DESIGNING A VISUAL COMPONENT OF COMMUNICATION WITHIN 3D AVATAR VIRTUAL WORLDS

by

NICHOLAS JOHN EARLE

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Arts and Education

December 2001

NICHOLAS JOHN EARLE
Designing a Visual Component of Communication Within 3D Avatar Virtual Worlds

ABSTRACT

Over the last few years 3D avatar virtual worlds (AVW) have emerged on the Internet. These are computer generated, multi-user, graphical spaces within which people meet, form social groups and interact with each other in real time, typically through the exchange of text or audio messages. Each user is represented within the space by a digital image known as an *avatar*, which is usually humanoid in form, and is predominantly under the control of the person it represents.

This thesis describes a creative project that is concerned with aspects of social communication between users of AVWs. In particular, an avatar is designed that is capable of performing body language, and a set of useful gestures are implemented that support aspects of social interaction and integrate with verbal discourse in a meaningful way. In addition to this, a number of scenic properties are derived that enable better comprehension of the non verbal communication, e.g. spatial arrangement, camera position and lighting effects.

The research consists of a number of interrelated design activities which include reviewing the literature on avatar design in order to locate goals and variety of the project, therefore building on the on the work of others; a comparative review of three popular 3D AVWs to explore the design problem; a study that aims to gain an understanding of the social dynamics involved; the adaptation of a diagrammatic technique for the purpose of modelling social interaction; the development of 2D and

3D prototype techniques exploring the application of the social interaction modelling technique; a body of creative work developing ideas for conveying non verbal communication and the appraisal of the effectiveness of this creative work.

The research contributes to the field of avatar design in a number of ways. Firstly, it develops our understanding of social dynamics in virtual worlds. Secondly, it postulates modes of non verbal communication for both individuals and social groups that supports multi-participatory social discourse. Additionally, a number of useful research techniques have been devised, such as a linear diagramming technique that can be used to represent the structure of conversation thereby facilitating the exploration and understanding of the dynamics of AVW social discourse.

The work is of interest to those working in the field of avatar and multi-user virtual world design. It may also be of interest to anyone thinking of using an avatar virtual world for the application of collaborative learning, collaborative games and conferencing.

LIST OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF CREATIVE WORKS

to my parents

ACKNOWLEDGMENTS

The reasons for undertaking this research are simple: a love of 3D computer generated worlds combined with a desire to contribute to knowledge.

My sincere thanks goes to:

my family for their unparalleled support and encouragement through all these years,

my girlfriend for standing by me and enduring a virtual partner during the final stages of my endeavour,

Colin Beardon, my supervisor, for inspiring me with the initial idea of doing a PhD and for providing me with so much invaluable advice and discussion,

Jeremy Diggle, my advisor, for providing crucial creative insight during the practical stages of the project,

and the University of Plymouth for financing my research.

AUTHORS DECLARATION

At no time during the registration for the degree of Doctor of philosophy has the author been registered for any other University Award.

This study was financed with the aid of a studentship from the University of Plymouth.

Relevant seminars and conferences were regularly attended at which work was often presented; two papers were prepared for publication.

Publications:

1. The Role of Obligation within Virtual Encounters, Proceedings of CVE 98, 17th-19th June, Manchester, UK, 1998 244

2. Designing a Non-Verbal Language for Expressive Avatars, Proceedings of CVE 00, 10th-12th September, San Francisco, USA, 2000 252

Presentations and Conferences Attended:

April 1997	Attended the Computers in Art & Design conference, Derby.
Nov. 1997	Presented at the VRML 2.0 Workshop hosted by Network Virtual Reality Centre for Art & Design, University of Plymouth, Exeter.
April 1998	Presented at the Centre for Visual Computing seminar, University of Plymouth, Exeter.
June 1998	Presented at the Collaborative Virtual Environments conference, Manchester.
Nov. 1998	Presented at the Art & Design PhD seminar, University of Plymouth, Exeter.
Nov. 1998	Attended the Avatar's 98 conference, in cyberspace.
Dec. 1998	Attended the Commedia Dell'Arte seminar, University of Plymouth, Exmouth.
Jan. 1999	Attended the Visualization and Virtual Environments Community Club seminar, Rutherford Appleton Laboratory.
Feb. 1999	Presented at the University of Wales College, Newport.
March 1999	Attended the Cross Currents conference, Phoenix Art Centre and the University of Plymouth, Exeter.

March 1999	Presented at the PhD research day, University of Plymouth, Exeter.
April 1999	Presented at the Computers in Art & Design conference, Middlesborough.
April 1999	Attended the Virtual Humans workshop, Digital Media's Futures Conference, Bradford University.
Nov. 1999	Presented at the PhD Research day, University of Plymouth, Exeter.
May 2000	Attended the Performing Virtualities Workshop, Cumberland Lodge.
Sept. 2000	Presented at the Collaborative Virtual Environments conference, San Francisco.
July 2001	Presented at the Mixed Reality Laboratory, Nottingham University.

Signed ...N Beale.....

Date .....20th Feb 2002

PART 1: BACKGROUND

Chapter 1:

Introduction

Chapter Overview

This chapter introduces the research project. It states the original rationale for it as well as the author's interests in avatar design. Next, the initial research questions are defined and the author declares an intent to address them through a design project. Following this, the general research strategy is presented, the relationship to the attached CD ROM is described and an overview of subsequent chapters is given.

1.00 Rationale

The reasons for undertaking this research are simple: love of 3D computer generated worlds combined with a desire to contribute to knowledge. Within recent years a number of avatar virtual worlds (AVWs) have emerged on the Internet. These are computer generated, multi-user, graphical spaces within which people meet, form social groups and interact with each other in real time, typically through the exchange of text or audio messages. Each user is represented within the space by a digital image known as an *avatar*, which is usually humanoid in form and is predominantly under the control of the person it represents. Already people are using AVWs to socialise, work, train, learn, conference and play games together, regardless of their geographical location. The author's interests have naturally extended to encompass this new medium, which has become the inspiration for this research project and it is the vehicle through which, it is hoped, new knowledge will be found.

1.10 Outset

AVWs have been designed as an expansion of the older text based multi-user systems known as MUDs (Multi-User Dungeons/Dimensions). The main development is the addition of a graphical dimension which makes possible non verbal forms of communication such as body language. In the real world of face to face communication, body language is used to transmit, often unspoken, indicators of social behaviour, for example the use of gaze to indicate whose turn it is to speak. When participating within an AVW session, one expects to see such visual communication both facilitate and complement social discourse, just as it does in the real world. However, as I will illustrate later in the thesis, this currently is not the case (Chapter 3). Communication is almost entirely limited to the exchange of voice, or more typically, text messages which are usually observable in a separate dialogue window on the screen. Despite being equipped with three or four poorly chosen gestures, in actuality the avatars do little more than represent a user's presence, location and direction of face within a novel visual representation. If the graphical dimension of AVWs is to become useful within the context of social communication, we must begin to address the poor design of visual channels such as body language. The research described within this thesis is therefore driven by the following question:

- How can multi-modal social communication that exploits the rich potential of body language, be achieved in 3D AVWs.

To address this important question, the author will design a 3D humanoid avatar and body language for it within the context of social interaction, that can be integrated with verbal discourse in a meaningful way. The avatar will be specifically designed for an AVW that would be used for socialising. This seems to be a good place to start because

the social dynamics in this type of world are the most apparent, therefore the requirements of a body language are the most obvious. In addition to this, the final solution will strive to be compatible with (but not led by) existing and near future technology. It is this author's belief that effective solutions can be reached without the need for technical advances because the problem to be addressed is due to inadequacies in the design, and not technical limitations.

1.20 General Research Method

The research question discussed in the previous section is addressed through the application of a creative design process. Figure 1-0 shows a simplified diagrammatic representation of a creative process. The reader should note that the model is consistent with the basic structure of many put forward and debated over the last fifty years by design methodologists and philosophers [Hertz 92, Lawson 90, Rowe 87].

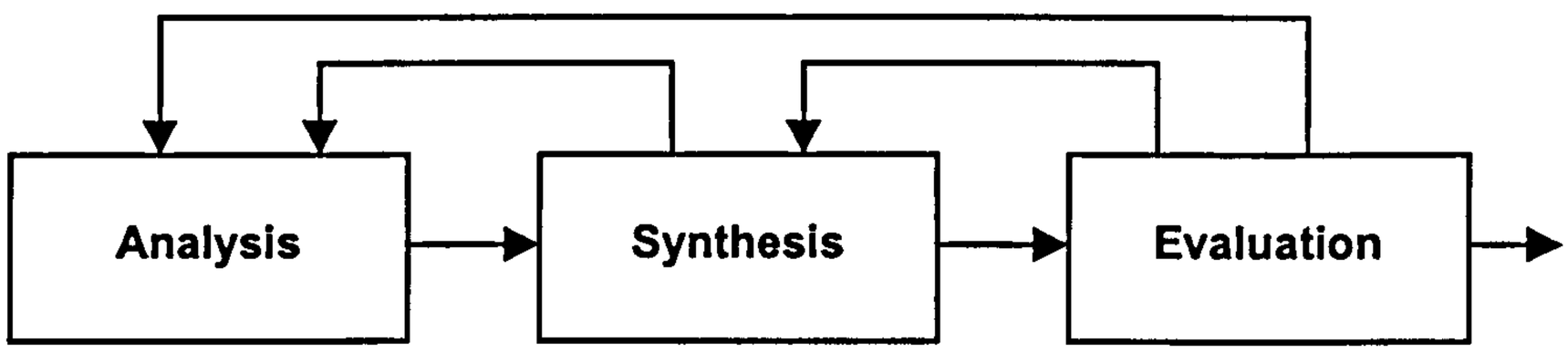


Figure 1-0: Simplified Diagrammatic Representation of the Creative Process.

The creative process can be understood as three sequential phases named *analysis*, *synthesis* and *evaluation*, which are linked in an iterative circle. Once the design problem has been identified, it is ordered and structured in the *analysis* phase. From this, a design brief is drawn up that defines the objectives of the project. In addition to this, formal background research is conducted that takes the form of a theoretical investigation. The aim of this is to better understand the problem and to increase the amount of information available. The second phase, *synthesis*, involves moving on to

create a response to the problem. Creative practice is used to develop ideas through practical prototypes. In effect, the final solution is found by trying a number of possibilities and developing them. The third phase, *evaluation*, is an appraisal of the solutions against the objectives set out within the brief. Even the simplest models however, need to allow for an indefinite number of loops back to any of the preceding phases. These are represented on the diagram by backwards arrows.

Within the design project described by the thesis, *analysis* involves the following separate activities: background reading into avatar design in order to understand the relationship of the project to the field and learn about the work of others; a comparative review of existing AVWs to achieve a better understanding of both the technology and the challenges to be addressed: empirical research which aims to gain an understanding of the social dynamics of virtual worlds; and an attempt to model social interaction in order to arrive at a set of useful behaviours. The *Synthesis* phase involves the creation of a number of experimental practical pieces of work that explore conceptual ideas. These mainly take the form of digital movies that simulate an AVW and through which the ideas are first implemented for evaluation purposes, and interactive movies that allow ideas to be tested from a user perspective. Often, it is the process of creating a piece that facilitates the generation of new ideas and enables the concept to be taken forward. *Evaluation*, is ongoing throughout the process, and is particularly active at the *synthesis* phase. Each piece of creative work is evaluated as it is being constructed against the specific objectives that it is trying to address, and again when it is complete, against those of the larger design activity. On completion of the creative work, a more rigorous appraisal takes place against the objectives set out within the design brief.

1.30 About the Accompanying CD ROM

Within subsequent chapters the reader will be referred to creative pieces of digital work, which are located on the accompanying CD ROM. It is important that the reader views these works in order to follow the text and understand the research. Some works are digital movies while others are interactive and require the reader's participation. Relevant instructions for viewing a piece are given within the text where reference is made to it.

Many of the files use *Apple's QuickTime 5.0.* If this is not installed then it can be download free of charge from *Apple's QuickTime* website [1]. Also, in order to ensure reasonable performance, it is strongly recommended that the directory *Earle_PhD* be copied from the CD to the computer hard drive. Eject the CD ROM and run the file *Browser.exe* located within the new *Earle_PhD* directory. If it is not possible to copy the directory to the computer then the files will run from the CD, however, there may be performance problems, particularly with *Demonstration 9: Final Piece.* A PC computer with the following minimum computer specification is required:

- Operation System: Microsoft Windows 95/98/2000/NT
- Processor: 350 MHz
- RAM: 64 MB
- Free Hard Drive Space: 600 MB (for copying)
- CD ROM Drive
- Sound enabled

1.40 Thesis Structure

The thesis consists of four parts which roughly correlate to the various phases of the creative process used in the design project. Part One, *Introduction*, locates the research project within avatar design and an element of analysis of the problem takes place. Part

Two, *Background Research*, is purely analysis of the design problem in the form of two studies conducted as background research. Part Three, *Addressing The Problem*, describes creative work in the form of a series of experimental pieces. Part Four, *Evaluation & Conclusions,* is an appraisal of the solutions against the brief, as well as against the larger research activity which is presented in the form of a set of conclusions. There follows an overview of the subsequent chapters.

PART 1: BACKGROUND

Chapter 2: Context & Related Research

The field of avatar design is defined and the research project located within it. This is achieved by giving an overview of the history of 3D AVWs, presenting some emerging applications of the technology and by examining related research across a number of disciplines which the field touches upon.

Chapter 3: Existing 3D Avatar Virtual Worlds

In this chapter a comparative review of the three most popular 3D AVWs is presented. This enables a better understanding of the design problem to be addressed and provides the reader with an overview of the state of the art in 3D AVW systems.

---

[1] Download *QuickTime 5.0* at http://www.apple.com/quicktime/download/

Chapter 4: Design Brief & Considerations

The design brief is described which defines the objectives of the larger design activity, and divides the task into two smaller and interrelated endeavours which are integrated later in the project. A number of design considerations are also discussed which it was necessary to address during the early stages of the design process.

PART 2: PRELIMINARY STUDIES

Chapter 5: Understanding The Social Dynamics of Virtual Worlds

A theoretical study is presented that aims to develop an understanding of the social dynamics of virtual worlds. Two dialogues generated from MUD sessions are subjected to discourse analysis in order to identify patterns of social behaviour, and to determine modes of communication that a virtual body language can complement.

Chapter 6: Modelling Social Interaction

This study attempts to build a model of social interaction in order to reduce a seemingly infinite number of possible participant actions and reaction during social discourse, to a finite and manageable set that are likely to be useful within the context of AVW social communication. The work includes the development of 3D pieces of work and a 2D rapid prototype technique which are used to explore the model in practice.

PART 3: CREATIVE PRACTICE

Chapter 7: Integrating Channels Of Communication

Practical work is described that addresses the first task of the design brief, which aims to integrate the various channels of communication, i.e. text and image. This work, almost unexpectedly, leads to the design of a non verbal communication of social groups.

Chapter 8: Designing An Avatar Body Language

Practical work is described that addresses the second task of the design brief. The aim is to develop a humanoid avatar with the necessary operators to carry meaning, and a body language for it that consists of a limited set of gestures which support multi-participatory social discourse.

PART 4: EVALUATION & CONCLUSIONS

Chapter 9: Evaluation

This chapter presents an appraisal of the solutions against the objectives set out within the design brief. To assist the evaluation, two creative pieces of work are used. The first aims to gain an understanding of how well the set of gestures developed convey intended meaning. The second, aims to determine how effectively the solutions of the various design tasks arrived at in the previous part, work together when integrated.

Chapter 10: Conclusions

This chapter concludes the thesis by summarising the important contributions to knowledge, discussing the project key outcomes and suggesting future research projects to carry the work forward.

1.50 Summary

Within this chapter the initial rationale for the research project has been stated, and the author's interest in 3D AVWs declared. AVWs have been defined as online, computer generated, multi-participatory, 3D graphical spaces where users, who are represented by a digital image predominantly under their control known as an *avatar*, can meet and interact with each in real time typically through the use of text or audio messages.

The observation has been made that if the graphical dimension of AVWs is to become useful within the context of group social communication, we must begin to address the poor design of the visual channels such as body language. The initial research question is therefore: how can multi-modal social communication be achieved in 3D AVWs that exploits the rich potential of the visual channel? To address this important question, the author will design a 3D humanoid avatar, and a body language for it that supports group social interaction, and that can be integrated with verbal discourse in a meaningful way. The solutions will also aim to be feasible within the confines of, but not led by, existing technology.

The general research strategy is to apply a creative design process which the author believes is an effective approach for developing intelligent and appealing visual languages. This process consists of three sequential phases named *analysis*, *synthesis* and *evaluation*, which are linked in an iterative circle with an indefinite number of

backwards loops to each previous phase. This chapter has also described the relationship of the accompanying CD ROM to the thesis to which the reader will be referred to view practical work, and has provided an overview of subsequent chapters.

