] is an internet-based multi-user virtual reality system where participants navigate in 3D space and see, meet and interact with other users and applications. It supports the development of virtual environments and applications based on shared 3D environments. The architecture focuses on software and networking solutions that enable high interaction at each participating site, which was why it was chosen as a platform for the *Internet2* project.

The software was implemented on a derivative of DIVE 3.3x [Frécon *et. al.* 2001: (32)]. This was recently ported to support spatially immersive systems [Steed *et. al.* 2001: (71)]. Participants in a DIVE world can be either a human user or an automated application process i.e. an actor. Furthermore, it reads the user's input devices and maps the physical actions taken by the user to logical actions in the DIVE system in most case the head and the right hand. It includes navigation in 3D space, clicking on objects and grabbing objects.

Furthermore, C/C++ plugins make DIVE a modular product by providing a set of symbols and functions recognised by DIVE. When the plugin is loaded into the DIVE process these symbols are looked up by the DIVE plugin interface and called. Such a plugin was constructed in C to animate the avatars’ body parts and is discussed in detail in upcoming sections. DIVE also supports import and export of VRML and several other 3D file formats which means ready made avatars can be imported and used.

It is possible to write Tcl scripts to run processes and animations in DIVE and this methodology was used in the past while animating agents. However it was felt that in run-time animation the use of a C-plugin would be faster. The avatars were moved to their respective positions in their corresponding training rooms at the start of the experiment with the aid of Tcl script. At the end of the training session the doors leading to the meeting room in which the conversation took place were also opened with the use of a Tcl script. The procedure undertaken in the eye-gaze experiments required the participants are kept in a separate virtual room until the start of the experiment. More on this procedure is detailed under section 6.4.

Since DIVE also supports import and export of VRML and several other 3D file formats, it was possible to import ready-made avatars from other projects [(65)], make them H-Anim [Web3D Consortium 2001□: (84)] compliant and use them.

6.3.2 HIERARCHY OF JOINTS: H-ANIM

As the use of Virtual Reality and 3D Internet continues to grow, there is an increasing need to represent human beings in online virtual environments. Achieving the goal will require the creation of libraries of interchangeable humanoids, as well as authoring tools that make it easy to create new humanoids and animate them in various ways. H-Anim humanoids can be animated using key-framing, inverse kinematics, performance animation systems and other techniques [(84)]. The design goals of the creators of H-Anim match those of this thesis in terms of avatar design.

- **Complexity:** They should work in any VRML compliant browser
- **Flexibility:** No assumptions should be made about the types of applications that might use the humanoids
- **Simplicity:** The specification can always be extended so any doubts for a specification exclude it out of the standard

Bender was created keeping in mind this standard. This partially solves the problem of connecting children joints and segments a top node when it moves. Examples of joints are elbow, wrist ad ankle. Segments are the bones in between like forearm, hand and foot. More realistic avatars were made to conform to the standard and the methodology used is discussed in section 6.6.

The diagram below shows the hierarchy employed in the use of H-Anim. The tact used here is to mirror real life connection of the various human bones. The *root* node of the whole bone structure is the pelvic node. The pelvic node has three *children* nodes namely the lower torso, the left leg and right leg. The lower torso, the left leg and the right leg are *sibling* nodes. So if the pelvic moves, the whole body follows. The legs and lower torso in turn have children nodes and so on. Any translations or rotations travels down the spine but never across or upwards. The figure below depicts the whole hierarchy.

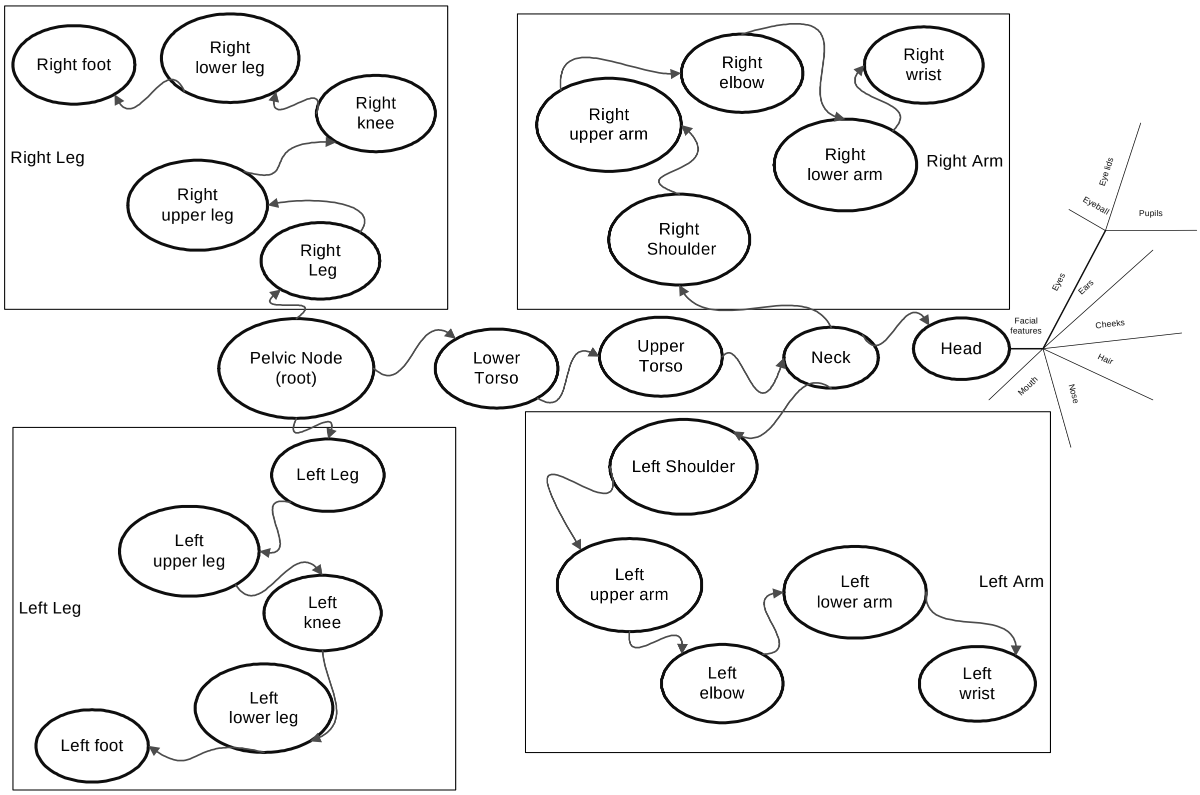


Fig. XV H-Anim hierarchy

The full figure is available from the H-Anim website [(84)] and contains the complete hierarchical structure defined in their standard.

6.3.3 BENDER'S APPEARANCE

Bender went through three appearance modification stages. The avatar had to look simplistic but humanoid. The proportions of Bender's body were made in accordance to the mathematical impression an artist uses in drawing the human figure. The proportions of the whole human body parts are a multiple of the human head. In theory, the average length of the human body is seven times that of the average length of the human head. However since artists do not think it looks quite right it has been regulated that the "*correct*" proportional length of the body should be eight times that of the head.

Lower torso	=	*1.0*	*	*(Head Length)*
Upper torso	=	*1.0*	*	*(Head Length)*
Upper arms	=	*1.5*	*	*(Head Length)*
Lower arms	=	*1.5*	*	*(Head Length)*
Upper legs	=	*2.0*	*	*(Head Length)*
Lower legs	=	*2.0*	*	*(Head Length)*

6.3.4 BENDER'S PLUGIN: C

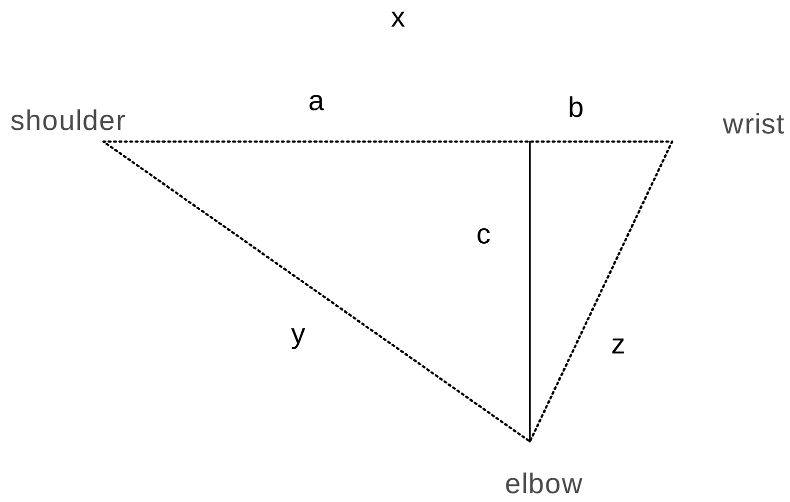
The idea behind this C-plugin was to create an animation methodology to be used in conjunction with any avatar independent of any one platform. At the moment the plugin has been designed to work with DIVE and recognises avatars that are H-Anim compliant. The following sections give details of how the avatar plugin has been implemented.

The implementation of the plugin has been slightly ad-hoc but this was mainly due to lack of time and preliminary planning. This initial experiment has not only provided valuable insights into what makes an avatar realistic but also pin pointed problems in the implementation of the plugin which are detailed a later sub section.

DIVE automatically updates the position of the right-hand and the head of the avatar. The C plugin was used to animate the rest of the avatar's body in order to maintain a visually consistent humanoid. This included inferring the position of the right elbow using inverse kinematics when the user's tracked hand moved, and deducing the position of the avatar's knees when the user bent their knees. There were also some deductions involved in the rotation of the head and body. The body was not rotated to the same direction as the head unless there was some translation associated with the user. This was to enable the user to nod, tilt and shake their head in the VE whilst in conversation.

6.3.4.1 Inverse Kinematics of the arms: Geometrical

In the case of the avatars, the only tracked body parts are the head and the right hand. This section deals with the deduction of the elbow position of the tracked arm. The shoulder moves with the body, which moves directly under the tracked head. The hand is tracked via a joystick. The elbow is deduced using the "*co*s" rule.



x, y, and z are known

$x = a + b$

$a^2 + c^2 = y^2$

$b^2 + c^2 = z^2$

$a^2 - b^2 = y^2 - z^2$

$a^2 - (x-a)^2 = y^2 - z^2$

$a = (y^2 - z^2 + x^2)/2*x$

$b = x - a$

Fig. XVI Figure depicting human arm

A point along the line between shoulder and the hand is pin pointed using the rule and then it is pushed perpendicular to the plane, which contains both end points. The elbow co-ordinates are used to find a vector direction to get the direction in which the upper arm and the lower arm should point.