Chapter 2:

Context & Related Research

Chapter Overview

The purpose of this chapter is to provide an overview of the technology and related research. While the research itself is rooted within Art & Design, it is necessary to extend the review to include a range of studies from other disciplines, in particular, the computer and social sciences. It is not the author's intention to address all of the challenges discussed within the chapter, but to describe the emerging field of avatar design and locate the design project within it. The chapter begins with an overview of the history of 3D AVWs (avatar virtual worlds) and presents some emerging applications of the technology. Next, related research is discussed which is organised under the following headings: Realistic Virtual Humans, Believable Agents, The Design of Shared 3D Worlds, Personal Representation, Non-Verbal Communication and Social Communication in Avatar Virtual Worlds. The latter being the context of the design project.

2.00 Background of Avatar Virtual Worlds

The concept of a networked computer-generated space where people interact with each other and with artificial beings has naturally attracted interest from the realms of science fiction. William Gibson's novel *Neuromancer* [Gibson 84] describes a future scenario where people are able to jack into a global network via neural implants. Once within the network people interact with each other, objects, software and artificial intelligence programs. This novel inspired many minds and has no doubt, been an influential factor in the emergence of AVWs.

Several years later Neil Stephenson published his novel *Snowcrash* [Stephenson 92]. In *Snowcrash* a similar scenario is described where people log into a shared space known

as the *Metaverse* to conduct their everyday business. The virtual landscape resembles the real world but provides a much cleaner and safer environment to work in. People are represented by realistic 3D graphical bodies known as *avatars* which, in terms of identity and communication, function in much the same ways as our bodies do in real life. This novel is widely considered to be the origin of the term *avatar* in its contemporary usage.

We are of course, a long way from the imaginations of such science fiction authors as Gibson and Stephenson. However, publicly accessible networked virtual environments where people are represented by avatars and can interact with artificial characters are already a reality. This has been made possible by the development of online multi-user text-based systems originally used for role-playing games similar to dungeons and dragons, 3D rendering engines adapted from arcade games, and virtual reality modelling technology.

The first online multi-user virtual environments were text-based and emerged in the 1970s. These environments known as MUDs (Multi-User Dungeons/Dimensions), allow people to interact with each other in real time typically through the exchange of text messages. Early MUDs were generally used for role playing adventure gaming. Later MUDs like *The Well* were used for conferencing (see Hafner 97), however the majority of newer ones are chat-orientated and are commonly used within e-commerce to create communities of people around commercial websites. An example of this is the *Microsoft Network* [2] where visitors can shop, search the web, sign up for a free email account and meet like minded people through chat-based MUDs.

[2] *Microsoft Network* is at http://www.msn.com

In 1985 Lucasfilm released the first graphical online multi-user environment known as *Habitat*. While the basic principle of operation was the same as the text only environments, 2½D imagery was used to represent both the environment and the people within it (2½D refers to the use of 2D graphical objects that can overlap each other). Participants were able to move their avatars about the environment and communication was achieved through the exchange of text messages. *Habitat* was essentially an extension of the MUD systems. For more information on *Habitat* see [Morningstar & Farmer 92].

Ten years on, a 3D Virtual Reality Modelling Language (VRML) for the internet emerged and was quickly accepted as the web standard. A small company in San Francisco known as *Worldschat*, used VRML to create a 3D space station where visitors could meet as avatars and interact with each other in real time [Damer 96]. The physical laws of the environment, such as gravity, correlated to those of the real world, with certain flexibility. The avatars themselves were simplistic, static images that indicated a participant's presence, location, direction of face and said something about the user's identity through the image that they selected for themselves. Since that time several other publicly accessible worlds have emerged such as *Active Worlds* and *Blaxxun*, and the avatars in them are 3D forms rather than static images, and most can be animated to some degree.

Today, numerous virtual environments of all three types (i.e. text only, 2D and 3D) are available on the Internet. These environments are accessible to anyone with a desktop PC, standard modem and Internet connection. New styles of text-based systems have emerged which include IRC (Internet Relay Chat), ICQ by *Mirabilis* (a desktop text-based paging system) and MOOs (MUDs with an object oriented programming

language used for navigation, creating places and objects with behaviours). There are several popular layered 2D online systems such as *The Palace* and *Worlds Away*. There are about half a dozen 3D virtual worlds systems which include *Active Worlds*, *Blaxxun*, *OZ Virtual*, *Online Traveller*. Generally the 2D and 3D systems host a number of sub worlds that are based on different themes. Public popularity of each type of virtual environment continues to grow as does interest within research communities. For further information on these systems please visit their respective web sites [3]. Links to all current publicly accessible graphical avatar virtual worlds can be found at the *Contact Consortium* web site.

2.10 Emerging Applications

While avatar virtual worlds are still a relative new technology and are used significantly for socialising, there are already a number of useful applications and others are being investigated. The most common application next to chat, is multi-player computer games such as *Quake 3* [id 99]. As well as the standard game play, *Quake 3* features a multi player mode were the user is able to either compete against, or team up with, other players over the Internet. As Steve Benford has pointed out, an intriguing aspect of many of these games is that although they are highly competitive, they can also encourage co-operation and team work [Benford *et al.* 97].

---

[3] The IRC web site is at http://www.irc.org
The ICQ web site is at http://www.mirabilis.com
The Palace web site is at http://www.thepalace.com
Worlds Away web site is at http://www.worldsaway.com
Active Worlds web site is at http://www.activeworlds.com
Blaxxun web site is at http://www.blaxxun.com
OZ Virtual web is at http://www.oz-inc.com/
Onlive Traveller web site is at http://traveler.onlive.com
Contact Consortium web site is at http://www.ccon.org

An important application of virtual environments is collaborative learning. This has been studied by researchers such as Amy Bruckman who has shown that the social implications can be used as a driving force to encourage people to learn [Bruckman 94]. Also, Howard Rheingold has discussed the power of play for encouraging learning in virtual worlds within his book *The Virtual Community* [Rheingold 94]. In 1996 the Contact Consortium initiated a project to develop the first virtual university known as *TheU* [4], hosted by *Active Worlds*. The aim was to create an international university where students and professors could interact, and bring together their cultural differences for the purpose of connective intelligence. The subjects available revolve around multi-media, the idea being to teach multi-media using multi-media. Other related projects include, Randolph Jackson and Eileen Fagan's study of how immersive virtual reality can be integrated into existing school curricula [Jackson & Fagan 00]; Laurie McCarthy's project to create a software agent that assists the training of students within an immersive multi-user virtual environment by instructing, guiding, monitoring and evaluating their progress [McCarthy *et al.* 98]; also, a number of links to case studies concerned with the potential and challenges of 3D avatar worlds as learning environments can be found on the *V-learn* web site [5].

There are a number of projects that are investigating the potential of AVWs for dramatic activities and interactive story telling. Debate is taking place about how avatar worlds can be used to create a new form of theatre. For further information on this the reader is referred to an online newsgroup that focuses on the topic [6]. Mel Slater has experimented with a 3D virtual stage for rehearsing live performances that are performed in the real world [Slater *et al.* 00]. Disney Imagineering has experimented with using virtual reality for story telling. In this project guests fly on a magic carpet

---

[4] *TheU* virtual university is at http://www.ccon.org/theu/index.html
[5] Case studies on *V-learn* web site at http://www.vlearn3d.org/library/case_studies.html.

around a virtual world based on Disney's animated movie *Aladdin* [Pausch *et al.* 96]. Isabel Machado and colleagues have developed *Teatrix*, a 3D multi-user world that combines drama activities and story telling to assist the development of children [Machado *et al.* 00]. Lastly, the *OZ project* at the Carnegie Mellon University is focusing on bringing interactive drama to multi-user worlds that feature intelligent software agents, in order to create a new kind of cultural experience that is engaging and compelling. For further information on the OZ project see [Mateas 97].

A field of study has emerged which is concerned with the development of multi-user virtual environments to support collaborative work activities. There are two main conferences that are facilitating deep discussion and research within the field. One of these is CSCW (computer supported collaborative work), which deals with a range of issues including culture and community, distant collaboration, mobile work and technology, awareness and attention, video technologies, collaboration infrastructures, privacy, and new interaction paradigms. For more information see the CSCW conference web site [7]. Closely related is the CVE conference (collaborative virtual environments), which explores similar themes and includes models and metaphors for collaboration, experiences and evaluation of collaborative environments and applications, systems architectures, potential applications, social action and interaction. For information on CVE visit the conference web site [8]. An important point to note is that, rather than using anonymity, users within CSCW applications present their *true* self, i.e. a user will use their real name, their avatar will look like them and they will act in the same manner as they usually would within work environment.

---

[6] The VW-Theatre online newsgroup is at http://www.ccon.org/lists/vw-theater.html
[7] The CSCW conference web site is at http://www.acm.org/sigs/sigchi/cscw2000/
[8] The CVE conference web site is at http://www.ai.sri.com/cve2000/.

In 1998 the *Avatars* conference became the first major convention to be held in an avatar virtual world (AVW). An impressive 3D conference suite was created where people gathered to discussed contemporary issues in specially designed podiums. For more information on the *Avatars* conferences refer to the *Wired* News Article [Brown 98] or the conference's web sites [9]. Already other conferences have followed suit such as *V-Learn 3D 2000* [10]. For an insightful paper on the challenges and potential benefits of virtual conferencing see [Jones 00].

Many researchers are anticipating that in the near future avatars will play an active role in online retail. Both the *Avatars 98* and *99* featured trade stalls much like those found at some conventional events. Sale representatives appeared as avatars at their stalls to talk with existing and potential customers. Some researchers such as Barbara Hayes-Roth believe that avatar sales representatives of the future will be controlled by artificial intelligence software that interacts with, and serves customers while exploiting multi-modal communication [Hayes-Roth *et al.* 99]. The potential advantages of this are that they will bring interest and warmth to the online experience through distinctive personalities, and build relationships with customers based on familiarity and trust. They will exploit the web's vast information resources and operate twenty-four hours a day, seven days a week. They can scale in numbers to offer one-to-one service and remember important features making customers feel valued.

2.20 Realistic Virtual Humans

The applications discussed previously demand advances in a number of related areas of research. One of these is the modelling of realistic virtual humans whose appearance

---

[9] Avatars Conference 1998, 1999, 2000 websites are at http://www.ccon.org/conf98, http://www.ccon.org/conf99, http://www.ccon.org/conf00.
[10] *V-Learn 3D 2000* conference web site is at http://www.vlearn3d.org/conference/

and behaviour needs to be a convincing representation of real people. The main applications of realistic virtual humans range from movie industry special effects, to computer games and ergonomic testing. Research in the field breaks down into a number of sub-problems. The two main ones are physical modelling to create a credible visual appearances, and the modelling of human behaviour from personal mannerisms to large scale strategic planning.

In the work of Nadia Magnenat-Thalmann we can see a tension between realistic physical modelling and the quality of the visual outcome. Her concerns are mainly the latter and she achieves this through modelling techniques that do not necessarily correspond to the way the human body is constructed. One example of this is the making of a pilot to test the feasibility of creating a ninety minute film about the recently discovered Xian Terra-Cotta Soldier's, who come to life within the animation. Rather than create a skeletal structure that supports muscle, fat and flesh, all enclosed by a skin with elastic properties, the soldiers' bodies are created by using and adapting a technique known as *metaballs* that emerged in the late 1980's. These are ellipsoid forms that are attached to each other to create rough impressions of each soldier's body. Details are achieved by adding, editing, transforming and adjusting various parameters [Magnenat-Thalmann *et al.* 95]. This technique has recently become widely popular for giving shape to not only human, but also animal and other organic life forms which are difficult to model using traditional geometric techniques. Another good example of her work is a virtual incarnation of Marilyn Monroe. A recent piece of animation that shows Marilyn receiving a *Golden Camera Award* in Berlin, demonstrates the success of her efforts to model realistic hair, skin and clothing through similar techniques of simplified physical simulation [Magnenat-Thalmann *et al.* 96]. These efforts have also

allowed her to create a virtual television presenter [Magnenat-Thalmann & Kalra 95] and virtual fashion models that look like real ones [Volino & Magnenat-Thalmann 99].

The work of Daniel Thalmann has a slightly different orientation. Apart from credible modelling he is concerned with the way in which a virtual human could exist within a changing world. In one example he attempts to make a virtual human play tennis with a real person, in the sense that it sees the ball coming, has to analyse its trajectory and calculate an appropriate return shot [Kalra *et al.* 99, Noser *et al.* 96]. In this example the virtual human does not have perfect knowledge of the world and must function on the partial information that it can glean from the situation. Also it must act in a world in which there are other autonomous virtual humans. In a project to create a virtual public garden containing numerous synthetic actors that are out walking and attempting to engage each other in conversation, each actor is provided with a character and description of its interpersonal relationship with the others. The actors interact through body posture which indicate their level of happiness. An actor interprets the posture of a conversation partner, and responds in a manner that is influenced by their own character profile [Becheiraz & Thalmann 96].

Rather than strive for realistic hair and skin, or convincing behaviour and individual mannerisms, Norman Badler is concerned with creating models that are bio-mechanically correct for simulation purposes and human factors analysis. He has created a software system, *Jack*, which is embedded with an understanding of physical limitations of the body such as reach, fit, visual perception, fatigue and strength. This enables Jack to be set goal directed tasks based on a set of objectives and rules [Badler *et al.* 93]. An example of an application for this kind of virtual human is to evaluate how easily a person of a specified size and weight can manoeuvre in and out of a virtual

space shuttle doorway. The resultant simulation allows designs to be evaluated and altered if necessary, potentially saving on production costs. One new and particularly relevant direction of the Jack software, although still in its prototype stage, is an online AVW that is used for education and training called *JackMOO*. In particular Badler hopes to use the system as a test bed for verbally controlled avatar animation e.g. *step forward, turn around, look at* and so on [Badler 99]. This research is however still in its early stage and its outcomes remain to be seen.

At present only limited knowledge from the field of realistic virtual humans is transferable to the design of avatars for online multi-user worlds. No doubt in the distant future this situation will change, however for the time being there are two major difficulties with the approaches discussed. The first is that they generally use high-end technology which exceeds the capabilities of the limited bandwidth and processing power of the Internet and desktop PCs of today, and will continue to do so for the foreseeable future. The second difficulty is psychological and repeatedly surfaces as an underlying theme in literature about virtual reality. When a photo-realistic virtual scene is encountered, it induces an expectancy for everything to appear, move and behave in a consistent manner. As soon as something does not, the participants' illusion of reality is broken (for examples of such literature see Badler *et al.* 93, Benford *et al.* 97, Slater *et al.* 00). It is reasonable to argue that until avatar technology is capable of meeting our expectations for realism, it is advisable not to attempt to design avatars intended to pass for real humans in terms of appearance and movement, particularly when low end technology, such as desktop PCs, is being employed.

2.30 Believable Agents

In the early 1990s there was an explosion of research into intelligent agents. These are computer programs that employ artificial intelligence techniques to provide active assistance to a user with computer-based tasks. Agents become competent by learning from the user(s) as well as other agents. They are useful as they help us to manage the overwhelming amount of digital information, particularly within a business environment. Successful prototypes have already been created that provide personalised assistance with meeting scheduling, electronic mailing handling, electronic news filtering and selection of entertainment [Maes 94]. For further information on intelligent agents see [Selker 94, Wooldridge & Jennings 95, Riecken 94 and Norman 94]. A sub field has emerged from this where researchers are embodying agents within virtual worlds and developing believable characters for them. The appeal of this is that it humanises the agent, thus making it easier for participants to accept and more natural communication can be achieved through the addition of non verbal forms of conversation.

Many researchers are anticipating that such agents will soon be common place in AVWs and will play two main roles. Firstly, they will be able to stand in for a participant and take action on their behalf whenever the person finds themselves unable to do so personally due to absence, overloading, distraction, etc. Secondly, they will coexist with human participants and play key roles in learning and interactive drama applications [Mateas 97, Pausch *et al.* 96]. The point to note is that an agent's avatar will look and operate in an almost identical manner to one that belongs to a human participant. The only difference is that the agent's avatar it is controlled by software that takes action without user intervention and can operate concurrently.

The field itself shares some challenges with that of modelling realistic virtual humans, in particular, the goal of creating the illusion of life. It differs however, in that the approach taken to achieve this does not necessarily depend on the use of the human form or realistic imagery. In fact, many researchers in this field opt for a cartoon like aesthetic and focus on creating believable characters. This is evident in the work of Joseph Bates who leads the *OZ* project. His approach to achieving this goal is to embed agents with human like emotions. He advocates that an emotionless character is lifeless and little more than a machine (based on a key principle used in Disney animation, that it is the portrayal of emotion that gives animated characters the illusion of life [Thomas & Johnston 81]). In one experiment he creates a simulated world containing several agents that are represented as self animating creatures called Woggles. These are simplistic constructions consisting of a coloured sphere with two eyes. Each Woggle is given a goal-directed task within the world, thus creating certain events to which they will have emotional reactions. For example, anger when a goal failure has occurred and one Woggle judges that this has been caused by another [Bates 94]. The experiment is successful at creating a strong sense of life without realistic imagery, and without other behaviour that we consider central to human intelligence such as reasoning and problem solving.

To Barbara Hayes-Roth, creating the illusion of life is just as important as the agent's ability to do its job. She has led a number of projects to develop interactive agent characters to staff the web and play helpful roles in Internet enterprise web sites. One example of these is *Merlin*, a 2D graphical character who guides interactive neighbourhood tours at Haight Street on the web. Through natural language interaction, he greets each customer, determines the details of the right tour, acts as an escort and offers value-added commentary [Hayes-Roth *et al.* 99]. *Merlin* is a more complex

construction than a Woggle, and a strong sense of character is achieved not only through the portrayal of emotion, but a range of mechanisms such as appearance, knowledge, opinions, social dynamics, and in particular manner of moving, gesturing and speaking. This work is a good example of how believable characters do not have to impersonate real humans, rather, they encourage the user to willingly suspend their disbelief through interactivity with interesting and warm personalities.

The *Persona Project* team at Microsoft have created an agent character named *PIDI* (Peedy) who provides access to a music data base. Over time *PIDI* learns what sort of music the user likes and so is able to make better suggestions. In order to enhance the believability of the character the group have experimented with a simulated episodic memory. Elements from earlier dialogue and past conversations are stored, then used to effect *PIDI's* social behaviour and make future dialogue more interesting [Ball *et al.* 97]. The notion of an autonomous memory for avatars that represent human participants is one that will need to be addressed at some point in terms of NVC (non verbal communication) for avatars. Memories of past encounters should automatically effect the visual properties of visual communication. For example, if on the last encounter two participants argued, then their body language and mannerisms are likely to be quite different on their next encounter.

One other relevant focus of this field is the study of how we socially interact with these believable characters. Much of this work is rooted in the development of system architecture, for example Bruce Blumberg's work to develop autonomous creatures for virtual worlds that are controlled by multi-level directional instructions [Blumberg & Galyean 95]. There have also been some useful user-based studies. For example, a study lead by Salvatore Parise which involved conducting a series of experiments to test

how co-operation with a computer agent was effected by pictorial realism, human likeness and likeability. In one experiment human participants played a game designed to produce data to assess their level of co-operation with the agent. Each subject played each of the four agent partners with which they interacted through a monitor, keyboard and audio link. One of these partners was a real human using a video link (although this was unknown to the subjects), the other three were agents in the form of a human face, a photo-realistic dog and a computer generated dog. The resultant data suggests that the participant's social behaviour is effected by the agent's appearance [Parise *et al.* 96]. This leads to intriguing questions, which have yet to be addressed, about how the graphical appearance and mannerisms of a participant's avatar will effect the way other people will react to them within multi-user virtual worlds.

2.40 The Design of Shared 3D Worlds

The design of avatars that are more sociable will demand social spaces for them to exist within. There has been considerable research over the years into the design of virtual environments for a range of purposes such as Tetsuya Fujii's integrated system for interactive planning and evaluation of a city's development plans [Fujii *et al.* 95], Christine and Adrian Clark's 3D worlds for interfaces to information systems [Clark & Clark 97], William Mitchell's project to create the pyramids of Kahun, Egypt, as part of a virtual museum [Mitchell 97] and gaming environments such as Sony's *Tomb Raider* available for the Play Station [Sawyer 97]. Despite this, the design of graphical shared social spaces that support online communities and work groups is still a relatively under researched area. This is somewhat surprising considering that widespread acknowledgment of virtual worlds as highly social spaces occurred a decade ago as a result of two well known papers. Pavel Curtis's report on the social phenomena that took place at LambdaMOO which led to new modes of behaviour not seen before

[Curtis 92] and, Michael Rosenberg's report on the social culture that developed at WolfMOO and its relationship to real life [Rosenberg 92]. One of the first efforts to address the poor design of shared worlds was a report released by the Centre for Electronics Arts at Middlesex University, that examined a range of issues [Boyd-Davis *et al.* 96]. A relevant conclusion made was that the design of virtual environments should be a creative rather than technical design process.

Since this report researchers such as Avon Huxor have been applying useful knowledge about the respect for social relations within architecture, to the design of virtual spaces that support CSCW (Computer Supported Collaborative Work). One of his key interests is to create virtual spaces that facilitate chance encounters, which have been said to account for much of the information flow in organisations in the real world. He believes that this can be achieved using two main principles which are *temporal* i.e. by designing spaces that people are drawn to more then they need to be, and *spatial* i.e. by the arrangement of specific task and meeting areas, known as locales, to bring appropriate people together when they pass from place to place. Huxor focuses on the latter in a project to provide support for part time students who can only meet in real life one day a week, through creating a space in Active Worlds (3D system) that supports their collaborative research. The space hosts design features such privacy areas, specific locales for each project that are arranged to create opportunities for chance encounters, and transparent walls to increase the awareness of others [Huxor 98].

Huxor's first principle, to create spaces that encourage people to spend more time within AVWs, is being addressed by a number of other researchers in the sense that they are attempting to create worlds that are more appealing to users. Ben Anderson and Andrew McGrath focus on this in their work with mutable virtual environments that

change with and without user intervention, just as they do in the real world. Their goal is to create worlds that users wish to return to even if there is nobody to talk to. This is important as in most systems, such as *Active Worlds*, users are aware how many people are in a world before they enter. If there is nobody there then they are unlikely to enter at all. In one experiment they create a 3D information garden that features real time growth of search trees, based on text searches on Internet sites. Using the metaphor of information plants the related sources are represented as stems and leaves which grow over a period of time. There is also an element of change and decay, and visitors can actually take cuttings to create new trees which encourages them to return and tender the garden [Anderson & McGrath 97]. An extremely valuable aspect of this work is that contexts need to be provided for conversations, and providing mutable objects is one way of doing this. For example, a conversation topic can be initiated with reference to objects in the surrounding information garden e.g. "oh my!... hasn't that tree grown since last week".

Bruce Damer has experimented with an innovation for designing social spaces through a method known as *interaction design*. The *Worldschat* environment attracted a regular group of visitors who developed a culture of their own to suit the environment. This trend is prevalent amongst all other 3D worlds that have emerged with the exception of *Active Worlds* which allows communities to adapt the space for their own needs. Registered members can claim ownership of small areas on which they can build constructions with almost no technical skill by creating, copying, and manipulating basic forms such as floors, and walls, and roofs, then applying image texture maps to them such as pictures of brick or glass. The problem with this is that, although the environment is developed by community members, they are acting as individuals. It has therefore generally resulted in a mass of unpopulated private locales in the form of

houses that regular visitors have built for themselves but seldom return to. While this leads to ineffective world designs in terms of social interaction, the notion of a virtual space that an online community can adapt to meet its own social needs would seem to be a good one. Damer has shown that if an online community is coordinated and has cohesion through a few loosely defined objectives, then allowing them to sculpt the surrounding environment is an effective design approach for creating social places. In one example he led an open community of people to construct a virtual village in *Active Worlds* based on the theme of Sherwood Forest [Damer 96]. A large wall was erected to define the village within which the entire design process took place through experimentation and group meetings. Community members were even assigned roles such as map maker, town crier, administrator and architect. The result was a highly social space consisting of a main communal square, a school of building, a therapy clinic, sacred spaces and homes, which are arranged in ways which facilitate aspects of social interaction such as chance encounters. In addition to this, regular events and activities were held which encouraged people to keep returning to the village and contribute to the community.

2.50 Personal Representation

An important part of avatar design is the notion of personal representation. This has been addressed by computer scientists from a practical view point. Steve Benford has made important observations from his experience of both the *DIVE* environment at the Swedish Institute of Computer Science [Carlsson & Hagsand 93], and *MASSIVE* at Nottingham University. His observations led to many fundamental questions such as: if an avatar can take any form, how can we make it clear that an object is an avatar?, whether it represents a human participant or an agent?, and is the same person behind a specific avatar as it was during the previous session? [Benford *et al* 97]. Avatar

standard groups such as *The Humanoid Animation Group*, *Livings Worlds* and *Universal Avatars* have emerged to address these very questions. For further information on these groups and their specifications visit their respective web sites [11]. Computer science however, tends to approach the issues from a formal stance, therefore it is technical improvements that are usually proposed, rather than suggestions for developing the visual languages. It is believed that the questions mentioned would be best addressed through the design of an avatar's appearance in relation to the virtual world theme and the other objects within it. For example, in some worlds features such as eyes and hands would distinguish the object as being an avatar.

Both the *DIVE* and *MASSIVE* systems are laboratory-based and used for investigating the potential for CSCW applications. They therefore strive towards *true* representation of the users (see Section 2.10). The anonymous nature of the publicly accessible environments however, provides social scientists with an opportunity for studying issues of virtual identity, which many have embraced. One such researcher is Sherry Turkle who has investigated the power of MUDs for reconstructing one's own persona. She believes that the anonymous and highly social nature of MUDs facilitates a series of second chances for adults who work and rework unsolved personal issues. They provide users with the chance to express multiple and often unexplored aspects of the self, to play with their identity, abandon characters and to try out new ones. Through ethnographic research methodology she has studied many cases where users have created new identities and discusses them in relation to who they are, and the effect on them in real life [Turkle 95, Turkle 96, Turkle 97]. Turkle and other researchers such as Howard Rheingold and Amy Bruckman have examined the popular activity of posing as

---

[11] The *Humanoid Animation Group* web site is at http://www.ece.uwaterloo.ca/~h-anim/
The *Living Worlds* web site is at http://www.vrml.org/WorkingGroups/living-worlds/
The *Universal Avatars* web site is at http://www.chaco.com/community/avatar.html

the opposite sex during MUD sessions [Rheingold 94, Bruckman 93]. Barbara Becker has sought a better understanding of the process of self creation. Through her work she has shown how the identities people create for themselves are profoundly influenced by the culture, language and technology of MUDs [Becker 97].

Avatars bring a whole new graphical layer to self creation. This has been studied by John Suler who has observed that they are a vehicle for increased expression and can effect the resultant social behaviour of people one meets. Based on observations of the popular 2D world *The Palace* he has been able to draw up a rudimentary taxonomy of the different types of avatars people choose for themselves. For example, *cartoon* avatars which people identify with or admire, *evil* avatars that allow people to safely and even creatively express their dark side and *animal* avatars which symbolise certain traits such as strength, wisdom, loyalty etc. The *seductive female* avatar is used commonly by both men and women, usually because of the attention it attracts from others [Suler 97], therefore increasing opportunities for social interaction.

An artist who also deserves a mention is Victoria Vesna who has proposed ways of taking the role of avatars and agents as a representation of the self, to a more sophisticated level in her work with the *Information Personae.* This is a mixture of agent and avatar technologies that incorporates disembodied information on the Web and defines itself by the data, links and community that it is connected to. It autonomously searches for other data bodies belonging to those who have similar interests, and functions as an extension of the physical self. Community members are updated and notified of conversations, contacts, events and any other information requested [Vesna 98]. Within this work the avatar becomes more than an empty shell through which communication is achieved, it represents other information on the web

about its owner. It increases the user's awareness of the information and allows them to track who is accessing it and for what purpose.

2.60 Non-Verbal Communication

The purpose of this section is to gain an understanding of what is known about non verbal communication (NVC). It turns out that this is a well researched area, and a firm grounding can be achieved from a handful of good sources. In a most useful paper that examines the advantages and challenges of communicating though virtual reality, Kirsi Heiskanen and colleagues define NVC as being 'the process of signalling meanings through behaviour which does not involve the content of spoken or written words'. Heiskanen *et al* also explains how this behaviour breaks into the following components: appearance and artefacts, facial expressions, oculesics, kinesics, proxemics, paralanguage, haptics, chronemics and olfactics [Heiskanen *et al*. 98].

Appearance and artefacts refers to the general aesthetic of a person e.g. body shape and size, facial features, hair colour, jewellery, clothing, etc. This communication is important as it often forms a basis for judgements on first encounters. Within a virtual world this can be understood as the choice of avatar that a user chooses for themselves, and in some systems, the props that can be added to achieve personalisation such as a top hat.

Facial expressions are primarily used to convey emotion and are the most thoroughly researched category [Chovil 92]. They are able to convey emotion vividly where words can only do it abstractly. Paul Ekman and Wallace Friesen's research specifically investigates the role of facial expression in conveying emotion, and argues that there are only six main ones which are surprise, fear, disgust, anger, happiness and sadness.

These six can then be combined to create more complex emotions e.g. anger and fear. To convey these emotions the face uses certain operators that work in conjunction with each other i.e. forehead, eyebrows, eyelids, cheeks, nose, lips and chin [Ekman & Friesen 75]. Other researchers such as Janet Bavelas and Nicole Chovil have investigated the linguistic function of facial expression. In particular their role in punctuating discourse, for example, raising both eye brows indicates that an utterance is a question. Face expression is in fact the only form of NVC fast enough to keep up with speech [Bavelas & Chovil 97].

Oculesics refers to gaze, which is a powerful channel of NVC and is different to facial expression although it involves facial operators. Argyle has categorised the different functions of gaze as information seeking, signalling, controlling the synchronising of speech, mutual gaze and intimacy, inhibition of gaze [Argyle *et al.* 73]. By observing eye movements and changes of facial expressions, information about a conversation partner can be gathered. Gaze can be used to signal personal attitudes. For example, Argyle has observed that conversation partners may gaze more at each other the more they like each other, e.g. lovers. Gaze can also supplement verbal content by giving emphasis, comments and illustration and displaying the structure of what has been said. It can control the synchronisation of speech by helping to regulate it, e.g. indicating whose turn it is to speak by looking at them at the end of an utterance. Gaze is used as a cue for intimacy and it declines with proximity.

Kinesics is the communication of hand gestures and body movements. Within the thesis this term is extended to include *body postures* which are static, but relate to the category as additional meaning can be carried by the transition from one to another. Within NVC they generally indicate the state and mood of a persons, e.g. tired or alert. Hand

gestures are the most comprehensive NVC skill, they can be flexible and can compensate for the lack of involvement of the other bodily channels. They can be classified as being either symbolic, metaphorical or pointing [Salem & Earle 00]. Symbolic gestures are context independent expressions such as signs (e.g. sign language), the message is contained fully within the gesture performed. Metaphorical gesture are used in parallel to a spoken or typed message for quantification, expression of virtual and interface objects (e.g. a handle and geographical navigation), the message, in this case, is gathered from the gesture itself and the context within which it is performed. An example is using the hand to specify the intensity of the light, by moving it up and down. Pointing gestures are used to select objects or to reduce the scope of a spoken statement (e.g. "this", "that"). This is a special gesture, since its meaning is always context dependant. Pointing could be performed to indicate a direction of displacement, an object or an option selection.

Proxemics is the communicative aspects of the use of space. This includes the personal space around a body to which one attaches ownership, as well as meanings generated through the arrangement and orientation of bodies in space. Before the relevance of the notion of space ownership to virtual world NVC is described, it is necessary to seek evidence that users have the same perception of ownership of the immediate space that surrounds their virtual body as they do in the real world. Fortunately such evidence exists in the work of Barbara Becker and Gloria Mark who have investigated the social conventions within different types of virtual world. Although they report that conventions vary slightly between the various existing systems, their observations indicate that users do have a sense of personal space in AVWs. For example participants comment when others are standing too close, also distances are kept on first encounters and reduced as intimacy develops [Becker & Mark 98]. The latter behaviour

has also being reported by Anthony Guye-Vuillème during an evaluation of an implementation of a NVC of avatars which is discussed within the next section [Guye-Vuillème *et al.* 98].

The remaining categories are both less researched and less important to the design project, however it is necessary to define them. *Paralanguage* is behaviour that indicate the true meaning of a spoken message. Typical operators are voice tone, tempo, rhythm, articulation and pitch range. *Haptics* is tactile communication such as shaking hands. *Chronemics* means the use of time in communication for example, punctuality and age. Lastly, *Olfactics* refers to the communicative characteristics of scent. Pleasant smells can increase communication while unpleasant smells might reduce it. This is obviously a problematic form of NVC to replicate in virtual worlds as the various media available are currently limited to audio, text and image.

2.70 Social Communication in Avatar Virtual Worlds

This field of study is the context for the design project, however being a relatively new area there was little published literature available at the start of the work. Previously, some social scientists had commented on aspects of social communication within their work such as John Bowers, who reports problems with turn taking in the *MASSIVE* system [Bowers *et al.* 96]. However, only a few specific projects can be found which are discussed with the aim of understanding what has been achieved in this area, identifying aspects that can be built on and mistakes that can be learned from.

Hannes Vilhjálmsson shares this author's belief that the development of non verbal channels of communication in AVWs is crucial to their credibility. This is evident in his work with the *BodyChat* system, a prototype 3D virtual world that allows users to

communicate while avatar's gaze behaviour and face expression are automatically animated to complement attention, salutations, turn taking and back channel feedback. This is achieved by setting parameters through buttons in the interface, and employing a text recognition system that scans message for keywords which trigger gaze behaviour. For example, if a user sets their avatar to 'unavailable for conversation mode', then it will automatically avoid eye contact with an approaching participant. Another important feature of *BodyChat* is that text messages are automatically broken down into words which are displayed one at a time above the message author's avatar, which tries to accompany it with an appropriate face expression [Vilhjálmsson 97]. The project has a number of useful aspects that can be built on. Firstly it uses gaze to regulate conversation flow, indicate attention as well as the state of the participant, e.g. available or unavailable for conversation. Secondly it realises, and addresses, the need for a better method of interfacing with an avatar's visual behaviours and this is achieved using a degree of autonomy. Lastly, the design process uses an interdisciplinary approach drawing on media studies, computer science, discourse theory and context analysis. The project has three major short comings. Firstly, the system is only designed for one-to-one communication. Multi-participatory scenarios are not accommodated, which is really what AVW interaction is all about. Secondly, only the facial expression and gaze channels of NVC are exploited, therefore not utilising its full power. Lastly, although an interdisciplinary design approach is adopted, creative practice is not a component. This author believes that this has led to weaknesses in the visual languages, such as the poor use of camera angles.