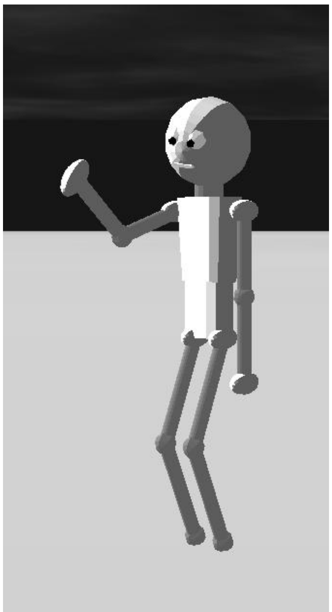
The direction is got in a straightforward fashion by simple vector subtraction but this only gives the Y-axis of the new rotation matrix of the stated body parts. However the direction has to be converted to a three-dimensional rotation matrix. One perpendicular vector is obtained by finding the cross product between the Y direction and the vector (0, 1, 0). By finding the cross product between the direction and the resultant vector of the previous cross product, a second perpendicular direction is found. This procedure results in three perpendicular vectors. However, if the Y-direction happens to be in the same direction as the vector (0, 1, 0), the cross product fails.

In order to deal with rotation matrices correctly, it has been decided to look into Quaternion[*] matrices. This is more thoroughly dealt with in section 6.5.2.

6.3.4.2 Deductions (Legs and Body)

Another deduction that is made frame by frame is the position of the knees in case of the body bending. If the head goes to a position lower than that of the total body length across the Y axis, the program assumes that the user is bending and moves the position of both knees by the appropriate distance in front of the body.

[*] Any number of the form $a + bi + cj + dk$ where $a, b, c,$ and d are real numbers, $ij = k$, $i^2 = j^2 = -1$, and $ij = -ji$. Under addition and multiplication, quaternions have all the properties of a field, except multiplication is not commutative



The face is moved in accordance to the translation and rotations of the head tracker as long as the length of the body is not exceeded. If the body length is higher that it should be, the face follows the rotations of the head tracker but not the translations.

Fig. XVII Arm Kinematics and Legs in action

The body moves right under the head of the user. However the rotations of the head is not completely mirrored by the body. The body is only rotated in along the Y-axis and then also only if the translation in the head tracker is more than about 0.075 meters. As already stated, this is to ensure the ability of the user to nod and shake their head without the body rotating as well.

6.4 EYE-GAZE EXPERIMENTS [SUB EXPERIMENT A, PART II]

An experiment was designed to explore the impact of eye gaze in a dyadic conversation while humanoid avatars represent the users. This is an immersive version of the previous experiment carried out my Garau et. al. [(33)] at BT[*]. The experiment had two purposes:

- To ensure that the positive results got from the previous experiments were due to eye-movement and not head-tracking.
- To study the impact of differing avatar representations.

As already mentioned in section 5.0, there were four mediated conditions in which the users are represented to their conversational partners as a different type of avatar (four in total):

- A realistic avatar (Jack/Nancy) with random eye animations
- A simplistic avatar (Bender) with random eye animations
- A realistic avatar (Jack/Nancy) with inferred eye animations
- A simplistic avatar (Bender) with inferred eye animations

[*] British Telecom

In the random-gaze condition the avatar's eye animations were unrelated to conversational flow. In the inferred-gaze condition, they were related to who were speaking and who weren't in the conversation. The animations were inferred from the audio stream.

The avatar plugin needed some specific changes made to make it suitable for this experiment. One of the main add-ons was the behaviour of the eyes. The following section deals with the modelling of the eye animations.

6.4.1 THE EYE ANIMATIONS

The eye gaze of human beings (and animals alike) play a great role in social communication. Eye-gaze signals turn taking in conversations and indicate how interested (or not) another person finds the conversation. It carries across emotions between people. A lot of information transpires both consciously and sub-consciously between people in conversations or simply between other types of interaction. This study focuses between the amounts of time a person *looks-at* and *looks-away* from another person in a dyadic conversation. It also takes into account of the magnitude of the *looking-away* eye-position and the velocity at which it gets there.

Before studying the models used to program the eye movement, it is necessary to know some definitions/conventions used in describing the models. The change in eye-direction is called a *saccade*. The amount of change in degrees between the eye-direction is called the *saccadic magnitude/amplitude*. The change in direction over the time it takes is called the *saccadic velocity*. The direction of the saccade is called the *saccadic direction*. When the eye returns to *look at* the eye of the conversational partner, it is said to be in *mutual glance* position. The time between the termination of a saccade and the beginning of the next one is called the *inter-saccadic interval*.

Two models were considered and programmed in the study. The original model, which was to be used in the experiment, is referred to as the *Original* model. The second model, Badler, was published while the preparations for the experiment were being carried out. On studying the Badler model, it was agreed that the eye-animations were slightly more naturalistic. So the saccade-directions were calculated from the Badler model. The saccade velocity was calculated using a methodology to get a result consistent with the Badler model. However the original timing model was used to determine how long the *look-at* and *look-away* glances were held.

In both implementations, a decision was made not to automate *at* gaze by targeting it at the other avatar. Rather, *at partner* gaze was targeted straight ahead in relation to the position and orientation of the head. In this way, the avatar could only seem as if it was looking *at partner* if the participant was in fact looking directly at the other avatar's face.

6.4.1.1 Original Model

This model assumes that the eye-saccades never exceed a maximum magnitude of 15 degrees [Bahill *et. al.* 1975: (7)]. That is the eyeball rotation is done within a 15-degree cone. It also states the mean time during the looking-at and looking-away glances. The timing model itself is an exponential distribution of random numbers around the mean times provided in this model. This is to avoid any repetitive animations of the eye.

Generating uniform random numbers and then distributing them exponentially using the formula below ensured a unique instantiation of the animation.

*Exponent = -(mean) * loge (uniform random number in the range [0-15]);*

It was concluded that the average time a person looks at their conversational partner while listening is more than the time a person holds mutual glance while speaking consistent with Argyle's findings [Argyle *et. al.* 1976: (4)]. Given below are the timing averages i.e. *inter-saccadic intervals*.

While speaking, average time period of mutual glance (looking-at)	*1800ms*
While speaking, average time period of away glance (looking-away)	*2100ms*
While listening, average time period of mutual glance (looking-at)	*2500ms*
While listening, average time period of away glance (looking-away)	*1600ms*

The saccade direction and magnitude are not done according to any particular formula so this was calculated randomly as long as it stayed within the 15-degree limit. This meant that the direction could be anything in the 2D circle controlled by a 15-degree cone with the central eye position being its axis of return.

6.4.1.2 Badler Model

The statistics for this model were taken from a similar study published by the University of Pennsylvania [(41)] [Becker 1989i: (15)]. The model states that the maximum eye-saccade is 27.5 degrees while speaking and 22.7 degrees while listening. The magnitude of each saccade is obtained using the following formula:

*Magnitude = -6.9 * log_e (uniform random num in the range [0-15] / 15.7)*

The uniformly generated number depicts the percentage of time that particular magnitude occurs and follows the distribution shown in the figure below.

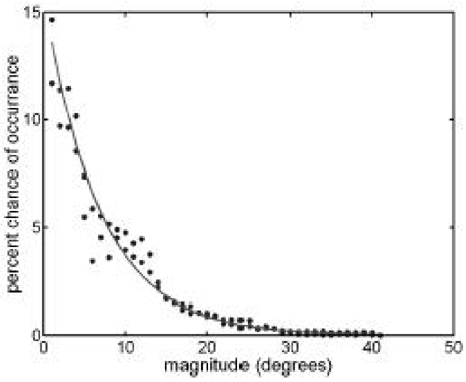


Fig. XVIII Saccade Magnitude distribution

Unlike the original mode, which had a lot of possible saccade directions, the Badler model has only eight basic directions, each spaced at 45 degrees around a circle.

A variety of magnitudes are applied along those directions to get the necessary saccade jumps. The model also gives a probability chart for the direction of the eye-saccade. According to the statistic derived in the paper [(41)], up-down and left-right movements happens at twice the frequency diagonal movements happen. Also up-down saccades happen as often as left-right movements. In the table below 0 degrees is to the right and the angle increases in anti-clockwise direction.

Direction (deg)	*0*	*45*	*90*	*135*	*180*	*225*	*270*	*315*
Percent (%)	*15.54*	*6.46*	*17.69*	*7.44*	*16.80*	*7.89*	*20.38*	*7.79*

The theory of *saccade velocity* was the most interesting part of the Badler model. It sprouts from work published in 1989 [(15)]. The model suggests that during a saccade the eye movement begins with an extremely high initial acceleration (30,00 deg/sec^2) and terminates with a similarly high deceleration. Peak velocities during a saccade can be between 400 and 600 degrees/second. The duration and the velocity of the saccade are functions of its magnitude. In this case the function is:

Duration = $D_0 + d * Magnitude$

'D_0' is the intercept or catch up time (25 milliseconds)

'd' represents the increment in duration per degree (2.4 milliseconds/degree)

In reality 'D_0' can vary between 20 to 30 milliseconds and 'd' can vary between 2 and 2.7 milliseconds/degree for a normal saccade jump of 5 to 50 degrees. Within this range of saccadic jump, the rate of increase of magnitude is constant hence gives the linear function above.

Once the duration of the saccade is determined, the velocity is determined using a fitted instantaneous velocity curve. The duration of the saccade is divided into six frames and an appropriate velocity is calculated using the model. The eye is then rotated along the appropriate intermediate magnitude in the saccadic direction needed frame-by-frame with the appropriate pauses to reflect the saccadic velocity. In comparison to the inter-saccadic intervals, these saccadic durations are very minute.

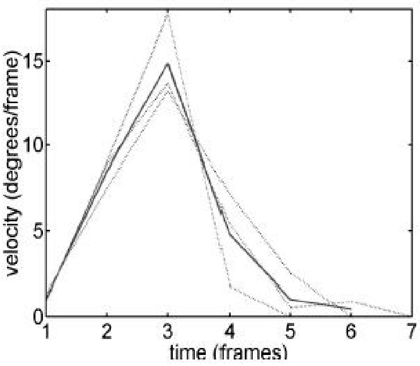


Fig. XIX Badler's Curve for instantaneous saccade velocities

Even though a polynomial is given in the paper, it was felt that it did not correspond with the instantaneous velocity curve, so the curve was approximated further to a Gaussian function.

$Y = 14 * e^{(-(PI/4) * (x-3) * (x-3))}$

Y: Instantaneous saccade velocity at a frame X

X: Frame number ranging from 1 to 6

Saccadic velocities with accordance to frame 1 to 6:

{0.604995, 6.383134, 14.0, 6.383134, 0.604995, 0.011920}

The *inter-saccadic intervals* were not changed since the paper states the timing as being consistent with the Argyle timing used in the original model. That is the times for mutual glance and away glance while speaking and listening were left as they were in the original model.