Anthony Guye-Vuillème *et al* have attempted to develop a better NVC interface for AVWs within the input constraints of a desktop computer system. This is necessary as participants have various tasks to perform in addition to social communication with

others, such as navigation and interaction with virtual objects. Current interfaces only allow one of these activities to be controlled at any one time. Guye-Vuillème's solution is to create an on-screen control panel containing and categorising graphical buttons that trigger the available gesture, expression and body postures [Guye-Vuillème *et al.* 98]. In an implementation using the VLNET system, we can see some important links with the work described within this thesis (for information on the VLNET system see [Çapin *et al.* 97]). For example, Guye-Vuillème has developed a method of drawing up a limited palette of actions based on meeting a set of criteria. His criteria seek gestures and expressions that are commonly used, documented in scientific papers, different enough to compose, as culturally unbiased as possible, able to be performed in the standing position and where a graphical representation of the actions is available. Through an evaluation of this work by surveying a group of subjects who freely used the system, Guye-Vuillème has offered a number of useful observations. For example, the attentive pose was the most used posture as it was the only way the subjects could show that they were listening to each other. Because a first-person view had been used, a subject was not able to see his or her own avatar. This resulted in commenting on the need for confirmation of the behaviour of one's own avatar, and this suggesting that a third-person view may be better for AVWs. Shortcomings of this work are as follows. Firstly, there is an attempt at realistic modelling which I have already argued (see section 2.20) is problematic as avatar technology is a long way from been able to meet our demands for realism. Secondly, the interface solution relies upon manual activation of the gestures which the author believes is why existing gestures in existing systems are rarely used (see Section 3.20). Thirdly, the control panel takes up a considerable portion of the screen. Lastly, the palette has been developed to provide users with a means to be expressive, but it does not support groups of more than two participants, the

social conventions that exist or complement the social discourse that is specific of AVWs.

A team at Microsoft lead by David Kurlander have created a 2D virtual world that is radically different from any other. *Comic Chat* utilises the rich visual language of comics to represent the social communication. As dialogue is produced, static panels are automatically generated relying on the rules of comic strip composition. The system chooses which avatars (comic characters) to include in each panel, determines their orientation, decides on the gesture and expression for each character, constructs word balloons of multiple types, places them in proper reading order, chooses an appropriate zoom factor, decides when to begin a new panel and changes background and scene elements to reflect the topic of conversation. All of this is achieved through the detection of key symbols within text messages [Kurlander *et al.* 96]. *Comic Chat* begins to address group conversation by orienting avatars towards the intended message recipients. It also minimises the need for user interaction with the system through automation, therefore relieving the participant and allowing him or her to focus on conversation. One limitation of the *Comic Chat* system is that, although it features a graphically rich communication, the cartoon like aesthetic invokes a specific kind of dialogue therefore limiting it to chat and role playing applications.

The relationship of the design project to the work discussed within this section, is that it hopes to contribute to this field by developing a visual component of communication that addresses not only the individual, but also the needs of social groups in 3D AVWs. The NVC developed here aims not only to achieve avatars that are visually expressive, but are also more socially aware by supporting the social conventions of AVWs and by complementing multi-participant discourse. Rather than strive for photo-realistic

imagery and anatomically correct movement, the visual language will attempt a grammar that emulates natural paradigms sufficiently enough to effectively convey meaning, but does not limit the kinds of dialogue that can take place e.g. chat, collaboration, learning, etc.

2.80 Summary

This chapter has provided an overview of the history of 3D AVWs and explained how they have been developed as an extension of the earlier text based systems. A number of applications have been presented which include socialising, gaming, conferencing, collaborative learning, drama activities and CSCW. The emerging field of avatar design has been defined through an analysis of relevant work from other related fields. Through this analysis it has been possible to define quite precisely, the research goal of the current project.

This analysis shows that many researchers are attempting to design avatars that convey a strong sense of life. One approach is to aim for photo-realistic imagery and anatomically correct movement. Another is to employ a cartoon like aesthetic and to focus on creating believable characters through strong conveyance of emotion. Other researchers are interested in aspects of personal representation which range from how to distinguish an avatar from other objects in a world, to how it can serve as a costume which can help one explore ones own identity though experimentation with self creation. Another relevant area is the design of virtual worlds. While researchers and designers are attempting this for a number of purposes, the design of online 3D spaces inhabited by avatars for the purpose of multi-participatory and multi-modal social communication remains a relatively under-researched area. We have also seen how real world NVC can be broken down into a number of channels which are: appearance and

artefacts, face expressions, gaze behaviour, hand gestures and body postures, proxemics, paralanguage, body contact, time and scent. Lastly, the design project has been located within the field of social communication and relevant work has been examined. From this we can see that the design of means by which an individual can communicate non-verbally in a social group has barely been addressed. The potential will be discussed in Chapter 3.

Chapter 3:

Existing Avatar Virtual Worlds

Chapter Overview

The role of this chapter is to provide the reader with an overview of the state of the art in 3D AVWs (avatar virtual worlds). Firstly, there is a description of the client-server architecture that is typically used to technically implement virtual world systems. Next, a comparative review is presented that was conducted during the early stages of the project. Its purpose was to gain a thorough understanding of existing AVW technology and to establish common features of the three most popular, publicly assessable, 3D systems, specifically within the context of social communication. This enabled the design problem to be explored in depth and relevant challenges, which are discussed within the last section of the chapter, to be identified.

3.00 Client Server Architecture

As with all virtual environments, the three reviewed are implemented technically using a client-server network architecture. It is important that the reader is familiar with this concept in order to understand the technology. Figure 3-0 shows a simplified diagram of a client-server set up. *Client* refers to a computer and software that a participant uses to access an AVW. Via the Internet, a client contacts a computer termed the *server*, which is a permanently online and running special software that allows multiple client's to connect to it simultaneously. When accessing an AVW a connection to the server is first established, and the client provided with all the relevant information needed to construct the 3D environment and the avatars within it. The server then continues to manage the flow of information between clients such as text or voice messages, avatar movements, arrivals and departures, etc.

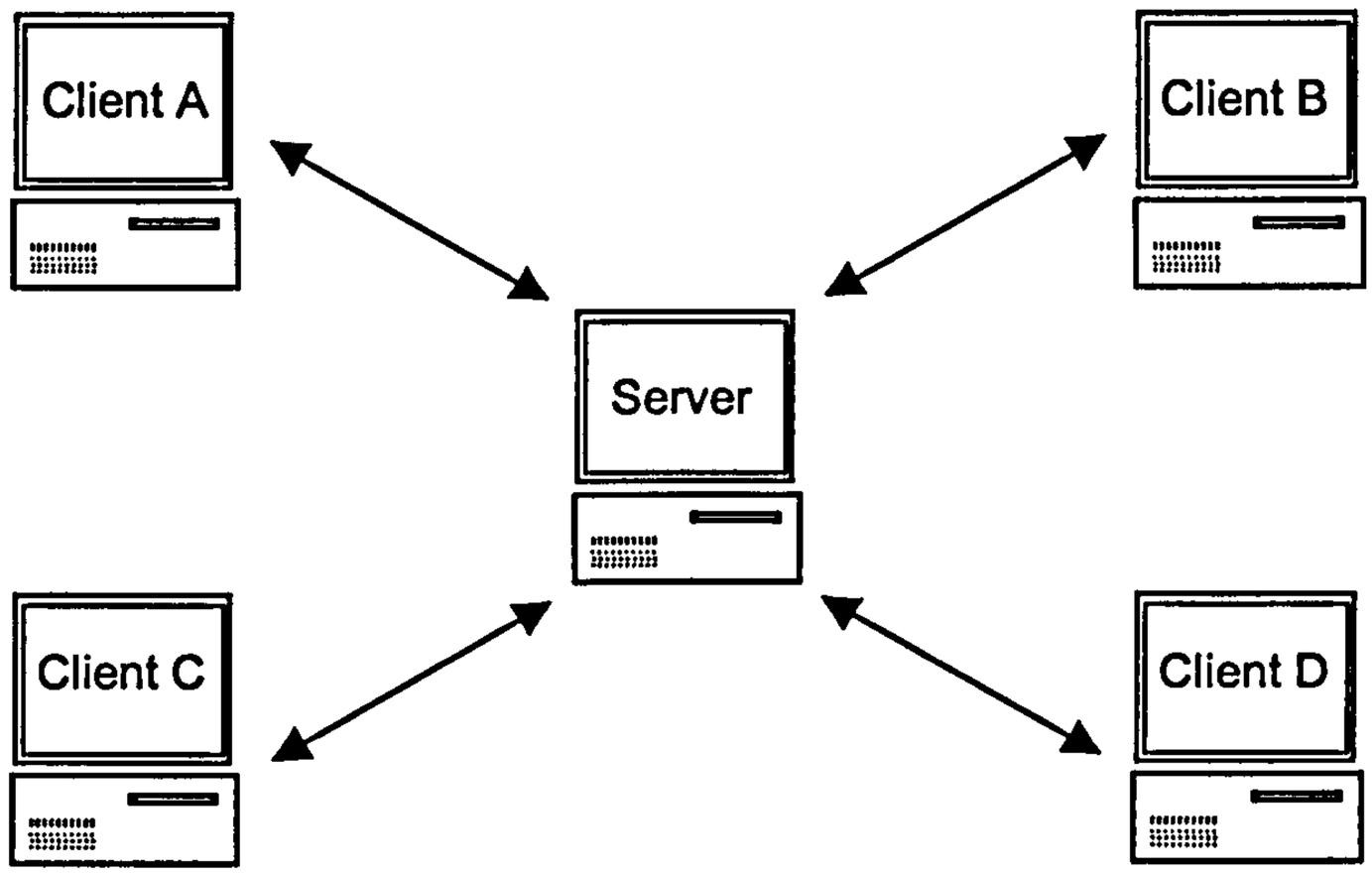


Figure 3-0: Simplified Representation of a Client-Server Network Architecture

Despite the limitations on the amount of information that can be exchanged over telecommunication networks in real time, there is considerable potential within currently available technology for the development of the visual experience for participants. The increased technical demands required to achieve an NVC (non verbal communication) with greater sophistication does not necessarily rely on vast increases in the available bandwidth. Rather, it is the client computers that will be required to process a greater amount of information, which the average home PC is already more than capable of doing.

3.10 Comparative Review of Three AVWs

The following sections present a comparative review of three popular, publicly accessible, 3D, AVW systems, specifically within the context of social communication. Conducted during the early stages of the research (1999) its purpose was to establish common features of AVWs, to better understand the technology and identify the shortcomings, and challenges, concerning social communication that could be addressed by the design project. The method used was to participate in a number of sessions

within three popular AVWs. During these sessions, in which the designer acted as naturally as possible, observations were recorded on paper in note form. Four aspects of social interaction were of particular interest which are representation, communication, existing body language and avatar control.

Representation refers to how participants and software agents are embodied, and identified, in each of the systems reviewed. *Communication* refers to how verbal discourse is achieved as well as more general aspects of group interaction such as how members are made aware of arrivals and departures. *Existing Body Language* also refers to communication but specifically what hand gestures, face expression, body postures, etc., avatars are equipped with and how they are used by the participants. *Avatar control* refers to navigation (this is important as difficulties have been observed which affect social behaviour (see Section 3.13)) as well as how participants activate existing avatar body language during conversation. Sections 3.11 to 3.14 present relevant notes made for each of the headings in the form of comparative tables, which are expanded in the text.

The three AVW systems used for the review are *Blaxxun* (BX), *Active Worlds* (AW) and *Onlive Traveller* (OT) [12]. Although there are others, it is these that attract the most visitors therefore these provided the greatest scope for observing social interaction. Each system presents visitors with a number of worlds from which they can choose to enter, however the style of interaction across them is usually the same. BX from *Blaxxun Interactive Inc.*, was the first to emerge of the three reviewed, and its community and features have continued to evolve. The client software version used for the review was *Blaxxun Contact V4.4*, which is a plug-in for a standard web browser

---

[12] *Worlds Away* web site is at http://www.worldsaway.com
Active Worlds web site is at http://www.activeworlds.com
Blaxxun web site is at http://www.blaxxun.com

e.g. *Netscape Navigator*. AW from *Active Worlds.com. Inc*, is the most popular of the three, probably because of the client software's simple interface and technical stability. Another reason may be because participants are able to manipulate basic shapes in order to build a dwelling for themselves. The version of the client software used for the review was *Active Worlds 2.2*, which is a stand alone piece of software. OT from *Onlive Inc*, is the only system of the three reviewed that features audio messages to achieve verbal discourse, although text is possible for private one-to-one conversations. The client software version used for the review was *Onlive Traveller V2.02*, which is a stand alone piece of software.

3.11 Representation

The table in Figure 3-1 shows how the representational features compare across the three systems reviewed in terms of how avatars are selected, what avatars are available, details of their appearance, how a greater sense of individualism is achieved and how participants are identified. There is also information on the representation of bots. These are software agents that provide assistance to users. Some bots are embodied by an avatar which operates in a similar manner to one representing a human participant. From a participant's perspective the only difference between the two is that the general behaviour of a bot's avatar is controlled by software, rather than another participant. Bots are not yet as advanced as the agents used to create believable characters (see section 2.30). They do not recognise individual users, learn from them, log details about each interaction or provide an emotional response. They simply offer a more appropriate interface to software tools such as help files, within the context of the primary virtual world activity, i.e. social interaction. The key point to note here is that a bot's avatar should ideally use body language when communicating with a human participant, and its library needs to contain a similar set of gestures and expressions.

	Onlive Traveller (OT)	Blaxxun (BX)	Active Worlds (AW)
Allow 3rd party Av's.	No	Yes	No
Types of Av's.	Humanoid & animal	Mainly humanoid but also animal & inanimate	Mainly humanoid
Common Av. Features	Heads only: face, ears, eyes, brows and mouths	Av's. have at least eyes, arms and hands.	All have head, torso, arms, hands, pelvis, legs & feet but few have facial detail.
Customisable Av's.	Yes	No	No
Identification method	Name is highlighted in list when Av. is clicked on	Name is highlighted in list when Av. is clicked on	Name always visible above Av. head
Textual profile	Yes	Yes	No
Bot avatars	No	Yes	Yes
Identifying bots	NA	Tag attached to bot name in list of names	Bot name precedes message in dialogue window & is encased within square brackets e.g. "[bot] can I help?"

Figure 3-1: Comparative Table of the Representational Features Across Three Publicly Accessible Online 3D Avatar Virtual Worlds

All three systems offer participants a choice of avatars from a predefined library which can be assessed at any time. As discussed within the previous chapter (see section 2.60), appearance is an important part of NVC. Therefore, to increase variety and the sense of individualism, BX allows participants to select an avatar from third party libraries or use one that has been created by the participant and uploaded to a URL, therefore achieving greater variety. OT has taken a different approach to this by providing tools for customising the avatars within its library by adjusting basic parameters such as colour, stretch, squash and face expression.

The avatars available across the three systems can be categorised into three main groups which are humanoid, animal and inanimate object. While AW insists upon humanoid avatars of a certain style in order to maintain a theme though consistent visual imagery, BX and OT's libraries consist of combinations of the three. Design features of avatars native to an AVW system often reflect the channels that are available for

communication, for example, OT avatars all have a mouth and ears which reflects the use of audio to achieve verbal discourse. Many AW avatars have no facial detail as communication is achieved through text and, to some extent, hand gesture.

On entry to AVWs, a participant chooses a nickname through which he or she will be known by others. People rarely use their real names as they prefer to remain anonymous. In AW nicknames are displayed above the avatars heads which makes it possible to identity participants at a glance. BX and OT require the user to perform a mouse action such as clicking on an avatar, which results in the owner name being highlighted in a list. BX and OT also feature a personal profile in the form of additional textual information to further increase the sense of identity, true or fictional. This is also accessible though a mouse action and may include nickname, real name, email address, URL of personal web site, etc.

3.12 Communication

The table in Figure 3-2 shows how communication features compare across the three systems reviewed in terms of how verbal discourse (which refers to either text or audio messages) is achieved, authored and displayed, how communication with bots is achieved, and whether macros are featured. In additional to this there is information on more general aspects of group communication which are how participants are made aware of arrivals and departures, how conversations are initiated, and how conversation partners and social groups are identified.

Communication is achieved through the exchange of either text or voice messages which occur so rapidly that communication is considered to be real time. Text messages are typically authored by typing into a text field within the client software.

They are sent to the server by pressing the *Return* key, which subsequently sends them to the other clients. BX displays text messages within a separate dialogue window as does AW, which also displays messages above the avatars. Within OT, audio messages are authored by holding down the *ctrl* key and speaking into a microphone connected to the computer. Audio is streamed to nearby participants and heard through loud speakers or headphones. Stereo and volume levels are used to create a basic sense of spatial orientation and users can disguise their voice by applying effects such as pitch, which helps preserve participant anonymity. Text messages are possible within OT for private one-to-one conversation, which are viewed within a floating dialogue window.

To increase the speed of message authoring BX features macros. These are predefined text messages that can be selected from a pop up menu, that describe physical actions performed by the body in the real world, e.g. 'Apple waves hello to everyone'. Macros appear within the dialogue window and are distinguished from standard messages through a typographical difference, i.e. coloured italic text.

Within all three worlds one-to-one conversations can be visually observed by two avatars standing close together and facing each other, however social groups are more difficult to make out. Only within OT are they sometimes observable through a circular arrangement of avatars. This behaviour probably occurs in OT due to the need to be close to group members in order to hear their audio messages, and lip synchronisation features put emphasis on seeing other's virtual faces.

	Blaxxun (BX)	Active Worlds (AW)	Onlive Traveller (OT)
Verbal discourse	Text (text to speech option)	Text only	Audio (Text 1-to-1 only)
Message authoring	Typed into a text field & sent with 'return' key	Typed into a text field & sent with 'return' key	Pressing 'Ctrl' key & speaking into microphone
Message viewing	Text appears within a separate dialogue window	Text appears within a separate dialogue window & above Av's. Heads	Heard by those nearby. lip synchronisation show who is speaking (text viewed in floating window)
Macros used	Yes	No	No
Arrivals & departures	Alert in form of text message from a disembodied bot e.g. "Groovy Mover has joined the group"	Indicated by participants text messages e.g. "bye all"	Audio alert
Initiating conversations	1. Move into view, 2. perform gesture, 3. send text message	1. Move into view, 2. perform gesture, 3. send text message	1. Move into view, 2. perform gesture, 3. send audio message 4. send text message
Identifying one-to-one conversations	Participant avatars close together and facing each other	Participant avatars close together and facing each other	Participant avatars close together and facing each other
Identifying social groups	Not possible, Avatars usually scattered	Not possible, Avatars usually scattered	Groups occasionally form circular formations
Communication with bots	Text only - conv. initiated by including bot name within message.	Text only - conv. initiated by including bot name within message.	NA

Figure 3-2: Comparative Table of the Communication Features Across Three Publicly Accessible Online 3D Avatar Virtual Worlds

Conversations can be initiated by taking one, or a combination, of the following steps. Firstly, navigating one's avatar into view and range of the recipient. This will be taken as a signal that one wishes to begin conversation. Secondly, a visual behaviour can be performed for someone who is in visual contact, e.g. waving hand. This method may also be used to reinforce the first method. Lastly, by sending a message addressed to the person one wishes to contact. In the case of OT, recipients must be within hearing range, unless one-to-one conversation is initiated by sending a text message. Communication with bots is achieved through text, and interaction is usually initiated by including a bot's name within a message. Further messages are scanned for

keywords which are used by the bot to retrieve relevant information from a data base. Feedback is returned in the form of a text message addressed to the participant.

3.13 Existing Body Language

The table in Figure 3-3 compares the avatar body language across the three systems. This includes the forms of NVC used, what a typical set consists of, what the visual properties are, how the behaviours available are used by participants during dialogue, what body language bots use and what lifelike behaviours have been implemented.

Most avatars are capable of animation and can perform a few basic gestures or facial expressions. Most humanoid avatars in BX and AW use hand gestures to convey basic expression and occasionally to support social interaction, e.g. wave to offer conversation. They are short in duration, exaggerated to make them more visible, and begin and end in a natural position. OT uses heads for avatars, so it can concentrate resources on face expression and lip synchronisation, which can also indicate who is speaking. The expressions are used to convey mood, therefore they are set by the participant and remain so until changed. Despite the available behaviours, once the novelty has worn off, participants rarely use them during conversation. Bots do not use body language at all other than moving towards a participant who has requested assistance.

Lifelike animations are a technique used by AVW creators to break away from the early static models and give avatars a greater sense of life. AW avatars have as many lifelike animations as they do meaningful gestures for example, scratching the head and looking around the environment. Lifelike animations are performed at random intervals and are beyond the control of the participants.

	Blaxxun (BX)	Active Worlds (AW)	Onlive Traveller (OT)
No. gestures/emotions	5 - 10 gestures	3 - 5 gestures	4 emotions
Typical set	Hello, hey, yes, smile, frown, no, not, laugh & goodbye	Happy, angry, wave, jump, fight & dance	Happy, sad, angry & surprise
Visual properties	Exaggerated, short duration, fast, minimal face expression, begin and end in a neutral position, they can only address one person at a time	Exaggerated, short duration, fast, no face expression, begin and end in a neutral position, they can only address one person at a time	Face expression only, emotion stays set until changed by participant, they can only address one person at a time
Usage by participants	Performed infrequently. They are used to convey basic expression and allow the avatars to appear more lifelike.	Performed infrequently. They are used to convey basic expression and allow the avatars to appear more lifelike.	Used infrequently. They are used to convey mood.
Bot body language	None	None	NA
Lifelike animations	None	Several random animations involving whole body e.g. look around environment	Avatars blink randomly

Figure 3-3: Comparative Table of Existing Body Language Across Three Publicly Accessible Online 3D Avatar Virtual Worlds

3.14 Avatar Control

The table in Figure 3-4 compares the features provided to achieve avatar control. These include how a participant navigates into a position to offer conversation or join a group, what features are provided to assist this process, how awareness of the surrounding environment is achieved and how existing body language is controlled.

In all three AVWs participants are able to navigate their avatars around the virtual environment to find an individual to talk with, or a group to join, using the keyboard cursor keys or the mouse. Barbara Becker and Gloria Mark have reported that the proxemics of AVW correlates to real life [Becker & Mark 98], therefore when making an offer of conversation the spatial orientation and distance between avatars is an important factor. However, navigation in AVWs is a difficult process which is due to a number of factors such as the absence of shadows and bad camera views, which provide

vital information about the surrounding environment. Achieving the correct positioning when making an offer of conversation is subsequently an awkward task. To address this BX and OT feature methods that assist the process of navigating known as *beaming* and *go near*. By clicking on the recipient's name within the list, also on the their avatar in OT, a participant is automatically transported to face their potential conversation partner.

	Blaxxun (BX)	Active Worlds (AW)	Onlive Traveller (OT)
Methods to approach potential conv. partners	1. Move into position using keyboard or mouse 2. click on a recipient in list of names to be *beamed* over to face them	1. Move into position using keyboard or mouse	1. Move into view & range using keyboard or mouse 2. Click on a recipient in list of names to *go near* 3. Click on recipients avatar to *go near*
Features to assist navigation	Collision detection can be disabled to pas through avatars and world objects	Can pass through other avatars but not world objects	Avatars can pass through world objects but not other avatars
Environment awareness	1st and one 3rd person camera view	1st and a number of good 3rd person camera views	1st person view and a number of awkward 3rd person camera views
How gestures/emotions are triggered	Manually using a tool bar	Manually using a tool bar or menus	Manually using a tool bar

Figure 3-4: Comparative Table of Avatar Control Features Across Three Publicly Accessible Online 3D Avatar Virtual Worlds

A feature provided to assist in standard navigation is to pass through other avatars or world objects and this is handled differently within the three systems. BX allows collision detection to be turned off so both objects and avatars can be passed through, thus saving the participant from the need for complex manoeuvring. AW allows participant to pass through other avatars at all times. This behaviour is not usually considered rude. Within OT participants are able to pass through world objects, and not avatars, thus giving other participants a greater sense of presence.

Each of the systems allows the user to choose from a number of preset camera views in relation to their avatar. First-person perspective is usually the most effective for navigating the environment, however AW offers a number of useful third-person views. Within BX and OT third person views are generally only useful when stationary, to glean information about one's surroundings. One would expect the third person view in OT to assist in the task of navigating into a position to join a social group, but this is not the case due to a poor camera position.

Once engaged in conversation participants are able to use the preset gestures and emotions available for their particular avatar. At the time the review was conducted all three systems required the user to manually activate these through buttons located on tool bars within the client software. Since that time however, BX has developed *CyberTown*, which features an autonomous activation system based on the detection of keywords within text messages. For example if the word "hello" is detected, the avatar waves its hand. Each gesture has an average of three trigger keywords, which can be customised by the participant.

3.20 Discussion of Shortcomings

The comparative review identified a number of graphical shortcomings in AVWs and challenges for the design of advanced systems in respect to social communication and body language that the design project will address. Participants of AVWs use the gestures or emotions available only when new to an AVW system when exploring the various features of the client software. As soon as the novelty wears off and they become engaged in dialogue, participants cease using these behaviours and the avatars have little interactive and communicative value. There are three probable reasons for this.

Firstly, in all three worlds reviewed the behaviours available are few and poorly chosen. There are one or two that are useful in a handful of potential social scenarios, for example, *waving* the hand can be used to place an offer of conversation, or to indicate that a conversation is over. However, most of the gestures available in BX and AW are suitable for conveying basic expression (e.g. *jump* for joy), but are rarely used by participants to directly complement verbal messages. In OT the emotions available are not used at all, perhaps because it is easier to convey them via the audio channel in which, for example, laughter indicates happiness. In addition to this, the visual properties of existing behaviours are such that they can only be used to address an individual. If a participant wishes to greet all members of a social group using gesture, he or she is usually required to undertake complicated manoeuvring of their avatar involving the activation of an appropriate behaviour for each participant in turn, e.g. wave. In order to increase the usefulness of the graphical dimension, a richer and more varied body language could be designed that conveys the often unspoken indicators of social behaviour such as turn-taking, that supports an individual interacting with a social group, and that integrates with verbal discourse in a meaningful way.

The second probable reason gestures and emotions are rarely used within existing AVWs is that group members are not usually visible to one another. Avatars engaged in one-to-one conversations are sometimes observable by virtue of their orientation and close proximity to one another, but members of larger groups are usually scattered around the environment (the reader can observe the haphazard arrangement of the social group's avatars in Figure 3-5). The significance of this is that there is no point in one performing gestures and emotions because it is unlikely that they will be visible to one's conversational partners. In addition to this, the scattering of avatars means that it is not

PAGE NUMBERS CUT OFF

IN

ORIGINAL

possible to determine what social groups exist, who has group membership and to what group members belong.

Figure 3-5: Screen Shot of a Typical Scene in Alpha World, Active Worlds Inc.

The scattering of avatars is caused by two main factors. Firstly, navigational difficulties can make manoeuvring into a group formation a difficult and frustrating task, thereby discouraging participants from attempting the task. The current *beam* and *go near* features offered by BX and OT are adequate for approaching an individual, however, they break down when one wishes to approach a group of more than two. An attempt to do so will place an avatar on top of, or between, the existing conversation partners, obscuring their view and invading their personal space. Secondly, and of great concern, is the disjunction of the various channels of communication caused by the use of separate dialogue and visual windows. Participants are not required to physically approach each other to make contact, and whilst communicating their attention is focused on the text. Hence, in the first case the visual is not utilised effectively and in the second attention is drawn away from the visual. Although AW also displays text messages above the avatars, they are usually impossible to make out as the

conversational partners are either too far away or the environment is crowded with overlapping messages. A solution needs to be found to make group members visible to one another as a precondition for effective avatar body language. In addition to this, it would be helpful to find a way of integrating the text and image communication channels so attention can be more easily divided between looking and reading.

The third probable reason why the existing gestures are rarely used is the inefficiency of activating them from tool bars and menus. In real life many of the visual clues that we provide during conversation are spontaneous and unconscious. The activity of authoring messages however, is very different to that of using the mouse to select tool bar buttons which requires a participant to take conscious action (with the exception of *Cyber Town*). A better method is required to control an avatar's body language, ideally one that integrates gesturing with messaging. As the research here is primarily concerned with deriving and implementing a more useful body language, this issue will not form the main focus of the project. It is noted here because it is an important issue which will need to be addressed at some point in future work if an avatar body language is used by participants. Furthermore, as we shall see, there are aspects of the work described later in the thesis which are formed by ideas that have emerged during the course of the project on how to address this problem (see sections 4.13 and 9.20).

One additional problem with existing avatar body language is that the lifelike behaviours often bear no relevance to the conversation taking place and can consequently can result in misleading or even conflicting signals. For example, in AW, one of the lifelike animated behaviours is to randomly gaze around the environment. If this is performed when an avatar is addressing a nearby participant, the action of looking away from the intended recipient is likely to conflict with the content of the

message, i.e. one would usually expect to see the head turn towards the recipient rather than away. While the author is in favour of lifelike behaviour because it can achieve a stronger sense of the presence of other beings, there appears to be a need to rethink how it is used within the context of group social interaction. What role will this kind of behaviour play within the body language that is to be designed?

3.30 Summary

This chapter has provided the reader with a thorough understanding of existing 3D AVWs and how they were implemented at the commencement of the project. The client-server architecture used to technically implement these systems has been described as a computer network, where the server is permanently online and running special software that allows multiple clients to connect to it simultaneously. The server then continues to manage the flow of information between clients such as text or voice messages, avatar movements, arrivals and departures, etc.

A comparative review was presented that established common features of AVWs, and enabled identification of a number of shortcomings and challenges specifically within the context of social communication. The review focused on the three most popular publicly assessable systems, and observations made while participating in world sessions were documented on paper in note form under four headings: representation, communication, existing body language and avatar control.

The various visual design challenges which will be addressed by the design project have been discussed and are summarised as follows. Firstly, in order to increase the usefulness of the graphical dimension, a richer and more varied avatar body language will be designed that conveys the often unspoken indicators of social behaviour such as

turn-taking, which supports an individual's interaction within a social group and complements verbal discourse in a meaningful way. To achieve this, it is desirable to integrate the different channels of communication, i.e. text, audio, proxemics and body language, so that multi-modal communication can be achieved.

Chapter 4:

Design Brief & Considerations

Chapter Overview

This chapter defines the design task which aims to address the various challenges of social communication raised within the previous chapter, through background research and creative practice. Firstly, the design brief is described which breaks the endeavour down into two smaller, manageable and related tasks, which will be integrated in the later stages of the project. Following this, there is discussion of a number of general design considerations that it was necessary to address at the start of the project. This leads to a set of design principles that gives the creative process form and ensures that a feasible solution is reached within the limitations of current and imminent AVW (avatar virtual world) technology.

4.00 The Design Brief

The design brief primarily seeks to derive and implement a more useful avatar body language within the context of social communication. This endeavour is broken down into two smaller, manageable and related tasks which are addressed separately, and will be reintegrated in the later stages of the project:

1. In order to create a framework for multi-modal communication, find a solution to make group members visible to one another while integrating the various channels of communication, i.e. text, image and audio.

2. Design an avatar capable of performing body language, and a limited set of behaviours for it that supports group social interaction and integrates with verbal discourse in a meaningful way.

The first task aims to create a framework onto which multi-modal communication can be built. In order to achieve this there are two main objectives. Firstly, to make group members visible to one another. This must be achieved before a virtual body language is developed, otherwise it will not been seen by the relevant group members when used. It is anticipated that the solution lies in the gathering and arrangement of group members' avatars, as well as determining a number of scenic properties such as camera position and lighting effects. The second objective of this task is to integrate the text and visual channels of communication, so they begin to work together rather than compete for participants' attention. It will be necessary to explore the visual properties of the text, e.g. how will it be displayed and how can it be made clear who has authored which message.

For the second task an understanding of both the social dynamics of AVWs, and of the relationship of human body mechanics to social meaning, will be required. From this, it will be possible to determine the minimal set of operators necessary for an avatar model in order to effectively convey meaning. There is also a need to consider what the best forms of NVC (non verbal communication) are for avatars, how to arrive at a set of useful behaviours, what their visual properties should be and how the set can be technically implemented without too great a compromise of the designer's ideal solution. This is important as the highly technical nature of AVW technology and 3D modelling software can lead the designer to the final outcome.

4.10 Design Considerations

Before addressing the design brief it is necessary to consider a number of design constraints and principles that will influence the end solution. Constraints are limitations on design that are generally imposed by AVW technology. Principles refer

to decisions made about the approach that will be taken to certain aspects of the design, for example whether to use text or audio to achieve verbal discourse. In the remaining sections of this chapter relevant design considerations are examined in turn under the following headings: polygon budget, visual style, text or audio to achieve verbal discourse, technically achieving body language and useable forms of NVC.