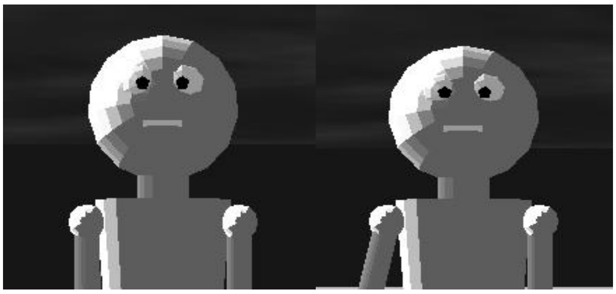


Fig. XX Eye-animations on Bender (At and Away)

The Badler model is consistent with timing expectations from the literature, but adds valuable new probabilities for gaze direction during *away* fixations that were absent in our initial model. Both the original model and the more recent one were implemented and compared. The more recent model was selected for the eye-gaze study as it yielded more satisfying results in the immersive setting.

6.4.1.3 Random Model

The eyes were rotated once every two seconds at a random direction with the 15-degree cone determined as the natural cone of view for a person in dyadic conversation.

6.4.2 ADAPTING EXISTING AVATARS: JACK & NANCY

One basic problem that was faced was that avatars come from different sources and have different names for the similar body parts. This happens despite the existence of common interchange formats for avatars such as H-Anim. This meant that the avatars (Jack and Nancy) had to be *broken* into the appropriate number of body parts and renamed into the names that are recognisable by the plugin. The body parts were then stripped of their rotation matrices. This doesn't mean the simple deletion of the rotation matrix. The body part had to retain the rotation as part of the polygon's vertex definition. So the body part had to be placed in the H-Anim standard rotation.

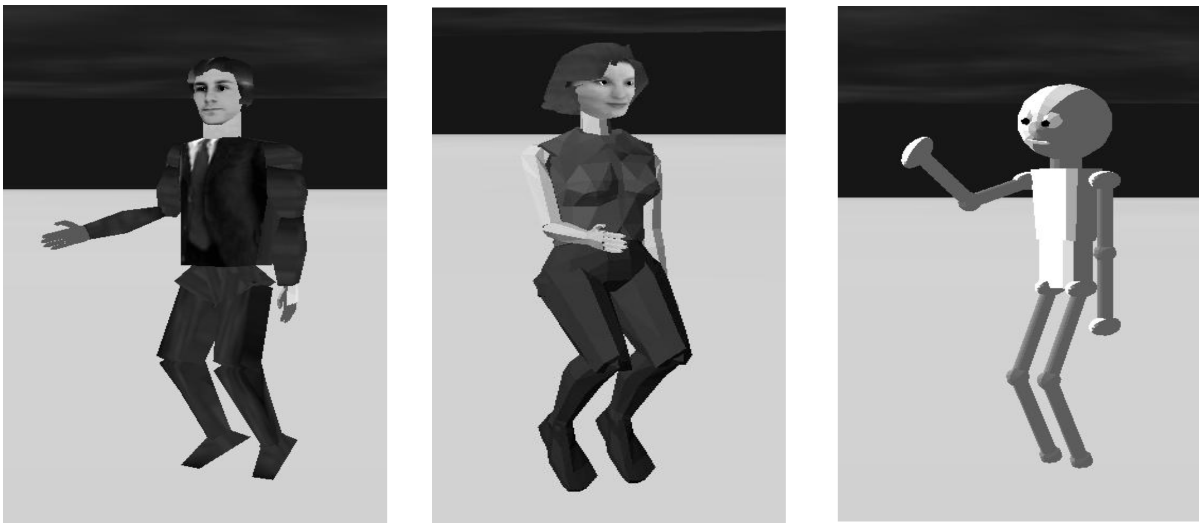


Fig. XXI The Avatars used

6.4.3 THE HARDWARE BEHIND THE SCENE

The idea was that two subjects would be placed in a shared virtual environment to converse on a set theme. One of the subjects was placed in the ReaCTor and the other was placed in a Head Mounted Display (HMD). A machine called *Exodus* runs the ReaCTor and a machine called *Genesis* runs the HMD. A Linux box called *Alpha* runs the visualisation for the HMD. Since *Genesis* does not have a sound card, the audio pipeline for the HMD had to be sent through another machine called *Romulus*. Both participants had wireless microphones attached to their clothing. These were only activated just before the start of the conversation so that participants could not hear each other during the training period. Synchronising all four machines with all the correct settings for every experiment was a logistical nightmare in itself.

6.4.3.1 The ReaCTor

The CAVE'-like system used was a ReaCTor made by Trimension, consisting of three 3m x 2.2m walls and a 3m x 3m solid acrylic floor. It is powered by a Silicon Graphics Onyx2 with 8 300MHz R12000 MIPS processors, 8GB RAM and 4 Infinite Reality2 graphics pipes. The participants wore CrystalEyes stereo glasses, which are tracked by an Intersense IS900 system. They held a tracked 4-button joystick in their right hand. The frame rate was 45 Hz throughout.

Fig. XXII The ReaCtor

6.4.3.2 The Head-mounted Display (HMD)

The scenarios were implemented on a Silicon Graphics Onyx with twin 196 MHz R10000, Infinite Reality Graphics and 192M main memory. The tracking system has two Polhemus Fastraks, one for the HMD and another for a 5 button 3D mouse. The helmet was a Virtual Research V8, which has true VGA resolution with 640x480x3 colour elements for each eye. The V8 has a field of view of 60 degrees diagonal at 100% overlap.

Fig. XXIII The Head Mounted Display

6.4.4 AUDIO DETECTION

Since this was a collaborative task a lot of unseen potholes were uncovered as the experiment dates approached. One such issue was that of audio detection. As can be gathered from earlier sections of this report, the behaviour of the eye varies in a dyadic conversation and one of the control variables is the role being played by the user in accordance to speaking/listening.

The user will tend to *look at* the conversational partner more when listening than when speaking. In order for the correct eye animations to be triggered in the avatars it is inevitable that the program detect when the user is speaking.

The audio panel maintained audio streams input into *Romulus* and *Exodus*. Existing software, written by David Swapp, was recalibrated to recognise when there was a speech signal coming through from the microphone. This was done with the aid of a threshold. Any signal below the threshold was discarded as noise and anything above it was speech. A wrapper script was written to reset an event called the *message* event, which is recognisable by DIVE.

The plugin was then re-routed to check for an event named *chattering* to detect if it indicated a speech stream. Depending on the presence or absence of a speech stream the right animations were triggered. The only changes in the eye behaviour are the *inter-saccadic intervals.*

6.4.5 SPATIAL RECORDING

The amount of space a person maintains in between their self and the conversational partner could potentially provide an important insight into the user sub-conscious regard towards the avatar. Since this was easily retrievable and cheaply available information, the plugin printed out the spatial information of the avatar every 1.5 milliseconds.

'Time ms:μs' || *'Head rotation'* || *'Head position' 'Is user Speaking'*

6.4.6 THE EXPERIMENT

There were two participants in each experiment. One participant was in a system similar to the CAVE called the ReaCTor and the other participant was in a HMD. The subject in the CAVE held a navigation device with 4 buttons and an analogue joystick that was tracked. All buttons except for the joystick were disabled to stop participants from manipulating objects in the virtual room. The subject in the HMD was given a similar 5-buttoned joystick. All the buttons except two are disabled. The joysticks were used to move around the VE, with pointing direction determining the direction of movement enabled for the horizontal plane only.

There were 24 validated pairs of subjects in total, 12 pairs of female subjects and 12 pairs of males. Even though 56 subjects were run through the experiment, only 48 subjects were able to participate in the experiment without hitches. Also some subjects had to be discarded because they knew too much about the research.

6.4.6.1 Pre-Experimental procedure

There were three experimenters of which during the experiments, there would be two at any one time, one for each subject. Each pair of subjects were separated and kept a part before the experiment so that they did not form an opinion of each other before they went into the VR. On arrival the subjects were given an instruction sheet briefly informing them about an outline of what the procedure would entail. They were then asked to sign the standard consent form similar to the one in the Internet-2 project. Their permission to be audio taped and video taped were also obtained.

The subjects were then asked to complete a personal background questionnaire, which asked them for their age, gender, occupation, familiarity with programming, computers, virtual reality, games and 3D images. After that they were asked to complete a psychological questionnaire (SAD: Social Avoidance and Distress), which was designed to gather their general outlook towards social situations. It was designed to assess their social anxiety and is appended in section 8.

The questionnaires were online following in the steps of the Internet-2 project. Since some of the subjects were not familiar with the use of computers or not well versed in the English language, the experimenter stayed with the subject throughout the procedure.

Role-Playing Activity

As detailed in section 5.0, there were two roles in the task, a mayor and a baker. Depending on which role the subject played and their gender, they were given a role-play card to read. The scenario informed them of the role-playing activity. The roles played by the subjects were equally distributed i.e. there were equal instances of the subject in the HMD playing the role of the mayor or the baker as there were in the ReaCTor. The figure below shows how the people are related.

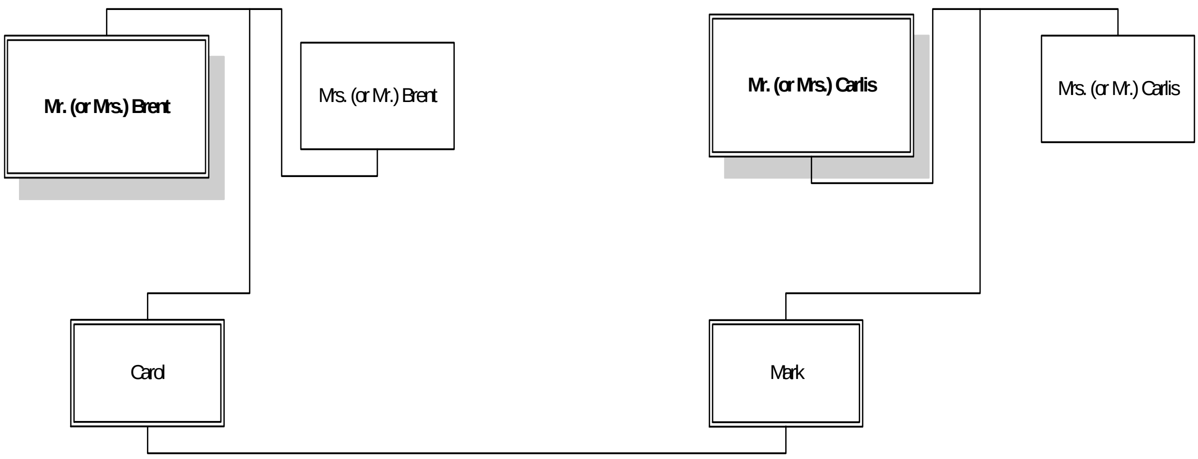


Fig. XXIV Relationship Tree

The mayor, Mr./Mrs. Carlis, was told he/she has a son, Mark, and that the local baker, Mr./Mrs. Brent, had called to talk about their pregnant daughter, Carol, and Mark. The baker also mentions a letter to the press written by the baker's spouse. The mayor calls up Mark who promptly says that he is aware of the pregnancy but that he does not think that they were in an exclusive relationship.