4.11 Polygon Budget

Bandwidth and processing power limitations demand that avatar models are kept to a minimal file size. To understand how this is achieved, it is necessary to describe an avatar in terms of geometry. An avatar is a 3D geometrical form that is defined by groups of polygons. A polygon is the surface generated between three or more coordinate points in 3D space that are joined by straight lines. Everything is made up of these flat surfaces therefore the more detailed an object, the more polygons the rendered model will contain. The number of polygons, however, is respective of the file size of the avatar, i.e. the more it contains, the larger the file size. Part of the challenge of avatar modelling is therefore to create forms that are aesthetically appealing while working to a polygon budget. This ensures reasonable download times and processing performance. Texture maps also increase file size and are therefore not common in avatar models, nor are they used within the design project. These are images that are wrapped onto the surface of 3D objects in order to achieve detail and, often, photo realistic image quality. Because the designer wishes to work within the confines of existing and imminent technology, the technical limitations will be respected and therefore the avatar will consist of a reasonable number of polygons (e.g. 1000). For more information on avatar construction see [Wilcox 98].

4.12 Visual Style

The visual style of an avatar can be categorised as realistic, symbolic or naturalistic. Realistic avatars strive to achieve a convincing image of the user, or other character that one wishes to temporarily adopt, e.g. a celebrity. Symbolic avatars are abstractions of a participant's real, or self-created, character. For example, one may choose an owl which can symbolise wisdom. Naturalistic avatars are usually humanoid in form, of a degraded level of detail, but emulate natural paradigms just enough to achieve recognition of familiar features.

The designer may experiment with more than one type of avatar at different stages of the project, however the final solution will be naturalistic for two reasons. Firstly, photo-realistic avatars are currently too demanding for existing and near future technology. Secondly, it makes sense to begin by developing a gestural language that is based on that of the real world in order to achieve one that is intuitive to users at any level of experience with AVWs. If this can be achieved, then future work can investigate how it can be transferred to other kinds of avatar, e.g. inanimate forms such as a clock.

4.13 Text Or Audio Verbal Discourse

The designer is in favour of avatars whose design reflects the media available in the AVW it which it is used, therefore it is necessary to decide whether to design for a world that uses text to achieve verbal discourse, or one that uses audio. Both present a number of advantages, for example text is far easier to process and demands fewer resources, enabling more to be assigned to a NVC with greater sophistication. On the

other hand, using audio eliminates the need for a separate dialogue window and is therefore closer to achieving a multi-modal communication.

It is decided to design for a text system for two main reasons. Firstly, it is anticipated that eliminating the need for a separate dialogue window will be achieved with relative ease. Secondly, it is anticipated that text systems will eventually offer a better method of controlling avatar body language than the existing tool bars, i.e. through an autonomous system based on the detection of keywords within messages. This would be far more difficult to achieve with audio, and in any case, if voice is really desired then text to speech software can be used. It is also decided that there will be one gestural complement per message, rather then per word. This will either convey the spirit in which the message is written, carry unspoken meanings or support verbal discourse.

4.14 Technically Achieving Body Language

Currently there are few practical methods for technically achieving avatar body language in a desktop system with limited input devices. One inexpensive solution is to use a web cam to continuously stream video footage of a participant's face onto their avatar's head [Çapin *et al.* 97]. The advantage of this approach is that the avatar will be able to display the participant's full array of facial expressions in real time. In practice however, this is demanding on technical resources and the result is usually a crude and jerky image. Beyond this technical limitation, which will no doubt be resolved in the near future, there is another difficulty in that the user's body resides in a space that is radically different from their virtual body. If the participant turns to look away from the screen, their avatar's face shows the side of their real head.

A second method that also employs a web cam is model-based coding of facial expression [Pandzic *et al.* 94, Thalmann *et al.* 95]. Instead of using streamed digital video, the moving image is analysed by the client computer and a set of parameters describing facial operators is extracted, e.g. head rotation, head inclination, aperture of eyes, eyebrow elevation, jaw rotation, mouth aperture, mouth stretch/squeeze. These parameters are then imitated in an avatar model which responds to changes in real time. Unlike other similar systems there is no need to position on the face white markers used to assist the automatic extraction of parameters. A considerable range of facial expressions can be achieved and the data can be mapped onto any avatar with corresponding operators so it does not need to look like its owner. The system however is limited to tracking facial expression and shares the difficulties of the real and virtual body residing in very different environments as in the streaming video approach.

One possible method of extending the two systems discussed is to modify them to also track the user's hands. This has been successfully implemented by Kristinn Thórisson in her work with J.Jr., a software agent that uses natural language to recognise when it is its turn to speak [Thórisson 94]. A camera is attached to the computer so the agent can monitor the user's hands. If there is a pause in speech J.Jr. checks the user's hand behaviour. If they are also stationary then the agent recognises that it is its turn to speak. This system could be developed to track actual hand gestures which can be mimicked by the avatar. The only flaw in this idea is that it may prove problematic if a participant's hands spend the majority of the time on the keyboard authoring text messages.

The method chosen for the design project, is to utilise the library of predefined behaviours approach that is favoured by existing AVWs such as *Blaxxun*, *Active Worlds*

and *Onlive Traveller*. The advantages of this are firstly, that the body language developed can exploit all channels (i.e. face expression, gaze, hand gesture, body posture, etc.), thus being closer to achieving multi-modal communication. Secondly, it is possible to achieve correct avatar gaze direction although there is still the problem of tracking the participant. Lastly, the behaviour animations are portable and can be loaded into other naturalistic avatars, providing they are of similar form. As discussed in section 3.20, the problem with existing implementations is that the gestures are simplistic, non-varied and are not integrated with other channels of communication.

4.15 Useable Forms Of NVC

It is necessary to consider the question: what are the best forms of NVC to implement in avatars? The various forms of NVC that were described within Chapter 2 (see section 2.60) are now examined for their suitability for AVWs. Firstly, appearance and artefacts will not be considered for the project as this form of communication is already handled by the participant's choice of avatar and props that he or she adds to it. Facial expression appears to be a well suited form of NVC to use, however it is subtle and changes are fleeting. This is not a problem in real world communication, however it is suspected that on a computer screen facial expression could be overwhelmed by other information such as text messages, particularly when participating in group conversation. *Worlds Away*, a 2½D AVW, has addressed this issue by giving avatars enlarged heads and exaggerated facial features, thus making them more visible. *Worlds Away* also uses sound alerts to inform participants of a face expression change. Although effective, this principle produces a cartoon-like aesthetic which conflicts with the naturalistic style used in the design project. While being a powerful vehicle for expression, the suitability of facial expression for avatars needs further investigation. It

is anticipated that a better understanding will be reached by the end of the design project.

Body posture and hand gesture would seem to be the ideal forms to exploit in avatars as they involve larger movements making them easier to see and, between them, they can compensate for the absence of other forms. Conclusions from Vuillème's experiment (see section 2.70) suggest that hand gestures will be used more by participants than body postures. He speculates that this is due to gestures being more consciously activated than body postures which people forgot to change in the experiment [Vuillème 98]. Body postures convey mood, and changes in a participant are difficult for the computer to automatically track, therefore hand gestures will be the focus of the two. Gaze should be an important part of the NVC to develop, however it cannot be used in the same elaborate manner as it is in the real world. It is hoped that gaze code can be simplified and used to help resolve issues of turn-taking and to indicate who is talking to whom. Proxemics (space) can be exploited within AVWs, particularly when arranging group members avatars. However, at this early stage of the project it is unclear exactly how this may be implemented due to the navigation difficulties which make exact positioning of avatars a difficult task.

Body contact is used less frequently than the other forms of NVC discussed previous, and whilst it is technically possible to implement it in an AVW, it raises many difficulties. For example, for two avatars to shake hands they must manoeuvre into an exact position relative to one another and extend their hands to meet at a point in space. These coordinates will change for avatars of different sizes and form, which makes such an action difficult to achieve when employing a library of predefined gestures. Given these difficulties and the limited role of body contact in communication it is reasonable

to practice effort focus on the more frequently used forms of NVC, e.g. hand gestures. Therefore physical contact will not be employed in the body language designed. Paralanguage is only relevant for systems that use voice to achieve verbal discourse in which case operators such as voice tone, tempo, rhythm, articulation and pitch range, are carried with the message. Some aspects of communication through time will be handled by participant behaviour, e.g. punctuality, while others will be handled by appearance (choice of avatar), e.g. age. Using scent is not feasible at present as the available media are limited to audio, text and image.

4.20 Summary

This chapter has described the design task, the constraints imposes upon design and a number of principles that will be adopted. A design brief has been drawn up which aims to develop an avatar capable of multi-modal social communication, through background research and a series of creative pieces of work. The endeavour has been broken down into two smaller, manageable and related tasks. Firstly, to integrate the various channels of communication into a single space/window. The solution must ensure that group member's avatars are visible to one another in order that body language can be seen. Secondly, to design an avatar capable of performing body language, and a limited set of behaviours for it that supports group interaction and social discourse.

A set of general design principles has been defined which are as follows. The design will take place for AVWs that use text to achieve verbal discourse. The final avatar will respect the technical limitations and will meet a polygon budget. The avatar will be naturalistic in style, i.e. humanoid, of a degraded level of detail, but will emulate natural paradigms sufficient to achieve recognition of familiar features. There will be one

gestural complement per message that will either convey the spirit in which the message is written, carry unspoken meanings or support verbal social discourse. The body language developed will consist of a limited number of predefined behaviours that are located within a library. These behaviours will exploit four forms of NVC for AVWs which are face expression, gaze, hand gestures and proxemics.

PART 2: PRELIMINARY STUDIES

Chapter 5:

Understanding the Social Dynamics of Virtual Worlds

Chapter Overview

This chapter presents the first of two studies conducted as background research to inform the primary design activity. The objective in undertaking the study is to identify aspects of text-based communication that an avatar body language might complement and this is achieved through an analysis of participants' social behaviour in multi-participatory virtual worlds. It was necessary to conduct the study described within this chapter because there was little published material of direct relevance on the subject available at the time. Two dialogues generated from MUD sessions are subjected to analysis in order to identify patterns of social behaviour and a diagrammatic technique is developed to assist the interpretation of the data gathered.

5.00 The First Study

The background research begins with a study that focuses on aspects of the discourse between participants of multi-participatory virtual worlds. As discussed in section 3.20, existing body language available in AVWs (avatar virtual worlds) rarely supports verbal discourse which is one probable reason why it is not used during conversation. It is hoped that by developing an understanding of participants' social behaviour within virtual environments it will be possible to identify specific aspects of textual discourse that can be complemented through an avatar body language. To better understand how an avatar body language can be more useful, two dialogues generated during multi-participatory virtual world sessions are selected and subjected to discourse analysis. In particular, it is hoped that the study will point to some of the challenges of text-based communication within the context of social interaction that can be addressed by the design of other, non verbal, channels. MUDs are used for the analysis as they are completely text-based.

5.10 Overview of MUDs

Apart from being text-based, MUDs operate in the same manner as AVWs. Each participant is identified by a self-chosen nickname and will be able to see a list of the names of the other people present in that room which is updated each time there is an arrival or departure. Many MUDs allow participants to make and name a room of their own which will operate in the same way as a room provided by the multi-user server software. The only difference is that these rooms cease to exist when the room becomes empty.

The medium has certain specific features. Firstly, in these environments the real participant usually chooses to be anonymous (i.e. only their MUD nickname will be known to other participants). Secondly, communicating through a textual interface removes many of the additional channels of communication that usually accompany discourse such as body language. Thirdly, technical stability can not be guaranteed in practice. A common event in these environments is that a technical breakdown occurs and one or more participants lose their connection to the multi-user server without warning. This is also known as getting *booted*.

5.20 Method

The chat environment used to gather data for the study was *Talk.com* by *Wired* magazine. This environment no longer exists, however it was one of the more popular ones operating from the United States. It was mainly used as a social meeting place and had several rooms which a participant can choose to enter. Because participants are aware of everyone who is logged into the environment it was not possible for a designer to be present as an uninvolved observer. This requires the designer to decide how to

behave when observing the dialogue. The decision was taken, at the initial stage of the study, to try to act as naturally as possible: to enter the environment as someone who had recently discovered the MUD and had returned for another chat as this was the real-life position of the designer. There was therefore no particular objective in mind other than to chat and no statements or questions were prepared in advance.

The primary interest of the background research described within this chapter is to increase understanding of the patterns of behaviour and social interactions that are taking place in the MUD environment and, because the only medium of communication is through the exchange of typed messages, the key to this is an analysis of conversations that have taken place. After participating in and observing six sessions, two dialogues were selected for further analysis (see Appendices A and B).

The first dialogue is named *One-to-One* because it consists of a single conversation between two people with no other participants entering the room. The duration of this conversation is thirty-five minutes. An analysis of the text of *One-to-One* is conducted at both word and message level. It is hoped that this will provide a base-line understanding of the mechanisms at work when two people in a MUD environment interact with each other. It should be noted that this example describes an unusual conversation and is used in this research as an illustration of fundamental mechanisms, rather than as a demonstration of their general importance within MUDs.

The second dialogue is named *First Steps*. The duration of this session is one hour and twenty minutes and it involves a total of nine participants, with a minimum of three at any given time and a maximum of five. Unlike the *One-to-One* dialogue, it does not represent one continuous conversation from beginning to end. Instead it represents

several conversations that develop and, to some extent, intertwine. The method involves seeing how the observations that were made in *One-to-One* might apply when more than two people are involved, i.e. within a small group. To achieve this a text analysis of the dialogue was conducted. To further aid the identification and understanding of the patterns of conversation a linear diagram was constructed which is explained in Section 5.51.

One aspect of the method used that needs to be documented, is that the designer did not declare his intention to observe and record what participants say to each other. If one of the essential elements of these environments is a sense of trust and obligation, then keeping this hidden means that the designer is immediately acting in bad faith unlike (it is assumed) the other participants. On the other hand, if the designer declares his true intention then this is likely to greatly affect the dialogue that follows.

5.30 Additional Terminology

In order to preserve the anonymity of participants, they have each been given an alias, i.e. the name of a fruit. Italicised text is used when referring to a participant in the text e.g. *Apple* said "I love to eat". To avoid both gender connotations and cumbersome text, the use of the pronouns *him* and *her* are varied. Many people in MUDs present themselves as the opposite sex so it would not be possible to know the gender group of each participant from the data gathered.

The term *message* refers to a single continuous piece of text that is sent as one unit to the MUD address by a participant, and is subsequently received by all other participants. It is important to note that the structure of messages is not based upon

linguistic units, such as sentences. A message may contain several sentences, or a single sentence may be split over two or more messages (Figure 5-0).

Apple: in fact i love to eat
Orange: they give a lot of energy but don't make you hyper
Orange: and are very good for ya
Kiwi: so do i! i only ate 1 time today or yesturday and that was like 3 pierogies
Orange: you can get um at health stores
Kiwi: thats cool! i will have to try that!

Figure 5-0: A Typical MUD Conversation

When a message is read it is usually, though not always, seen by a recipient as referring to one or more previous messages, in which case there is said to be a (directional) link from the later message to the earlier one. The term *conversation* is used to describe a sequence of messages linked to each other in this way. The start of a new conversation is usually marked by a message that does not refer to any previous message.

5.40 Analysis of the *One-to-One* Dialogue

The *One-to-One* dialogue (Appendix A) is, in some senses, very straight forward and provides an opportunity to develop a simple analysis which can be applied to more complex situations. Although it raises no real surprises, it is worth documenting carefully as it provides a base line from which complexity can be measured. The dialogue forms a narrative which divides naturally into three parts, which are labelled *Start*, *Middle* and *End*. In the *Start* section, the participants engage in dialogue, try to get to know each other and negotiate a meaningful topic of conversation. In the *Middle* section the participants undertake a specific activity. In the *End* section the participants negotiate a way to disengage from the conversation.

5.41 Word Level Analysis

To conduct the word level analysis, each message within the *One-to-One* dialogue (see Appendix A) was examined for individual words and symbols that are used to augment discourse. As they were identified, they were noted on paper, and their role within the larger activity of social communication was observed (see Appendix C). From the analysis it is clear that the participants had employed devices to enhance their communication in a text-only medium. These can be organised into three categorises which are *performative words*, *acronyms* and *punctuation marks*. These contribute to social communication by compensating for the forms of body language missing in this context. In a graphical system where these absent forms of body language can be provided, we might seek to revise the need for these linguistic adjuncts through NVC.

Performative words describe a bodily action or sound and are usually identifiable within the text through the use of two asterisk symbols, for example ' *wave* ' (e.g. message 02, Appendix A). An acronym is an abbreviation for a string of words commonly used when communicating in a virtual world, for example 'LOL', which is an acronym for 'laughing out loud' [13] (e.g. message 14, Appendix A). It is likely that participants use acronyms to decrease the length text messages, thereby increasing the speed of message exchange and the sense of social interaction. However, many of these, such as 'LOL', also describe a physical action or a gesture. Some punctuation marks used in traditional forms of grammar are also observable, but not necessarily used in the traditional manner. For example, the full stop is used to create between two and five consecutive dots (i.e. '.....') either in the middle or at the end of a message. In the middle of a message they usually represent hesitation or silence such as interruption during speech. At the end of a message they usually imply that the message is unfinished (e.g.

messages 01 and 49, Appendix A). Some implications of this will be discussed in the next section.