Mark is a top lawyer and the mayor hopes that Mark will become a judge. A scandalous issue like this one could ruin Mark's future.

On the other hand, the subject playing the baker, Mr./Mrs. Brent, is told that he/she has a daughter, Carol, who is an aspiring actress has just informed the baker of her pregnancy. Carol says that she is in love with Mark who is responsible and intends on keeping the baby. However the baker is not financially secure and wants to ensure the baby and Carol will be okay. Apart from that pregnancy, the baker's better half as written a letter to the press (with no financial gain) and is threatening to send it. The letter would ruin Carol's reputation and serves no foreseeable gain. It will surely alienate the mayor and Mark.

The mayor and the baker are to meet in the VR, do a small training session and then converse about the situation. They are to come up with solutions, which entail some sort of agreement. Some alternatives given to the subjects were: marriage, financial pay off or child support. The time given for the resolution of the problem is ten minutes.

6.4.6.2 During the Experiment

After the subjects were both ready they were taken into the training world separately in either the HMD or the ReaCTor. The subjects were trained on how to navigate in their respective immersive systems and then positioned in front or a door leading to the meeting room.

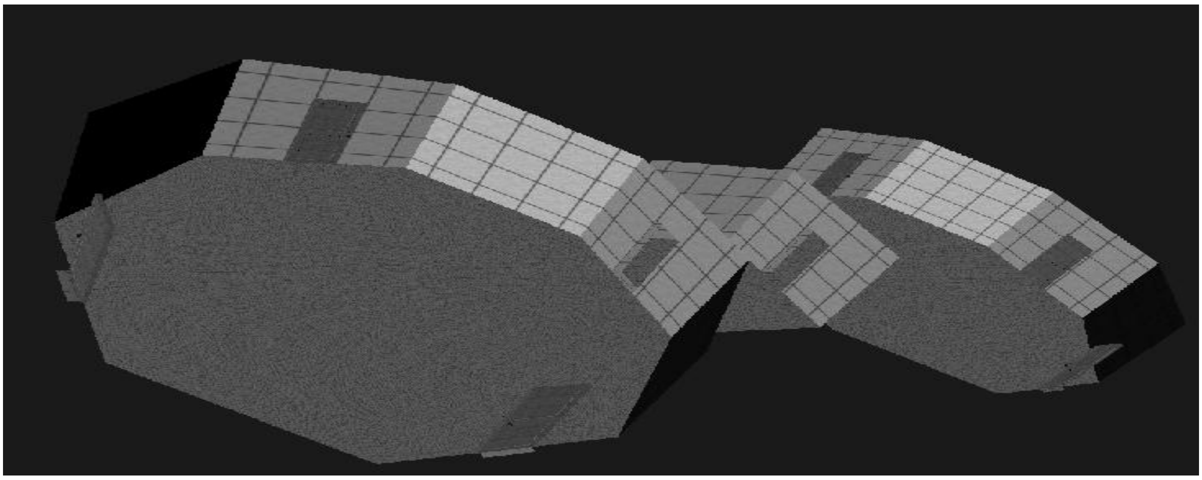


Fig. XXV The Virtual World

The doors were both opened at the same time in order to ensure both participants see each other at the same time. This was done with the aid of an embedded Tcl script. The participants were left to chat about the issue while the experimenters took notes on how much their subject gesticulates, nods, shakes their head, talks etc.

A visually identical avatar represented both participants in each instantiation of the experiment to avoid differences in facial geometry affecting the impact of the animations. As stated, there were three avatars used in the experiment. In the lower-realism condition a single, genderless avatar was used to represent both males and females. For the higher-realism avatar, a separate male and female avatar were used. Both participants were gender matched in each instantiation of the experiment. All three avatars had identical functionality.

Fig. XXVI The avatars used

DIVE automatically updates the position of the right-hand and the head of the avatar. The C plugin was used to animate the rest of the avatar's body in order to maintain a visually consistent humanoid. Each avatar was independently driven for each user. The participant in the HMD was isolated from the physical surroundings of the lab, butt could see the hands and feet of their avatar while the participants in the CAVE could only see their own physical body. This meant that participants never saw their own avatar in full, so the fact that both were visually identical was unknown to them.

6.4.6.3 Post-Experimental procedure

After the subjects had completed their task, they were asked to answer two immediate response questions.

- Whether the subject felt he/she was really in the virtual environment or in the real world
- Did the subject feel he/she was with another person?

After the subjects answered the immediate response questions, they were asked to complete the final post questionnaire. This questionnaire consisted of groups of questions dealing with various aspects of how the subject felt about different parts of the VR and is appended in section 8.

Once the questions were answered, the experimenter had a short interview with the subject to cover any areas that might have been unique to that particular experimental run. It was also noticed that most of the experimental subjects had a lot to contribute verbally and wanted to meet their conversational partner in reality.

6.4.7 OUTCOME: EYE-GAZE EXPERIMENTS

The results from this experiment as with others were analysed by Prof. Mel Slater using a statistical procedure called *logistic regression*.

6.4.7.1 Methodology: Data collection

As already mentioned the answers to the online questionnaire were collected and stored electronically and analysed. The questions covered some background information about the subject. It took a measure of their social anxiety. The post questionnaire covered their feelings on the VR experience. Spatial data from both users were collected and time stamped during the experiment as well. The subject in the ReaCTor was video taped and all the interviews held with the subjects were audio taped.

In this study 'quality of communication' was obtained with the same response variables as it was in the earlier study [(33)]. These response variables are (n ≡ number of questions used to obtain the measure):

- *Face-to-face:* The extent to which the conversation was like a real face-to-face conversation (n=6)
- *Involvement*: The extent to which the participants experienced involvement in the conversation. (n=2)
- *Co-presence*: The extent of co-presence[*] between the participants. (n=2)
- *Partner Evaluation*: The extent to which the conversational subjects positively evaluated their partner, and the extent to which the conversation was enjoyed. (n=5).

In the earlier study these were each based on a series of questions on a 9-point Likert scale. The only difference here was that a 7-point scale was used. In addition for this study, as explained earlier, a number of additional response variables were measured. Perceived co-presence was split into three different components

- The sense of being with another person (simply called 'co-presence'),
- Spatial co-presence – the sense of being in the same shared space as the other person

- The degree of experienced personal contact.

These were in turn each measured by grouping a series of 7-point Likert scale questions. The components were as follows:

Co-presence (n=5)

- Alone I felt alone
- Together I did not feel my partner and I were together
- Company I had a sense of being in the company of my conversation partner
- With I had a sense of being with the other person
- Behavealone I behaved as if there was nobody in the virtual room with me

Spatial Co-presence (n=2)

- Sharedspace I felt my partner and I were in a shared space
- Roomspace It did not feel as though my partner and I were in the same room

Personal Contact (n=2)

- Personal I had a real impression of personal contact with my conversational partner
- Interpersonal The interaction did not seem very personal to me

In addition, perceived co-presence was used, which was a merger of these three (n=9). Perceived avatar realism (n = 4) was also measured as the extent to which the subject perceived the avatar as real. This was constructed from the following individual components:

- Realappearance The avatar's appearance was realistic
- Realbehaviour The avatar's behaviour was realistic
- Expressive The avatar was not expressive
- Respond The avatar did not always seem to respond appropriately to me

Avatar facilitation (n=8) was the extent to which the avatar helped the subject in understanding aspects of their partner's behaviour. It included:

- Helpagree Come to an agreement with your partner
- Helpcontact Build a sense of personal contact with your partner
- Helpfeeling Understand how your partner was feeling
- Helpimpression Form a clear impression of your partner

---

* The sense of being with and interacting with another person rather than with a computer interface

- Helpknow   Get to know your partner
- Helpnatural   Have a natural conversation
- Helpreaction   Understand your partner's reactions to your conversation
- Helptrust   Trust your partner

As human (n=2) was the extent to which the avatar was perceived as being like a human. It consisted of two components:

- Likeperson   Did the avatar seem more like a person, or more like a computer?
- Personfront   It was as if I had a person in front of me

There were other questions asked from the subjects but they were not significant.

6.4.7.2 Methodology: Analysis

Each response variable is constructed from a set of 'n' questions. For each question the number of 'high responses' i.e. 6 or 7, where the maximum score for that response variable is 7. The negative set questions (i.e. questions with the word *not*) were reversed in scale to get them all in the same direction. This can then be analysed using *logistic regression*, as in the previous study, and as has been used in several previous studies. This is a conservative method of analysis, and has the advantage of never using the ordinal questionnaire responses as if they were on an interval scale.

The response variables may be thought of as counts of 'successes' out of n trials, and therefore naturally have a binomial distribution, as required in logistic regression. The right-hand-side of the regression consists of only one of two factors (the type of avatar and the type of gaze animation).

In this regression model the *deviance* is the appropriate goodness of fit measure, and has an approximate chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model is a good fit (at the 5% significance level). The change in deviance as variables are deleted from or added to the current model is especially useful, since this indicates the significance of that variable in the fitted model. Here a large change of deviance indicates the degree of significance, i.e., the contribution of the variable to the overall fit.

6.4.7.3 Results

The detailed version of the quantitative results is given in section 8. The type of avatar, type of gaze model, age of participant and participant role (mayor or baker) were significant in this experiment.

Just as in the first experiment [(33)], for this response variable, the person who played the role of the baker tended to have a lower face-to-face response count than the person who played the mayor.

The impact of the gaze model is different depending on which type of avatar is used. For the non-realistic avatar, the inferred gaze behaviour reduces face-to-face effectiveness. For the photo-realistic avatar the inferred gaze behaviour increases effectiveness. It was inferred that there must be a consistency between the visual appearance of the avatar and the type of behaviour that it exhibits with respect to eye-gaze.