5.42 Message Level Analysis

Part of the study is to increase understanding of the social dynamics of virtual worlds because the designer believes that such an understanding should inform the design of an avatar body language. With this in mind the designer became interested in what happens at different points of a conversation, therefore the message level analysis is aimed at investigating the following questions: how do people get into conversation?, how do people set up parameters for the conversation? and how do people get out of a conversation?

When the two participants first get into conversation they typically acknowledge each other's presence in the room by way of a greeting. Their mere presence, it may be assumed, means that they feel an obligation to find something to talk about. Within the *Start* section smaller sections could be identified that explore potential topics of conversation. These explorations often reference the few basic assumptions that it is reasonable to make about other participants. For example, it is reasonable to assume that the other person is in front of a computer monitor, so one of the first comments is "blinking at the screen..." (used by *Pear* at message 03, Appendix A).

After exploring several avenues, the conversation returns to one of the initial topics, the idea that the two participants should create a story. Its acceptance by the other participant creates a new conversation context, one of increased commitment to each other, within which the participants have agreed to co-operate in the construction of a

[13] such abbreviations have become standardised, see http://www.acronymsonline.com/chat_acronyms.htm

narrative. From this point, the structure of the dialogue changes radically in line with the new conversation context. The messages become much longer (i.e. contain more characters), there is a very regular pattern of exchange of messages and as a result the dialogue itself becomes fluent and easier to understand (see message 54 onwards, Appendix A). Why is this and how are the rules, parameters and conventions for such behaviour established?

Within the dialogue *Apple* suggests that they construct a story together and proposes two basic rules: (a) to make up one line at a time (message 52, Appendix A), and (b) to take it in turns (message 53, Appendix A). This is a clear and explicit offer, which *Pear* accepts but not explicitly. Rather, he responds with the first line of a story, "once I was a little boy" and adds "[your turn]" at the end of the message (message 54, Appendix A). This message, through its content, is an acceptance of the offer to go into story-telling mode, and the indication that it is unfinished invites the other participant to respond. The next time *Pear* sends a message, "[your turn]" is replaced by the consecutive dots symbol that is identified above and which is used consistently throughout this middle section of the conversation.

It is reasonable to believe that the commitment the participants have to each other is further increased by the nature of the activity (i.e. constructing a narrative). Narrative is normally thought of as having a beginning, a middle and an end and because of this the participants have an obligation to complete each message in a way that both continues the narrative and drives it forward to reach a point where the story can be completed. The rules and conventions for playing this game are firmly established. In all six of the dialogues gathered, participants tend to use short messages (this is evident in *First*

Steps, Appendix B), probably to increase the sense of social interaction through rapid message exchange. The heightened sense of commitment during the story telling activity creates a sub-environment in which the participants can compromise the convention of using short messages, allowing them to become more descriptive in their use of language. The messages are semantically connected and are provided in a fixed order (i.e. in turn). Greater fluency is achieved in the conversation as the pair develop and work within a committed environment.

The data suggests that during the storytelling activity the participants level of commitment to the interaction task (i.e. creating a narrative) increases. Also, it would not seem appropriate to end a conversation abruptly, say by simply informing the other that one has to go. In this dialogue, these two distinct obligations run into each other (i.e. commitment to the task and commitment to the social convention for disengaging from a conversation). So how do the participants indicate that they wish to end the game without compromising their commitment to complete the story? At message 68 (Appendix B) *Pear* offers an ending to the story. *Pear's* method for ending the story is ingenious: he brings a character into the story who is also named "Pear" and who intends to leave soon ("... to watch Columbo on television"). *Pear's* construction means that, in one move, he addresses the commitment to complete the narrative, and he is able to end the conversation.

Having completed the construction of the narrative, the participants still do not immediately leave but conclude their conversation by saying "thank you" and "good-bye" to each other. It is likely that this is not just a signal that the conversation is over, but also an acknowledgement it is satisfactorily concluded.

5.43 Summary of Main Observations on *One-to-One*

1. The participants begin with a greeting to acknowledge each other and initiate the conversation.

2. There is a process of negotiation about potential topics, based upon common knowledge and shared aspects of their circumstances.

3. In middle section the participants settle on an activity for which a set of conversational rules are negotiated. Here the level of interaction is at its highest.

4. The participants form a working relationship with commitment.

5. Participants signal to each other confirmation that the conversation is over and to acknowledge that it is satisfactorily concluded.

5.50 Analysis of the *First Steps* Dialogue

As with the *One-to-One* dialogue, the designer was a participant of the *First Steps* dialogue (Appendix B) so it was necessary to analyse it without preconceptions. The first dialogue represents an uninterrupted conversation between two people, to gain an understanding of multi-participatory conversation, it is now necessary to see how the analysis works in an environment in which more than two participants are involved. In this context there is a need to describe where each conversation begins and ends, who is talking to whom, how many conversations are taking place at any one time, and how the conversations are intertwined. A method is developed to represent the structure of a conversation as a linear diagram which also enables a more rigorous definition of some central concepts.

5.51 The Diagram

The diagram is produced by creating a node for each message and arranging these in a vertical linear path corresponding to the order they appear on the screen. Each participant is given a letter, which is attached to each node representing the author of the

message. Starting at the beginning of the dialogue each message is considered in turn. Where a message is understood as a continuation or a direct response to some previous message (or messages) a link is created between the nodes. For example, in the diagram extract shown in Figure 5-1, *Kiwi's* message at 014 is considered to be a direct response to *Orange* at 013, which is considered to be a continuation of 011 and 010. *Kiwi's* message at node 012 is considered to be a direct response to *Apple's* message at node 009.

Ap 009 *Apple: in fact i love to eat*
Or 010 *Orange: they give a lot of energy but don't make you hyper*
Or 011 *Orange: and are very good for ya*
Ki 012 *Kiwi: so do i! i only ate 1 time today or yesturday and that was like 3 pierogies*
Or 013 *Orange: you can get um at health stores*
Ki 014 *Kiwi: thats cool! i will have to try that!*

Figure 5-1: A Comparison of Dialogue Text and Diagram

The purpose of this technique is to identify key structural elements of dialogues in general, rather than present accurate representations of particular dialogues. It is not always possible to link a message into a conversation. Some messages start a new topic which is not followed up, some are an open question to all participants and some are simply difficult to understand in the context of the ongoing dialogue. The judgements made about to which message a link should be made to, are subjective to the designer, but, although there are a few that might be open to interpretation, it is believed these would not greatly effect the outcome. Appendix D shows the decision made for the first thirty two messages of the *First Steps* should the reader wish to examine them. Other information in the original diagram includes a number to indicate how many people are in the room at any particular time, a horizontal line to indicate the point at which the whole group is booted from the multi-user server and some written observations and comments.

The patterns created by the directional links enabled the identification of dialogue features. Figure 5-2 shows three examples of the kind of patterns that can be observed. Model 1 shows one-to-one conversation. Here the participants exchange messages almost tit for tat. Model 2 shows two parallel one-to-one conversations taking place, i.e. A and C, B and D. Model 3 shows a multi-participant conversation. Here the links appear less regular and each participant is interacting with each of the others within the group.

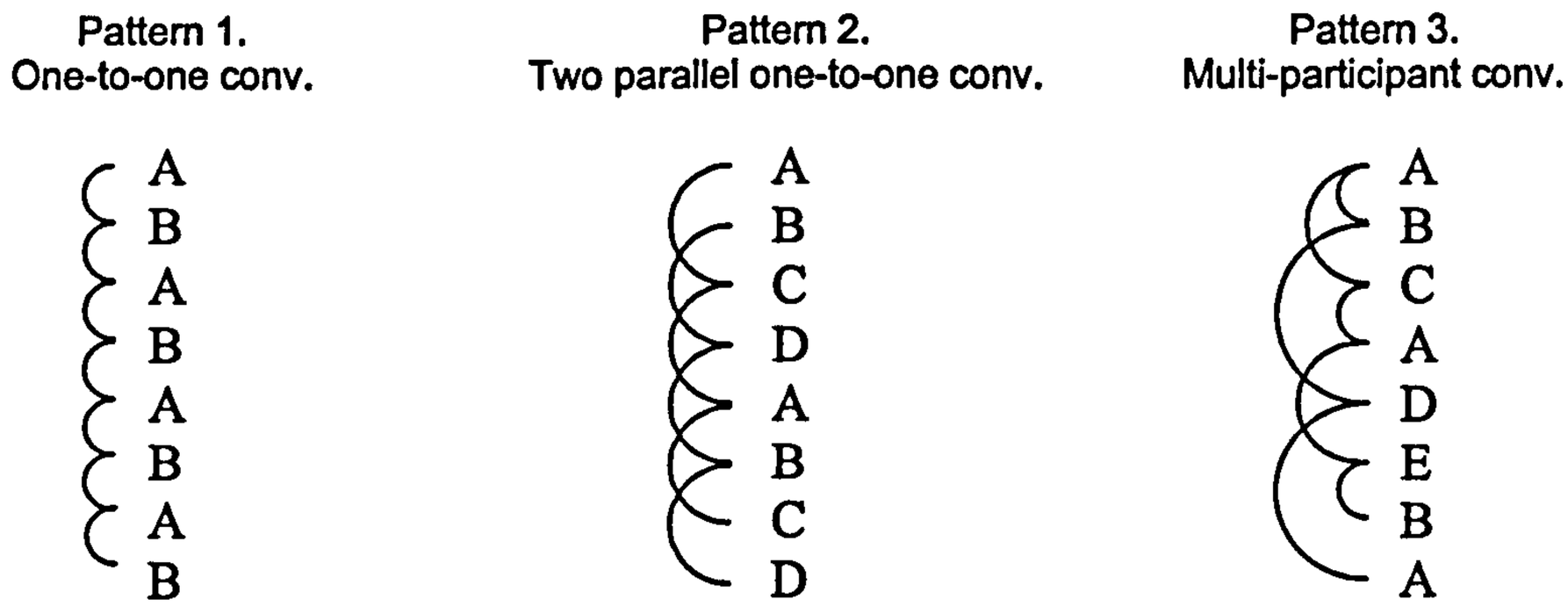


Figure 5-2: Three Examples of the Kind of Patterns that can be Observed in the Diagram

By using this diagrammatic technique we can better understand how many people are involved in a particular conversation, how long a conversation has lasted, how many conversations are going on simultaneously and which participants are not in a group.

5.52 A Third Participant Joins a Group of Two

When there are only two people in the MUD room, the links are easy to follow. When there are three or more participants present, the patterns of conversation are very different. From our first analysis we have some understanding of how two people interact and this raises the question of what happens when a third person enters the room and wishes to join a group of two for conversation?

Two main ways in which a participant may join a group are identified. She may either wait until a present member directs a message at her, such as a greeting, or she may intervene in the present group's discussion with a message such as "what's up everybody?" and await a response. There is no identifiable initiation ceremony to join these social groups [as in Morris 67]; the observation is made by the designer that the newcomer usually has to be brought into the group by one or more of the present members in order to be accepted. If the newcomer does not receive an invitation into the group then he or she is likely to leave the room completely.

It is reasonable to assume that once accepted into a group, the newcomer is expected to participate in the conversation. Two occasions in the data when a third person joins a group of two are considered. Figure 5-3 shows that the flow of conversation is very different in the two examples. Why might this be?

In the first example, *Apple* tries to join *Kiwi* and *Orange*. The latter two are already engaged in conversation when *Apple* arrives and he is brought into the group by *Kiwi* who continues to include him in the conversation. In the next three consecutive messages *Kiwi* continues her previous conversation, responding to *Orange* and then responding to *Apple* in consecutive messages. This pattern continues for the next eighteen messages. The diagram shows that when *Apple* enters the room he responds only to *Kiwi's* messages, not *Orange's*. If the paths of conversation are followed at this point in the diagram, we see that *Kiwi* is really having two parallel conversations, one with *Orange* and one with *Apple*.

In the second example, *Mango* tries to join *Apple* and *Coconut*. *Mango* takes the initiative in this case and intervenes, inviting the other two to include her in the group.

She is given an initial greeting by one of the other two group members but they immediately return to their conversation and she finds no major role to play. The reader will observe the continuous stream of links that represent the conversation between *Apple* and *Coconut* in occurrence two (Figure 5-3). Mango submits messages at a much lower frequency then the other two participants and, despite her persistence, does not manage to attract conversation her way. This is a very different behaviour. So why are the behaviours different? One possible explanation is that at message 169 in the dialogue all participants are booted from the server. Before that event *Melon* is trying to get involved in a conversation. He receives little response from *Apple* but *Coconut* addresses *Melon* on several occasions (the situation is, thus, similar to the first example). This would seem to eliminate the possibility that *Apple* and *Coconut* were unwilling to include a newcomer in their group.

Link	Number	Participant	Message
	001		*Kiwi: anything that comes to mind really!*
	002		*Kiwi: Right now diet pills*
	003		*Orange: get some gensing*
	004		*Apple: thats something I dont know much about*
	005		*Orange: pills*
?	006		*Kiwi: um well i stole these*
	007		*Kiwi. neither do i Apple*
	008		*Kiwi: its all good*
	009		*Apple: in fact I love to eat*
	010		*Orange: they give alot of energy but don't make you hyper*
	011		*Orange: and are very good for ya*
	012		*Kiwi: so do I! I only ate 1 time today or yesterday and that was like 3 pierogies*
	013		*Orange: you can get um at health stores*
	014		*Kiwi: thats cool! i will have to by that!*
	015		*Orange: i got some they are cool*
	016		*Apple: whats pierogies???*
	017		*Kiwi: I am not fat though! you understand???*
	018		*Orange: yes*
	019		*Apple: so what if you are fat*
	020		*Kiwi: they are potatoe pockets filled with like mashed potatoes! they are really good*

Occurrence 1

	173		*Coconut: also mine*
	174		*Apple: how did u cal for Apple in the list of rooms window?*
	175		*Apple: i couldn't find the dust room either where did that go?*
	176		*Mango: whats up everybody?*
	177		*Coconut: i created a room and named calin for Apple*
	178		*Coconut: by the command /n*
	179		*Apple: hey Mango*
?	180		*Mango: me too*
?	181		*Coconut: ithink it was died*
	182		*Apple: is it easy to create a room ??*
	183		*Coconut: I will creat one now*
	184		*Apple: wicked*
	185		*Apple: that a nice tip if you loss some one*
	186		*Mango: that kicks*
	187		*Apple: you must be well practiced at this chat stuff*
	188		*Coconut: that was easy no*
	189		*Apple: what other MUDs do you use*
	190		*Mango: this is only my second time*
	191		*Coconut: just this one, mostly!*
	192		*Apple: whys htat?*

Occurrence 2

Figure 5-3: Two Examples of a Third Joining a Group of Two

Apple and *Coconut* are engaged in a one-to-one conversation when they are booted from the MUD server. What happens next seems very significant. *Apple* and *Coconut* not only log back in immediately but they also seek each other in order to continue the conversation. This is initially frustrated because the room that they were previously in had been created by *Kiwi* and therefore ceased to exist when the group were booted. *Coconut* therefore creates a new room and names it *Calling for Apple* with the hope that *Apple* will log back in, see and enter the room, and find *Coconut* in order to resume the conversation. *Coconut's* plan is successful. *Coconut* and *Apple* continue their interaction in the new room. This experience serves to confirm and enhance the bond between them. They become so engaged with each other that by the time *Mango* enters the room she finds it difficult to become accepted within their conversation. So, it would seem that the sense of engagement exists at various levels of commitment and if it is sufficiently high, it may provide a barrier to a newcomer when trying to become involved.

5.53 Newcomers to a Group of Three or More

The major aspects of the situation when a newcomer joins a group of more than two have been covered in previous examples, but a few useful observations can be made. In groups of four or more there is sometimes a different mechanism to initiate a conversation. As well as exploring a number of potential topics, participants will occasionally ask a question that is clearly directed to all group members, such as "how old are you all?" (message 201, Appendix B). This is similar to the *Start* dialogue in *One-to-One*, but addresses a wider audience. If all the participants respond, this can serve to bring together a group as a whole and initiate a conversation which, in turn, creates a new group bond. From there on the number of conversations that are taking place will vary and the number of people involved in those conversations will also vary.

This model will tend to hold until the next open question or until the number of people in the room reduces to three.

Specialised language can create a further barrier to a newcomer. When *Cherry* enters the room in the *First Steps* dialogue, two of the group members are using a lot of acronyms in their messages. She indicates that she feels excluded from to the group with the message, "I need to learn the lingo" (message 067, Appendix B). The use of expressions such as "LOL" causes *Cherry* to feel, temporarily, excluded from the rest of the group.

5.54 Leaving the Group

How does the process of negotiating an exit work in a group of more than two? Figure 5-4 illustrates the typical situation. The pattern is that a participant gets into a position where she is not directly involved in the conversation and then signals to the group an intention to leave. This is not sufficient, however, and is rarely the last message that she submits. The person leaving needs the same sense of completion that was identified in the *One-to-One* dialogue. Half way through the *First-Steps* dialogue, *Kiwi* signals that he intends to leave (message 108, Appendix B). All of the participants, except *Apple*, respond by saying "good-bye". *Kiwi* does not leave at this point, but appears to wait for *Apple* to respond. This suggests that the departure from a group requires the recognition of all group members. While the general background conversation continues, it is interleaved with every group member saying good bye to *Kiwi*.