In the original paper [(33)] it was found that the 'inferred gaze' model results in a higher face-to-face response than the 'random gaze' model. However, it was not known if the head tracking had a more significant role to play in the outcome of the results. In this study the same head tracking was used in all conditions. The present result therefore resolves the ambiguity of the original study: independently of head tracking, in the case of the photo-realistic avatar, the 'inferred gaze' model results in significantly higher face-to-face than the 'random gaze' model.

The SAD (Social Avoidance and Distress) variable is based on a standard questionnaire that assesses the degree of social anxiety of the person. Since this was a social encounter it was important to factor out influences of people's differing responses to social situations. In this model the greater the degree of social anxiety the lower the face-to-face effectiveness.

This fit of the model excludes one individual who scored unusually highly on the SAD questionnaire (though actually the results are the same even if this individual is included).

Finally, age was found to be significant, and positively associated with the response – older people are more likely to have rated their experience as like face-to-face. The relationship between face-to-face and the independent variables 'type of avatar' and 'type of gaze' is the same whether or not these additional explanatory variables (age, SAD) are included.

These variables are included in an attempt to statistically factor out significant differences between people that may have contributed to the overall result. The logistic regression analysis suggests that for 3 out of the 4 response variables, there is a significant interaction effect between type of avatar and type of gaze.

- For involvement, there is no significant effect of the avatar and gaze models.

- For co-presence and partner evaluation, there is a significant interaction effect between these two variables.

In each case the realistic avatar has a higher response when used with the inferred gaze model. The overall conclusion is that when the avatar is simple, the better (inferred) gaze model does not improve quality of communication, and indeed there seem to be some cases where it may make things worse.

However, for the more realistic avatar, the inferred gaze model improves quality of communication. The evidence suggests that there should be some consistency between the type of avatar and the type of gaze model that is used: the more realistic the avatar appearance the better the gaze model that should be used.

Some of the comments from the subjects are:

- I enjoyed the experience. The avatar made the situation slightly comical, which I felt actually made our conversation more comfortable. With respect to the scenario I felt that the whole experience would have been
- More realistic avatar would have been better.
- It was a good experience!
- When I went through the door, I was preparing to meet a person and carry out a conversation. I did and it was fun.
- I could not entirely forget about being in a laboratory because of the appearance of the avatar. Perhaps I would have done if it had not been a big blue stick-man. I really felt like I was sharing the experience and room space with my partner, but I was disappointed that she did not challenge me more in the conversation, I had prepared myself for a fight or at least a row!
- The environment would have seemed more real if there had been accessories like furniture etc. (items of daily living) in the room. Also, the avatar did not seem to appear life-like at all; and his movements did not seem natural; the face was expressionless. The sense of movement and presence were good and so was the conversation!
- A large part of a normal conversation - especially given the delicate nature of the subject, involves a lot of facial expressions & gestures, which play a large part in the conversation. With the absence of the two, voice and the subtle changes in tone became a much more important part of conveying feeling and mood. After realising the fact that the avatar conveyed very little of the persons actual physical/emotional state, it became even less believable, and I found myself concentrating more and more on the voice. Also, the slightly jerky display distracted from the sense of reality and "being there" - especially as all movements were completely virtual.
- Cool. I would like to see the video of my and the avatar's movements. Has increased my interest in doing a post grad study or a study in virtual reality computer technology. I felt it had a good chance of being improved to a level that where a person can't tell the difference between the real world and the virtual reality world.

- I wondered whether the other person could realise that I could not stop smiling at his ridiculous claims and comments!
- Towards the end of the conversation my partner was looking around the room and the ceiling, which gave me the impression she was not concentrating, the response I had to this was very realistic, as if a real person was not concentrating on what I had to say.

This points to the overall hypothesis: Does the same result hold for other aspects of avatar behaviour than eye gaze. It cannot be claimed, that results from the eye-gaze experiments will generalise to avatar behaviour, but the results found for eye-gaze substantiates the hypothesis that the same might be true for posture.

6.5 SHORT COMINGS REALISED IN PLUGIN SO FAR

Over various pilots run to get this experiment running, some pitfalls were uncovered. The original simplistic avatar, Bender, ran the plugin without a problem. However when the realistic avatars, Jack and Nancy, were adapted to fit into the plugin, the animations did not *look* satisfactory. A lot of small adjustments had to be made to the avatar and ideally these adjustments should be made in a general sense to the plugin.

6.5.1 TREE: DEFINE HIERARCHY OF BENDER

One such adjustment is the storage of the hierarchy and the translations that relate the children and sibling joints to the root node. At the moment the plugin does not have to do the calculations to keep the hierarchy. In theory, if the local co-ordinates (in the form of a matrix) of an end-effector are known, then multiplying the local matrix with all the local matrices up the hierarchy all the way to the top node should give the world co-ordinates of the foot.

At the moment, DIVE allows the use of updating the positions of the body parts via world co-ordinates as well as local-co-ordinates. This feature was used through the whole of the plugin. This makes debugging problematic. The plugin has to be rewritten to only update in world co-ordinates. In order for this to be made possible, the hierarchy of the avatar structure should be stored in the plugin in form of a data structure similar to a *Tree*. The standard rotations of the body parts in accordance to H-Anim should also be stored in this data structure.

6.5.2 QUATERNIONS

At the moment, the avatar's rotations and translations are kept in separate matrices. There is no scaling in order to make the avatar suitable for people shorter than about 145 centimetres. There was also the problem mentioned under the geometrical IK* procedure used in the elbow positioning under section 6.3.4.1.

The use of quaternion mathematics also gets rid of the *Gimbal lock* problem. *Gimbal lock* is the name given to a problem that occurs with the use of Euler† angles. Since the final rotation matrix depends on the order of multiplication, it is sometimes the case that the rotation in one axis will be mapped onto another rotation axis. This problem has happened in the arms of the avatars in some poses of the arms.

A quaternion is an extension of the complex number. A complex number is an imaginary number that is defined in terms of *i*, the imaginary number which is the square root of –1 i.e. *i* * *i* = -1. In a quaternion, instead of just *i*, there are three numbers that are all square roots of -1, denoted by '*i*', '*j*', and '*k*'.

The quaternion, q, can be represented as

*q = w + x*i + *y*j + zk ; w *is a real number, and* x, y, *and* z *are complex numbers.*

q=[w,v] ; v = *(x, y, z) is called a "vector" and* w *is called a "scalar".*

The variable '*v*' is a vector in 4D space, which is totally unintuitive to visualize. Quaternions are not vectors, so it is not possible to apply preconceived vector mathematics to quaternions. It is possible to visualize unit quaternions as a rotation in 4D space where the (x,y,z) components form the arbitrary axis and '*w*' forms the angle of rotation.

All the unit quaternions form a sphere of unit length in the 4D space. A 180-degree rotation of a quaternion is obtained by inverting the scalar '*w*' component. Only unit quaternions can be used for representing orientations. A unit quaternion has the following property

$$w^2 + x^2 + y^2 + z^2 = 1$$

The plugin will be incorporating the use of quaternions to represent rotations and translations of body parts in the future and the necessary studies in this area is ongoing.

* Inverse Kinematics

† Euler angles are the set of rotation angles, which specify the rotation in each of the X, Y and Z rotation axis.

6.5.3 CONSTRAINTS NECESSARY

There should be some kind of collision detection incorporated into the avatar to stop the body parts from *going into* the other body parts of the avatar. It is also necessary to define regions around the avatar and animate body parts differently under different regions. For instance if the elbow is behind the shoulder, it is unlikely it is bending the wrong way. The geometrical IK functions in the plugin at the moment get a solution, which makes the elbow appear double-jointed.

6.6 LAB DEMONSTRATION

Slightly less research based than the other experiences, this academic duty was more for the development of *people's skill*. The duties undertaken in the three-month job involved demonstrating (with the aid of a colleague) the use of C++ to program the manipulation of various commonplace data structures, objects and functionalities. It led to the development and improvement of organisational, management (both time, anger and work), communication skills, which aid one to work in a team under stressful and demanding conditions with seemingly fast approaching deadlines. It also gave an alarming but invaluable insight into what lecturing a group of *not* very enthusiastic student might entail.

6.7 IMMEDIATE WORK PLANS

There a few immediate changes that needs to be made to the plugin. This is to enable the plugin to be ported into a platform independent of DIVE.

- All the DIVE commands have to stripped out
- All the updates in co-ordinates and rotations of the body parts have to be in terms of world co-ordinates
- A tree structure has to be made to hold details of the transformations of the avatar along with holding the hierarchical structure of the avatar.
- All the mathematics done on the transformation matrices in the avatar body parts has to be done in quaternions.
- Psychological survey of posture and the effect emotions have on poses.
- …

At the end of this PhD a thesis of the following structure is foreseen.

- Abstract
    - Keywords
    - Copyright statement

  - Acknowledgements
- Table of Contents
- Chapter 1: Introduction
  - Overall aim and motivation
  - Hypothesis and Discussion
- Chapter 2 - 3: Background (2 chapters)
  - Graphics Literature Review
  - Psychological Literature Review
- Chapter 4 - 6: Experiments (3 chapters/sub chapters)
  - Aims and objectives
  - Methodology
  - Data collected
  - Analysis
  - Results and Conclusions
- Chapter 7: Final Conclusion
- Chapter 8: Future Work
- Appendices
  - Acronyms
  - References
  - Extra data and results
  - Papers Published

7.0 CONCLUSION

It has been experimentally inferred that realistic avatar require behaviour realism in order to increase the levels of co-presence and realism for virtual environment users. This was done using the non-verbal behaviour inferred from eye-gaze. It has also been inferred from the Internet-2 project that emotions can be conveyed across a virtual environment using basic head movements and tone of voice.

This work has paved a foundation to carry on this research into extending behaviour communication to body posture. In the second year of this thesis, it is hoped that the basic building blocks to determining and computational model relating emotional state of mind and various poses will be available. Most of the second year of this thesis will involve researching psychological archives and papers to build a system, which relates emotions to posture. The third year can be used to test the results from the second year.

This year has provided me with a lot of skills and improved many others. These skills include: organisational, research, project management, team skills, time management, paper writing, presentation, communication and report writing skills. It has been very satisfying and academically profitable. It is hoped that future research and continuing study aids in the fulfilment of this thesis and in the procurement of further invaluable skills.

8.0 ADDENDUM

8.1 MORE ON THE INTERNET-2 PROJECT

A questionnaire was designed to assess the behaviour and views of the subjects. This questionnaire was administered online to each of the UCL subjects straight after their experience. It was also given to two of the UNC helpers after their first experience with a UCL subject. The questionnaire obtained responses on each of the following:

- *Demographic information*: such as age, gender, status etc.
- *Task Performance*: assessments of self and other's performance, the degree of harmony and cooperation between the participants.
- *Co-presence*: the sense of being together rather than interacting through a computer interface.
- *Similarity to real life*: the extent to which the experience was similar to moving an object together with someone in real life.
- *Mood assessment*: Assessment of the mood of the other person.

Each of these (apart from the demographic data) was assessed on a 1 to 7 scale, as shown in the Questionnaire. There was also an open-ended question where subjects could write their answer:

> *Please enter your comments. Things you could consider are:*
> *Things that hindered you or the other person from carrying out the task;*
> *What you think of the person you worked with;*
> *And any other comments about the experience and your sense of being there with another person.*
> *What things made you "pull out" and more aware of the computer...*

The purpose of this was to try to get behind the purely quantitative results to what the subjects were thinking about their experience.

8.1.1 QUANTITATIVE RESULTS

Co-presence was assessed from the following questions:

- To what extent, if at all, did you have a sense of being with the other person?
- To what extent were there times, if at all, during which the computer interface seemed to vanish, and you were directly working with the other person?

- When you think back about your experience, do you remember this as more like just interacting with a computer or working with another person?
- To what extent did you forget about the other person, and concentrate only on doing the task as if you were the only one involved?
- During the time of the experience, did you think to yourself that you were just manipulating some screen images with a mouse-like device, or did you have a sense of being with another person?
- Overall rate the degree to which you had a sense that there was another human being interacting with you, rather than just machine?

Each was measured on a 1-7 scale, and in the analysis the directions adjusted so that 7 always means the highest co-presence, and 1 the lowest. The overall mean co-presence was 3.8 ± 1.1.

In order to examine the association between other variables and co-presence, we score each subject by the number of 'high' scores on the 6 individual questions. A 'high' score in answer to a question is one that is above 4 out of 7.

Hence the overall score is actually a count of the number of high scores out of the 6 questions - e.g., if the result is 4 then it means that in 4 out of the 6 questions the subject responded with a score that was 5, 6 or 7. This is a conservative way to treat the results, and has been used several times before, for example [(33)].

We do this in order to carry out a logistic regression analysis between the co-presence results and the other variables. Here we report only results that are significant at the 5% level.

In the following results the means and standard deviations for the explanatory variables are shown in brackets after the question statement. Co-presence is most significantly and positively associated with two other variables in one single fitted model:

- Please give your assessment as to how well you contributed to the successful performance of the task. (4.2 ± 1.3).
- Please give your assessment as to how well the other person contributed to the successful performance of the task. (3.9 ± 1.1).

In each case the higher the score the higher the co-presence score (and these two explanatory variables are uncorrelated). In other words this demonstrates a link between co-presence and the subjectively assessed level of task performance. The coefficients for the two variables are almost equal, indicating that self- and other-performance were equally weighted.

Other variables that are *positively* associated with co-presence, but individually, not within the same overall model:

- To what extent were you and the other person in harmony during the course of the experience? (3.9 ± 1.7).
- Think about a previous time when you co-operatively worked together with another person in order to move or manipulate some real thing in the world (for example: shifting some boxes, lifting luggage, moving furniture and so on). To what extent was your experience in working with the other person on this task today like the real experience, with regard to your sense of doing something together? (2.9 ± 1.3).
- Please give your assessment of how well you and the other person together performed the task (4.3± 1.4).

The second question is quite important, since it gives an overall view of how "*real*" the collaboration felt. The mean response (2.9 ± 1.3), indicating overall that the degree of similarity to moving an object in real life was relatively low.

The following was *negatively* associated with co-presence:

- To what extent, if at all, did you hinder the other person from carrying out the task? (3.3 ± 1.8).

In other words the more the subject believed that they had hindered in carrying out the joint task the lower the sense of co-presence. Interestingly co-presence was not correlated with the degree to which the subject felt that the other person had hindered the carrying out of the task.

Almost all of the variation in co-presence can be explained by just two variables taken together in one model:

- Please give your assessment as to how well you contributed to the successful performance of the task. (4.2 ± 1.3).
- If you had a chance, would you like to meet the other person? (4.6 ± 1.7).

The higher the self-assessed contribution of the subject, and the more she or he wishes to actually meet the other person, the greater the degree of co-presence. UCL subjects were able to assess the mood state of the UNC person (recall that the UNC person was acting as happy or as depressed). The mean mood score (i.e., estimated degree of happiness) amongst those who experienced the depressed acting was: 2.5 ± 0.8, and for those who experienced the happy state: 3.9 ± 1.4. The difference is significant at the 5% level. The relatively high variance amongst the "happy" responses probably reflects that it seemed to be much harder for the UNC subjects to convincingly act in a "happy" manner than in a depressed manner.

8.2 PRE-QUESTIONNAIRE USED IN EYE-GAZE

The pre-questionnaire used was a standard psychological questionnaire used to measure an individual's prone social anxiety. It is called the SAD questionnaire. SAD stands for Social Avoidance and Distress. The questionnaire contains a number of 'yes/no' questions. The questions are listed below.

- I feel relaxed even in unfamiliar social situations
- I try to avoid situations which force me to be very sociable
- It's easy for me to relax when I am with strangers
- I have no particular desire to avoid people
- I often find social settings upsetting
- I usually feel calm and comfortable in social situations
- I am usually at ease when talking to someone of the opposite sex
- I try to avoid talking to people unless I know them well
- If the chance comes to meet new people, I often take it
- I often feel nervous or tense in casual get-togethers in which both sexes are present
- I am usually nervous with people unless I know them well
- I usually feel relaxed when I am with a group of people
- I often want to get away from people
- I usually feel uncomfortable when I am in a group of people I don't know
- I usually feel relaxed when I meet someone for the first time
- Being introduced to people makes me tense and nervous
- Even though a room is full of strangers I may enter it anyway
- I would avoid walking up to and joining a large group of people
- When my superiors want to talk to me, I talk willingly
- I often feel on the edge when I talk to a group of people
- I tend to withdraw from people
- I don't mind talking to people at parties or social gatherings
- I am seldom at ease in a large group of people
- I often think up excuses in order to avoid social engagements
- I try to avoid formal social occasions
- I usually go to whatever social engagements I have
- I find it easy to relax with other people

8.3 POST-QUESTIONNAIRE USED IN EYE-GAZE

The post-questionnaire used was more in depth to the experience the user had in the VE. Most of the questions on the post experimental questionnaire are graded on a scale of 1 to 7.

The questions dealing with how the subject evaluated their conversational partner:

- I could readily tell when my partner was concentrating on what I was saying
- I was able to take control of the conversation when I wanted to
- It was easy for me to contribute to the conversation
- The conversation seemed highly interactive
- I found it easy to keep track of the conversation
- There were frequent inappropriate interruptions
- This felt like a natural conversation
- I felt completely absorbed in the conversation
- I had a real impression of personal contact with my conversational partner
- I was very aware of my conversational partner
- My partner was friendly
- My partner did not take a personal interest in me
- I trusted my partner
- I enjoyed talking to my partner
- I would be interested in meeting my partner face to face

The questions dealing with general social presence and co-presence:

- This felt like a phone conversation
- I felt alone
- I did not feel my partner and I were together
- I had a sense of being in the company of my conversation partner
- I behaved as if there was nobody watching me
- The interaction did not seem very personal to me
- It was as if I had a person in front of me
- I felt my partner and I were in a shared space
- Most of my attention was focused on my partner's voice
- My partner and I frequently made eye contact
- The way my partner looked at me appeared natural
- I had a sense of being with the other person
- It did not feel as though my partner and I were in the same room
- It seemed as if my partner was observing me
- I behaved as if there was nobody in the virtual room with me

The questions dealing with the avatar:

- I tried to wave, smile and/or shake hands to the avatar
- The avatar had no impact on the conversation
- I paid less attention to the avatar as the conversation went on
- The avatar's appearance was realistic
- The avatar's behaviour was realistic
- The avatar did not always seem to respond appropriately to me
- I believe the avatar must have resembled my conversation partner
- The avatar was not expressive
- Understand how your partner was feeling
- Understand your partner's reactions to your conversation
- Form a clear impression of your partner
- Build a sense of personal contact with your partner
- Get to know your partner
- Have a natural conversation
- Trust your partner
- Come to an agreement with your partner
- When you first saw the avatar, to what extent did it become the person you were meeting with?
- Now consider your response over the course of the whole experience. To what extent did it become the person you were meeting with?
- In the beginning of the conversation, to what extent did you pay attention to the avatar?
- Now consider your response over the course of the whole experience. To what extent did you pay attention to the avatar?
- To what degree do you think the avatar's actions were controlled by your partner?
- To what degree do you think the avatar's actions reflected what your partner was actually doing?
- Did the avatar seem more like a person, or more like a computer?

The questions dealing with presence in the virtual environment:

- During the experience I was aware of background sounds from the laboratory...
- How dizzy, sick or nauseous did you feel resulting from the experience, if at all?
- I felt sick or dizzy or nauseous during or as a result of the experience...
- I had a sense of being there in the room...
- There were times during the experience when the room was the reality for me...
- When you think back about your experience, do you think of the room more as images that you saw, or more as somewhere that you visited? The room seemed to be more like...
- During the course of the experience, which was strongest on the whole, your sense of being in the room, or of being in the real world of the laboratory? I had a stronger sense of...
- Overall, how well do you think that you achieved your task? I achieved my task...
- During the experience I was thinking that I was really in the laboratory...
- During the course of the experience, how much were you aware of the experimenters?

Not all of the questions above were used in analysing the outcome of the experiment.

8.4 RESULTS FROM THE EYE-GAZE EXPERIMENTS

8.4.1 RESPONSE VARIABLES

There were two main questions of interest in this experiment. The first was to disambiguate the confounding between inferred eye gaze and head-movement generated by head tracking – a question that remained from the previous experiment.

The second major issue was to examine the impact of differing avatar representations. In the previous experiment, all eye gaze conditions were used with the same relatively photo-realistic avatar. What happens when the avatar is non-photo-realistic – in particular can high quality avatar behaviour compensate for low realism in avatar representation?

What are the relative impacts on quality of communication of both avatar realism and eye gaze animation? We take eye gaze animation as a specific (though important) instance of avatar behaviour. We cannot claim, of course, that results we find for eye-gaze animation will generalise to avatar behaviour, but results found for eye-gaze will generate hypotheses for studies of further aspects of avatar behaviour.

In this study we studied ‘quality of communication’ with the same response variables as in the earlier study. For convenience we repeat these response variables here:

Face-to-face: The extent to which the conversation was experienced as being like a real face-to-face conversation (n=6)
Involvement: The extent to which the participants experienced involvement in the conversation. (n=2)
Co-presence: The extent of co-presence between the participants - that is, the sense of being with and interacting with another person rather than with a computer interface. (n=2)
Partner Evaluation: The extent to which the conversational subjects positively evaluated their partner, and the extent to which the conversation was enjoyed. (n=5).

In the earlier study these were each based on a series of questions on a 9-point Likert scale. The only difference here was that a 7-point scale was used, otherwise the questions were the same. The n’s in the above table refer to the number of questions that were used to make up the corresponding construct (again, identical to the previous study).

In addition for this study, as explained earlier, a number of additional response variables were measured:

Perceived co-presence was split into three different components

- The sense of being with another person (simply called 'co-presence'),
- 'Spatial co-presence' – the sense of being in the same shared space as the other person
- The degree of experienced personal contact.

These were each measured by grouping a series of 7-point Likert scale questions. The components were as follows:

Co-presence (n=5)

Alone	I felt alone
Together	I did not feel my partner and I were together
Company	I had a sense of being in the company of my conversation partner
With	I had a sense of being with the other person
Behavealone	I behaved as if there was nobody in the virtual room with me

Spatial Copresence (n=2)

Sharedspace	I felt my partner and I were in a shared space
Roomspace	It did not feel as though my partner and I were in the same room

Personal Contact (n=2)

Personal	I had a real impression of personal contact with my conversational partner
Interpersonal	The interaction did not seem very personal to me

In addition, **perceived-copresence** was used, which was a merger of these three (n=9). **Perceived avatar realism** (n = 4) measured the extent to which the subject perceived the avatar as real. This was constructed from the following individual components:

Realappearance	The avatar's appearance was realistic
Realbehaviour	The avatar's behaviour was realistic
Expressive	The avatar was not expressive
Respond	The avatar did not always seem to respond appropriately to me

Avatar facilitation (n=8) was the extent to which the avatar helped the subject in understanding aspects of their partner's behaviour. It included:

Helpagree	Come to an agreement with your partner
Helpcontact	Build a sense of personal contact with your partner
Helpfeeling	Understand how your partner was feeling
Helpimpression	Form a clear impression of your partner
Helpknow	Get to know your partner
Helpnatural	Have a natural conmversation
Helpreaction	Understand your partner's reactions to your conversation
Helptrust	Trust your partner

As human (n=2) was the extent to which the avatar was perceived as being like a human. It consisted of two components:

Likeperson	Did the avatar seem more like a person, or more like a computer?
Personfront	It was as if I had a person in front of me

There were other variables used but they were not significant.

8.4.2 METHOD OF ANALYSIS

The same method of analysis was used as in the previous analysis. Each response variable is constructed from a set of n questions. For each question we count the number of 'high responses' (that is a response on the Likert Scale that is 6 or 7, where the maximum score for that response variable is 7. mention about reversing to get them all in the same direction. Therefore each response variable is a count out of n possible high scores.

For example, for the overall copresence variable, n = 9, so the response is the number of 'high scores' out of these 9 questions. This can be analysed using logistic regression, as in the previous paper, and as has been used in several previous studies. This is a conservative method of analysis, and has the advantage of never using the ordinal questionnaire responses as if they were on an interval scale.

The response variables may be thought of as counts of 'successes' out of n trials, and therefore naturally have a binomial distribution, as required in logistic regression. In the case where the right-hand-side of the regression consists of only one two factors (in the case the type of avatar and the type of gaze animation) this is equivalent to a two-way ANOVA but using the more appropriate binomial distribution rather than the Normal. Of course other covariates may be added into the model, thus being equivalent to two-way ANOCOVAR.

In this regression model the *deviance* is the appropriate goodness of fit measure, and has an approximate chi-squared distribution with degrees of freedom depending on the number of fitted parameters. A rule-of-thumb is that if the deviance is less than twice the degrees of freedom then the model is a good fit (at the 5% significance level). The change in deviance as variables are deleted from or added to the current model is especially useful, since this indicates the significance of that variable in the fitted model. Here a large change of deviance indicates the degree of significance, i.e., the contribution of the variable to the overall fit.

8.4.3 RESULTS

We consider the results of the face-to-face response variable in detail, and then analyses and results of all the other responses are very similar to this.

Table 1

Mean Face-to-face score (count out of 6) ± S.E.

Type of avatar/Type of gaze	Random Gaze	Inferred Gaze
Non-realistic avatar	4.2±0.52	2.9±0.47
More realistic avatar	2.2±0.43	3.9±0.56

Table 1 shows the raw means of the face-to-face score. An inspection suggests that there is a strong interaction effect – that within each row and column there is a significant difference between the means, but that there is no significant difference between the top left and bottom right cells. This is borne out by a logistic regression, on the independent variables.

Table 2: Logistic Regression with Face-to-face as the Response Variable

Variable	**Deviance** χ^2
Type avatar • type gaze	22.03 (+)
Age	7.8 (+)
Role (baker)	10.0 (-)
SAD	15.7 (-)

Overall deviance = 79.9, d.f. = 40

The deviance column shows the increase in deviance that would result if the corresponding variable were deleted from the model. The tabulated χ^2 5% value is 3.841 on 1 d.f. and all d.f.'s below are 1. The sign in brackets after the χ^2 value is the direction of association of the response with the corresponding variable.

Table 2 shows the results of the regression analysis for face-to-face as the response variable. Each of these terms is significant at the 5% level of significance (i.e., none can be deleted without significantly reducing the overall fit of the model).

Type of avatar, type of gaze model, age of participant and participant role were significant. Role refers to whether the subject is playing the mayor or the baker. Just as in the first experiment, for this response variable, the person who played the role of the baker tended to have a lower face-to-face response count than the person who played the mayor.

The formal analysis demonstrates the very strong interaction effect between the type of avatar and the type of gaze (denoted here by the • symbol). In other words the impact of the gaze model is different depending on which type of avatar is used. For the non-realistic avatar, the more realistic (inferred) gaze behaviour reduces face-to-face effectiveness. For the more realistic avatar the more realistic (inferred) gaze behaviour increases effectiveness. It is as if there must be a consistency between the visual appearance of the avatar and the type of behaviour that it exhibits – low fidelity appearance demands low fidelity behaviour, and correspondingly higher fidelity appearance demands a more realistic behaviour model (with respect to eye gaze).

In the original paper it was found that the 'inferred gaze' model results in a higher face-to-face response than the 'random gaze' model. However, in that study the result was confounded with head tracking. In this study there was the same head tracking in all conditions. Moreover, in the original study the avatar used was photo-realistic. The present result therefore resolves the ambiguity of the original study: independently of head tracking, in the case of the photo-realistic avatar, the 'inferred gaze' model results in significantly higher face-to-face than the 'random gaze' model.

The SAD (Social Avoidance and Distress) variable is based on a standard questionnaire that assesses the degree of social anxiety of the person. Since this was a social encounter we wished to factor out influences of people's differing responses to social situations. In this model the greater the degree of social anxiety the lower the face-to-face effectiveness. This fit of the model excludes one individual who scored unusually highly on the SAD questionnaire (though actually the results are the same even if this individual is included). Finally, age was found to be significant, and positively associated with the response – older people are more likely to have rated their experience as like face-to-face. The relationship between face-to-face and the independent variables 'type of avatar' and 'type of gaze' is the same whether or not these additional explanatory variables (age, SAD) are included. These latter are included in an attempt to statistically factor out significant differences between people that may have contributed to the overall result.

Table 3

Fitted Logistic Regression for the Original Response Variables

Non-significant terms are indicated by -

Fitted Variable	**Face-to-face Deviance χ^2**	**Involvement**	**Co-presence**	**Partner evaluation**
Type avatar • type gaze	22.03 (+)	-	9.7 (+)	5.0 (+)
Age	7.8 (+)	16.9 (+)	14.1 (+)	-
Role (baker)	10.0 (-)	-	-	6.2 (-)
SAD	15.7 (-)	-	-	-
Overall deviance	79.9	67.7	60.5	125.0
Overall d.f.	40	46	43	44

Table 4: Means and Standard Errors for the Response Variables

Response:	Random Gaze	Inferred Gaze
Involvement:		
Simple Avatar	1.3±2.9	1.3±0.2
Realistic Avatar	0.9±0.2	1.2±0.22
Co-presence:		
Simple Avatar	1.2±0.2	0.7±0.2
Realistic Avatar	0.3±0.1	1.1±0.3
Partner Evaluation:		
Simple Avatar	2.6±0.5	2.2±0.4
Realistic Avatar	1.8±0.5	2.8±0.5

This same overall pattern of results is repeated for each of the remaining response variables. Table 3 shows the results for all of the variables used in the original study.

The logistic regression analysis suggests that for 3 out of the 4 response variables, there is a significant interaction effect between type of avatar and type of gaze. For involvement, there is no significant effect of the avatar and gaze models. For co-presence and partner evaluation, there is a significant interaction effect between these two independent variables. In each case the realistic avatar has a higher response when used with the inferred gaze model. This can be seen also from the raw table of means and standard errors.

Table 5

Significant Fitted Logistic Regression for the New Response Variables

Non-significant terms are indicated by -

Fitted Variable	**Co-presence Deviance χ^2**	**Spatial Co-presence**	**Personal Contact**	**Overall Co-presence**	**Avatar Facilitation**
Type avatar • type gaze	4.0 (+)	15.9 (+)	10.8 (+)	20.3 (+)	5.5
Age	20.8 (+)	12.7 (+)	19.7 (+)	47.0 (+)	-
Role (baker)	-	-	-	-	-
SAD	19.2 (-)	-	-	18.1 (-)	-
Overall deviance	131.7	72.7	47.5	154.0	128.3
Overall d.f.	41	43	43	41	44

Table 5 shows the results of the logistic regression for the new response variables. In each case the same result is obtained: the significant interaction effect, with the realistic avatar requiring the inferred gaze model to get the higher response. Finally the response variable ‘as human’ shows the same pattern, but the significance level for the interaction effect on avatar type and gaze is just below 5% when ‘age’ is included as a co-variate (χ^2=3.8), and just above 5% when age is excluded.

It is clear that there is a highly consistent pattern of responses amongst many of the response variables that make up our notion of quality of communication. The overall conclusion must be that when the avatar is simple, the better (inferred) gaze model does not improve quality of communication, and indeed there seem to be some cases where it may make things worse. However, for the more realistic avatar, the inferred gaze model improves quality of communication.

The evidence suggests that there should be some consistency between the type of avatar and the type of gaze model that is used: the more realistic the avatar appearance the better the gaze model that should be used. This points to a further hypothesis: does the same result hold for other aspects of avatar behaviour than eye gaze – e.g., for walking, bending, general facial expression, lip-synching, and so on?

9.0 BIBLIOGRAPHY

[1] Allbeck M. J. and Badler I. N., "Consistent Communication with Control", *Workshop on Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents at Autonomous Agents*, 2001a

[2] Amaya K., Bruderlin A., and Calvert T., "Emotion from Motion", *Graphics Interface*, June 1996, pp. 222-229

[3] Argyle M., "Bodily Communication", Second Edition, 1998b

[4] Argyle M. and Cook M., "Gaze and Mutual Gaze", 1976

[5] Badler I. N. and Allbeck M. J., "Towards Behavioural Consistency in Animated Agents", *Proceedings of Deformable Avatars*, 2000c

[6] Badler I. N., Palmer S. M., and Bindiganavale R., "Animation control for real-time virtual humans", *Communications of the ACM*, 42(8), August 1999, pp. 64-73

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

[8] Baldes S., Literature Survey about Personality and Emotions

[9] Barrientos F., "Continous Control of Avatar Gesture", *International Multimedia Conference*, (Proceedings of the 2000 ACM Workshops on Multimedia), 2000, pp. 5-8

[10] Barrientos F. and Canny J., "Cursive: A novel interaction technique for controlling expressive avatar gesture", *User Interface Software &Technology*, 2001

[11] Basdogan C., Ho C.-H., Srinivasan M. A., and Slater M., "An experimental study on the role of touch in shared virtual environments", *ACM Transactions on Computer-Human Interaction*, 7(4), December 2000e, pp. 443-460

[12] Bates J., "The Role of Emotion in Believeable Agents", *Communications of the ACM*, 37, July 1994f, pp. 122-125

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

[14] Bates J., Loyall B. A., and Reilly S. W., "Integrating Reactivity, Goals, and Emotion in a Broad Agent", 1992b

[15] Becker W., Chapter 2: Metrics, In "*The Neurobiology of Saccadic Eye Movements*", R H Wurtz and M E Goldberg, 1989i

[16] Bickmore W. T. and Cassell J., "How about this weather? Social Dialog with Embodied Conversational Agents", *American Association for Artificial Intelligence*, 2000

[17] Cahn J. E., "Generating Expression in synthesised speech", 1990, MIT

[18] Capin K. T., Pandzic S. I., Thalmann M. N., and Thalmann D., "Integration of Avatars and Autonomous Virtual Humans in Networked Virtual Environments", 1998j

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

[20] Cassell J., Bickmore W. T., Campbell L., Chang K., Vilhjálmsson H. H., and Yan H., "Requirements for an Architecture for Embodied Conversational Characters", *Computer Animation and Simulation*, 1999b

[21] Cassell J., Bickmore W. T., Vilhjálmsson H. H., and Yan H., "More Than Just a Pretty Face: Affordances of Embodiment", 2000l,

[22] Cassell J. and Vilhjálmsson H. H., "Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous", *Autonomous Agents and Multi-Agent Systems*, 2, 1999c, pp. 45-64

[23] Cassell J., Vilhjálmsson H. H., and Bickmore W. T., "BEAT: the Behavior Expression Animation Toolkit", *SIGGRAPH*, 2001n

[24] Chopra S., "Where To Look? Automating Some Visual Attending Behaviours of Human Characters", 29-6-1998, Thesis/Dissertation

[25] Colburn R. A., Cohen F. M., and Drucker M. S., "The Role of Eye Gaze in Avatar Mediated Conversational Interfaces", 31-7-2000, Microsoft Research MSR-TR-2000-81

[26] Collier G., "Emotional Expression", Hillsdale, New Jersey, 1985

[27] Davitz J. R., "The communication of emotional meaning", New York, 1964

[28] EDS, PLM Solutions: Jack

[29] Ekman P., "Emotion in the Human Face", 1982

[30] Ekman P. and Friesen V. W., "Head and Body Cues in the Judgement of Emotion: A Reformulation", *Perceptual and Motor Skills*, 24, 1967p, pp. 711-724

[31] Eyematic, Eyematic

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

[33] Garau M., Slater M., Bee S., and Sasse M. A., "The impact of eye gaze on communication using humanoid avatars", *Conference on Human Factors and Computing Systems*, (Proceedings of the SIGCHI conference on Human factors in computing systems), March 2001q, pp. 309-316

[34] Graham J. A., Ricci Bitti P., and Argyle M., "A cross-cultural study of the Communication of Emotion by Facial and Gestural cues", *Journal of Human Movement Studies*, 1975, pp. 178-182

[35] Grassia F. S., "Believable Automatically Synthesized Motion by Knowledge-Enhanced Motion", 21-8-2000r, Thesis/Dissertation

[36] Gratch J. and Marsella S., "Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents", *International Conference on Autonomous Agents*, 5, 2001s

[37] Hewes G., "The Anthropology of Posture", *Scientific American*, 196, 1957, pp. 123-132

[38] Hilton A., Beresford D., Gentils T., Smith R., and Sun W., "Virtual People: Capturing human models to populate virtual worlds", *IEEE Computer Graphics and Applications*, 1999

[39] Jordan J., Mortensen J., Oliveira M., Slater M., Tay B. K., Kim J., and Srinivasan M. A., "Collaboration in a Mediated Haptic Environment", *Presence and Teleoperators and Virtual Environments*, 11, 2002t

[40] Kshirsagar S. and Magnenat-Thalmann N., "A Multilayer Personality Model", 2002u

[41] Lee S. P., Badler I. N., and Badler J., "Eyes Alive", *SIGGRAPH*, July 2002v

[42] Levine M. D., "A Mind at a Time", First Edition, 2002

[43] Loyall B. A. and Bates J., "Real-time control of Animated Broad Agents", 1993w

[44] Maes P., "Artificial life meets entertainment: lifelike autonomous agents", *Communications of the ACM*, 38(11), November 2002, pp. 108-114

[45] Maes P., Darrell T., Blumberg M. B., and Pentland A., "The ALIVE system: full-body interaction with autonomous agents", *Computer Animation*, April 1995, page 11

[46] Masip J., Non verbal Behaviour Links, *http://www3.usal.es/~nonverbal/introduction.htm*

[47] McCrae R. R. and John O. P., "An introduction to the five-factor model and its applications", *Journal of Personality*, 60(Special issue: The five factor model: Issues and applications), 1992y, pp. 175-215

[48] Mehrabian A., "Silent Messages", First Edition, Wandsworth, Belmont, California, 1971

[49] Met A., "Exploration of Persecutory ideation in virtual environments", 10-9-2001z, Thesis/Dissertation

[50] Mole J., Decoding Body Language, *http://www.johnmole.com/articles18.htm*

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

[52] Niu X. and Vassileva J., "Agent Persona and Modelling Emotions", 26-2-2002

[53] Ortony A., Clore G., and Collins A., "Cognitive Structure of Emotions", 1988}

[54] Patil R., Body watching, *http://www.lifepositive.com/Mind/psychology/body-language/body-gestures.asp*

[55] Pelachaud C., Badler I. N., and Steedman M., "Generating Facial Expressions for Speech", 1994,

[56] Perlin K. and Goldberg A., "Improv: A System for Scripting Interactive Actors in Virtual Worlds", *SIGGRAPH*, 1996

[57] Pertaub D.-P., Slater M., and Barker C., "An Experiment on Fear of Public Speaking in Virtual Reality", *Medicine meets Virtual Reality,* 2001€, pp. 372-378

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

[59] Picard R. W., "Towards Agents that recognise Emotion", 1998,515

[60] Polichroniadis T. P., "High Level Control of Virtual Actors", 2000, Thesis/Dissertation

[61] Rao A. S. and Georgeff M. P., "Modelling rational agents within a BDI-architecture", 1991

[62] Richard N., Codognet P., and Grumbach A., "The InViWo toolkit: Describing autonomous virtual agents and avatars", *Intelligent Virtual Agents*, 2001,

[63] Rosenberg G. B. and Langer J., "A Study of Postural-Gestural communication", *Journal of Personality and Social Psychology*, 2, 1965, pp. 593-597

[64] Salem B. and Earle N., "Designing a non-verbal language for Expressive Avatars", *Collaborative Virtual Environments*, Proceedings of the third international conference on Collaborative virtual environments, 2000

[65] Slater M., Howell J., Steed A., Pertaub D.-P., Garau M., and Springel S., "Acting in Virtual Reality", *ACM Collaborative Virtual Environments*, 2000a, pp. 103-110

[66] Slater M., Pertaub D.-P., and Steed A., "Public Speaking in Virtual Reality: Facing an Audience of Avatars", *IEEE Computer Graphics and Applications*, 19(2), March 1999,,, pp. 6-9

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

[68] Slater M. and Steed A., "A Virtual Presence Counter", *Presence: Teleoperators and Virtual Environments*, 9(2), 2000c, pp. 214-217

[69] Slater M. and Steed A., Chapter 9: Meeting people vitually: experiments in shared virtual environments, In "*The social life of avatars: presence and interaction in shared virtual environments*", New York, 2002

[70] SPARC, Examples of Body Language, *http://www.deltabravo.net/custody/body.htm*

[71] Steed A., Mortensen J., and Frécon E., "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, 2001, pp. 153-164

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

[73] Sun W., Hilton A., Smith R., and Illingworth J., "Layered Animation of Captured Data", *Computer Animation and Simulation*, September 1999

[74] Swedish Institute of Computer Science, The DIVE home page, *http://www.sics.se/dive/*

[75] Thalmann D. and Caicedo A., "Virtual Humanoids: Let Them Be Autonomous without Losing Control", 2000,

[76] Thalmann M. N., Kalra P., Moccozet L., Sannier G., Aubel A., and Thalmann D., "Real-time animation of realistic virtual humans", *IEEE Computer Graphics and Applications*, 18(5), 1998ˆ, pp. 42-55

[77] Thórisson R. K., "Communicative Humanoids: *A Computational Model of Psychosocial Dialogue Skills*", 19-7-1996‰, Thesis/Dissertation

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

[79] Velásquez J. D., "Modeling Emotions and Other Motivations in Synthetic Agents", *American Association for Artificial Intelligence*, 1997‹

[80] Vertegaal R. and Ding Y., "Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronisation?", 2002

[81] Vilhjálmsson H. H., "Autonomous Communicative Behaviours in Avatars", 8-10-1996Œ

[82] Vilhjálmsson H. H. and Cassell J., "BodyChat: Autonomous Communicative Behaviors in Avatars", 9-5-1998

[83] Vosinakis S. and Panayiotopoulos T., "SimHuman: A platform for real-time virtual agents with planning capabilities", *Intelligent Virtual Agents*, 2001Ž

[84] Web3D Consortium, Humanoid Animation Working Group of the WEB3D Consortium, *http://www.hanim.org/*