To underline the significance of this leaving protocol we can look at the departure of *Cherry*. When *Cherry* signals that he is leaving, there is a response from only one of the three members present. Clearly, *Cherry* does not feel that her departure from the

group has been acknowledged properly, i.e. "doesn't anyone else care?" (message 146, Appendix B). The others then respond and the obligation to *Cherry* is satisfied. Only at this point does *Cherry* leave.

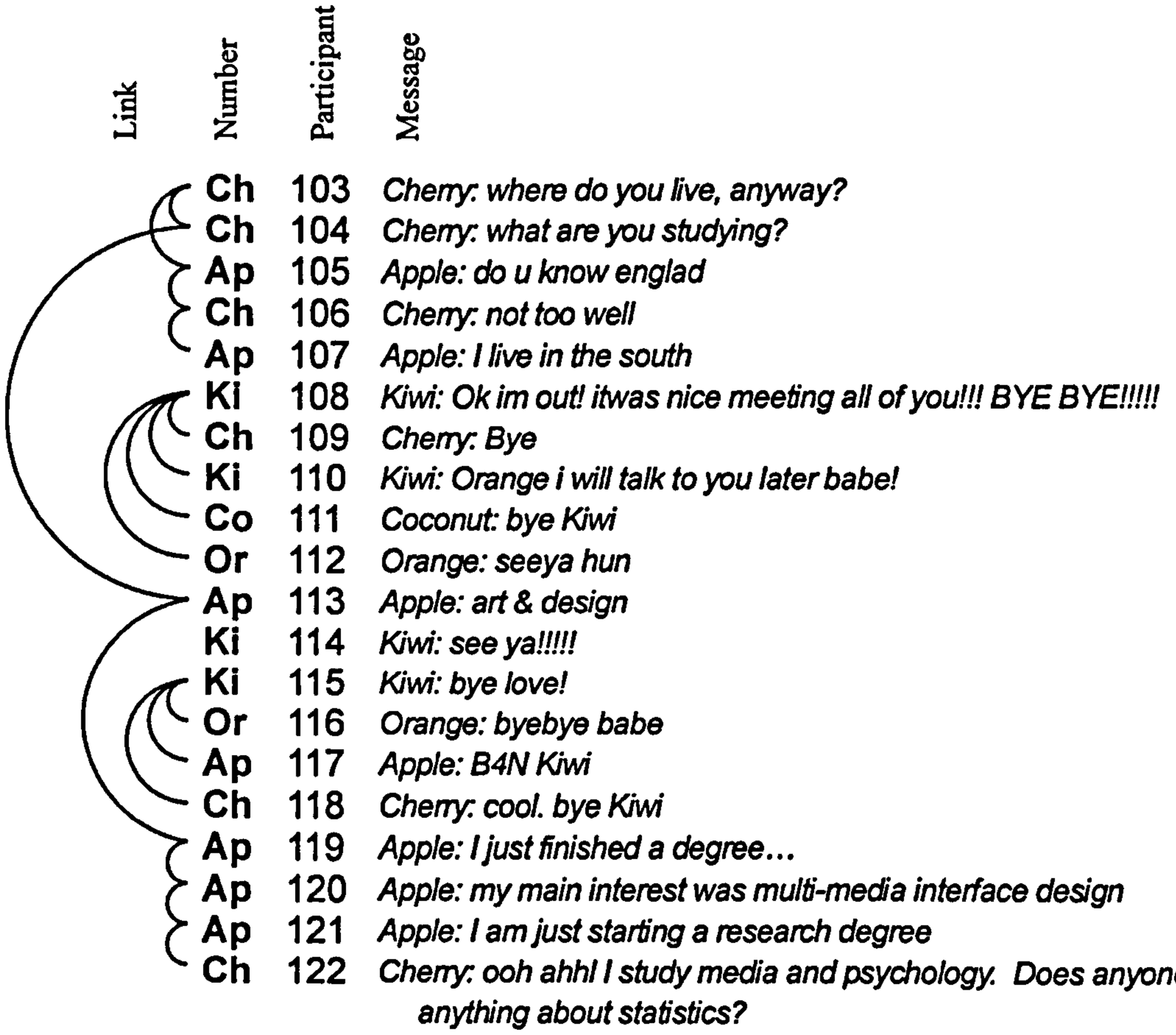


Figure 5-4: Leaving a Group

5.55 Summary of Main Observations on *First Steps*

1. A newcomer joins a social group by either receiving an invitation from at least one existing member, or by taking the initiative to intervene in conversation.

2. There is a sense of commitment to the dynamic of a conversation in a larger group, which can create a barrier to newcomers wishing in the conversation.

3. Specialised language can create a barrier to a novice user.

4. In multi-participant dialogue one way of initiating new topics and conversation is submitting an open question, that is clearly directed to all group members, e.g. "where are you all from?".

5. The analysis suggest that participants wait for some acknowledgement from group members before departing.

5.60 Conclusions Of The Study

Having investigated several MUDs dialogues, the conclusion is drawn that there are two main aspects of virtual world social communication which would benefit from the additional of a NVC (non verbal communication). First, NVC could complement devices by conveying actions that the participants employ to augment discourse, i.e. acronyms, punctuation marks and performative words. For example we could accompany the LOL acronym with a gesture that indicates laughter; a message containing a question mark could be accompanied with an appropriate gesture; and a waving gesture would accompany words such as *wave*. Second, NVC could support the three stages of social communication (i.e. start, middle and end). For example to explicitly express an interest in joining a group, to visually illustrate who has group membership and who has not, and to signal the imminent disengagement from the conversation.

One particular mechanism, which may not have been adequately covered by this analysis, concerns the transition from one-to-one conversations into genuine multi-participant conversations. We saw how some three-participant conversations were, in effect, two one-to-one conversations in parallel but not all conversations can be described in such terms. One mechanism that is identified is the open question, e.g. "how old are you all?", but it is suspected that other mechanisms exist. More data is required before this line of exploration can be pursued, however for the purposes of the larger design activity a sufficient understanding of the social dynamics has been achieved, at least for the time being.

5.70 Summary

This chapter has described a study designed to gain an understanding of how the development of NVC might support group social interaction and complement verbal discourse in AVWs. Of six dialogues collected, two dialogues were selected and subjected to analysis. The *One-to-One* dialogue presented a situation in which two participants were interacting. Analysis of this dialogue enabled a base line understanding of what happens when two participants interact with each other. Analysis of the *First Steps* dialogue enabled us to increase this understanding of social interaction by looking at multiple participants interacting. There a linear diagram was constructed to help identify the various conversations within the group. From the analysis of the dialogues an understanding of social dynamics of virtual worlds has been acquired and we have identified two features of social communication where NVC would be useful: (a) to supplement text devices that the participants employ to augment discourse by conveying actions and expressions, i.e. acronyms, punctuation marks and performative words and (b) to support behaviours during the three stages of interaction (i.e. start, middle and end).

Chapter 6:

Modelling Social Interaction

Chapter Overview

The chapter describes the second of two studies conducted as background research. In this study a model of social interaction in AVWs (avatar virtual worlds) is attempted, the aim here is to elaborate a number of states and behaviours that are likely to occur during the various stages of a conversation. To develop the model, a diagrammatic technique is borrowed from the field of Artificial Intelligence and adapted. Although the model is informed by knowledge gained from the analysis of the MUD dialogues in the first study, it is developed independently from discourse content. Another important difference is that it adopts a scenario-based approach which takes a user perspective. The chapter also presents practical work developed to explore parts of the model and help understand how it might be applied to practice.

6.00 The Second Study

The work documented within this chapter is further background research that aims to develop a model of social interaction based on unspoken behaviours. The model aims to increase understanding of what needs to be modelled, i.e. getting a better comprehension of what a conversation is within a social context. While also seeking to further understand social dynamics, this study is different to the objective analysis of the first in that it attempts to adopt a user perspective. It seeks to reduce a seemingly infinite number of possible actions and reactions of a participant during social interaction, to a finite and manageable set of states and behaviours that are necessary to achieve social interaction. If this can be achieved then a list of useful gestures and expressions that complement these states and behaviours can be drawn up and developed through creative practice. Knowledge gained from the previous study is applied as well as from the background reading discussed within Chapter 2.

To build the model a technique known as State-space Search is borrowed from the field of Artificial Intelligence, and adapted. During its early days, State-space Search Technique was developed to represent problems that involved a potentially large number of possible static *states* with a finite number of possible transitions between them. To represent a problem, the initial state is identified and the conditions under which a good state is to be recognised are described. One early application was in representing the possible moves of computerised opponents with two player zero-sum games such as chess. Through the use of a graph diagram the game can be represented to determine all the possible states of the board and operations between them. This seems to be a good technique to use to create a model of social interaction, which can also be thought of as a game in the sense that participants execute behaviours on a move for move basis with the aim of initiating meaningful conversation, and carrying it through to point of completion where the nested obligations have mutually been met. In additional to this, the first study showed how conversation can be understood as three states which are effected by specific social behaviours.

This approach is eventually found to be limiting, but it is carefully documented within the thesis for three reasons. Firstly, despite its failings it does lead to the identification of a number of states and behaviours that would benefit from a gestural complement. Secondly, it enables the break down of a complex problem into manageable parts. Lastly, a useful rapid prototyping technique is devised to test theoretical ideas and identify practical problems before considerable time is invested in constructing 3D implementations. This also enabled a number of important practical issues to be identified.

6.10 The State-Space Search Technique

State-space searches have been used since the early days of artificial intelligence, usually to build goal directed systems. The usual application domain is to represent systems that exhibit the characteristics we associate with intelligence in human behaviour. In particular learning, reasoning and problem solving. There are three main ingredients in the state-space (a game of chess is used as an example):

1. A starting state e.g. the initial state of the chess board
2. Conditions to recognise a final state, e.g. the rule for detecting checkmate in chess
3. A set of operations that can be applied to change the current state of the problem, e.g. the legal moves in chess.

To help develop a state-space search system there is a useful diagrammatic technique that represents the game. For each possible state of the board, a node is drawn on a graph and the operations between them are represented as arcs. For example, the start state of the chess board would be the first node on the graph. Each of white's possible first moves would then be an arc connecting this node to a new state on the board. Each of black's legal responses to each of white's first moves are thought of as operations that lead to new nodes that change the state of the board, and so on. If all the different states are drawn out then the shortest route to checkmate can be determined within the confines of the legal operations. It is this graph technique, sometimes known as a g*ame tree* when its represents two participants, that is adapted to build a model of social interaction.

Nodes are drawn to represent AVW participants at various states of social interaction, e.g. looking for conversation. The operations are the possible behaviours that can be performed, and are drawn on to the diagram as arcs, e.g. make an offer of conversation.

The arcs lead to new nodes that represent new states, e.g. being offered conversation. The various appropriate behaviours from these new nodes are drawn on as arcs, which lead to new states, and so on. Sometimes a behaviour will lead to a state that has already been identified resulting in a loop within the diagram. The behaviours are determined by the particular objective of a participant at each node and confined to those that are appropriate within the social context.

It is the process of drawing out the graph diagram that helps clarify states and behaviours that are likely to be useful at each of the various stages of social interaction. For example, Figure 6-1 shows the diagram created for two non group members who may be looking to become involved in conversation with one another. It was created by drawing a node to represent a participant (A) who is uninvolved in conversation. A number of likely states and behaviours which A might convey to a second participant (B) are considered and drawn onto the diagrams as arcs (e.g. ask B for conversation). At the end of each arc a node is drawn to represent B who is required to respond. For each of these nodes, likely subsequent reactions are drawn as arcs (e.g. accept and reject offer), and so on. When each of the states and behaviours drawn out can be followed to an outcome (i.e. no social interaction or to move onto the next stage of conversation), the diagram is finished and the states and behaviours have been derived. For further information on Expert Systems, state-space search technique and its application to two player games, see [Jackson 86] and [Barr & Feigenbaum 86].

6.20 Applying the State-Space Diagrammatic Technique

The model is first developed for two participants interacting with each other. If it can be made to work, then a third person can be introduced later in order to begin to model group conversation. The situation is considered where a non-group member, who is

present in a 3D AVW, is hoping to have a meaningful conversation with another non-group member. The model developed operates at two levels which are named *top* and *second*. At the *top* level of the diagram, shown in Figure 6-0, a node is drawn for each of the main stages of social interaction. These stages, shown as circles, represent the three consecutive stages, i.e. *start*, *middle* and *end*. In the diagram, the nodes are named *single*, *conversation* and *parting* respectively. The *single* node (S), represents the participant at the first stage, who is looking to become involved in conversation with another individual. The *conversation* node (C), represents the middle stage of an interaction where he is engaged in dialogue. The parting node (P), represents the *end* stage where the participant negotiates an end to the conversation in a manner that addresses the various obligations. To help readability of the diagram a letter is drawn on to each of the nodes which corresponds to the first letter of their name.

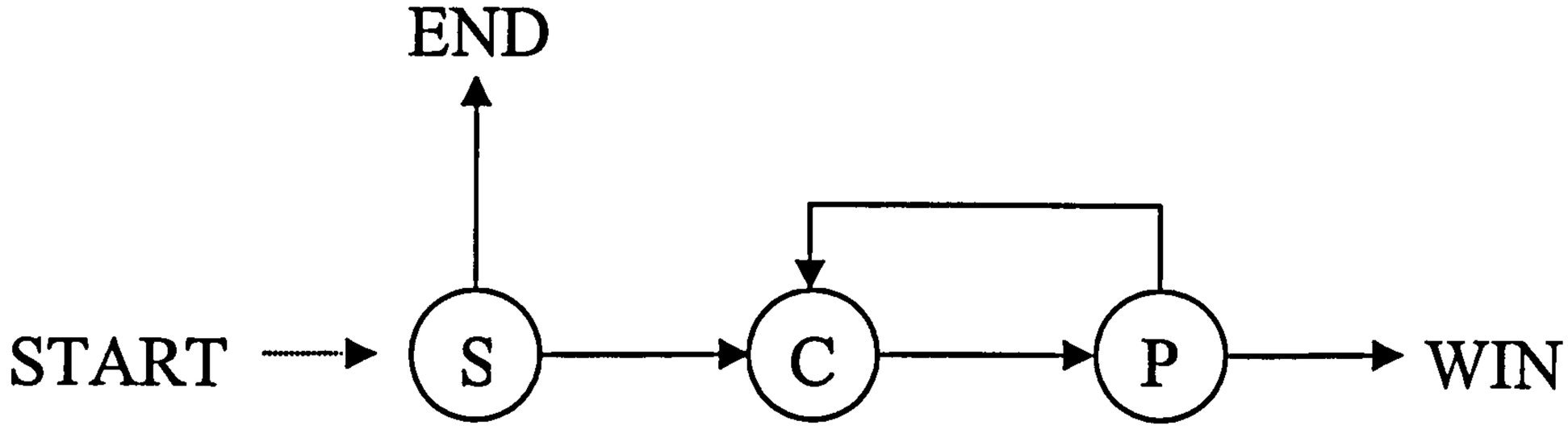


Figure 6-0: Top Level State-space Diagram of Social Interaction for Two Participants

The diagram, which is perceived as a game, has two possible outcomes, i.e. *win* and *end*. The objective is for the participant to start at S, and work through to P, where an end to the conversation can mutually be agreed, i.e. win. The *end* outcome will occur if the participant is unsuccessful at reaching C from S, which will occur if a conversation offer is rejected. While at C, the participant will at some point ask, or be asked to move on to P to negotiate an end to the conversation. If for whatever reason the recipient participant is not ready, the interaction will return to C. This event is represented on the diagram as a backwards loop.

Each of the three stages is broken down into separate graph diagrams in order to determine the necessary behaviours to achieve that stage of the interaction. These, second level diagrams, are discussed in the following sections that lead up to 6.30.

6.21 The Single Node: Initiating Social Interaction

This section of the chapter explains the second level diagram for S, which is shown in Figure 6-1. At S, the participant attempts to initiate a social interaction and if successful, will move onto C. If unsuccessful then the participant will arrive at the 'end' outcome. It is necessary to represent the state of the second participant on the second level diagram. The participants are referred to as A and B and are represented with a circle and a square respectively. The diagram is drawn from A's perspective, however their positions are interchangeable. The arcs represent the possible states and behaviours from each node and indicate the resulting state or outcome.

The start node represents A, who is present in the virtual world and looking to become involved in conversation with B who is nearby. The arcs leading from the first node represent four different states and behaviours that A can convey or perform. The first arc represents the behaviour of placing an offer of conversation to B (arc marked *Ask*). The second and third arcs are states that are informed by the work of Vilhjálmsson which was discussed within Chapter 2 [Vilhjálmsson 97]. Within his *BodyChat* system, participants are able to visually show whether they are available or unavailable for conversation. This is built on and the states are represented with arcs marked *Ava* and *Unav*. The final behaviour is a neutral mode which feeds back into itself. Eventually one of the other behaviours will be forced due to human nature which will only sustain such a position for a limited period of time.

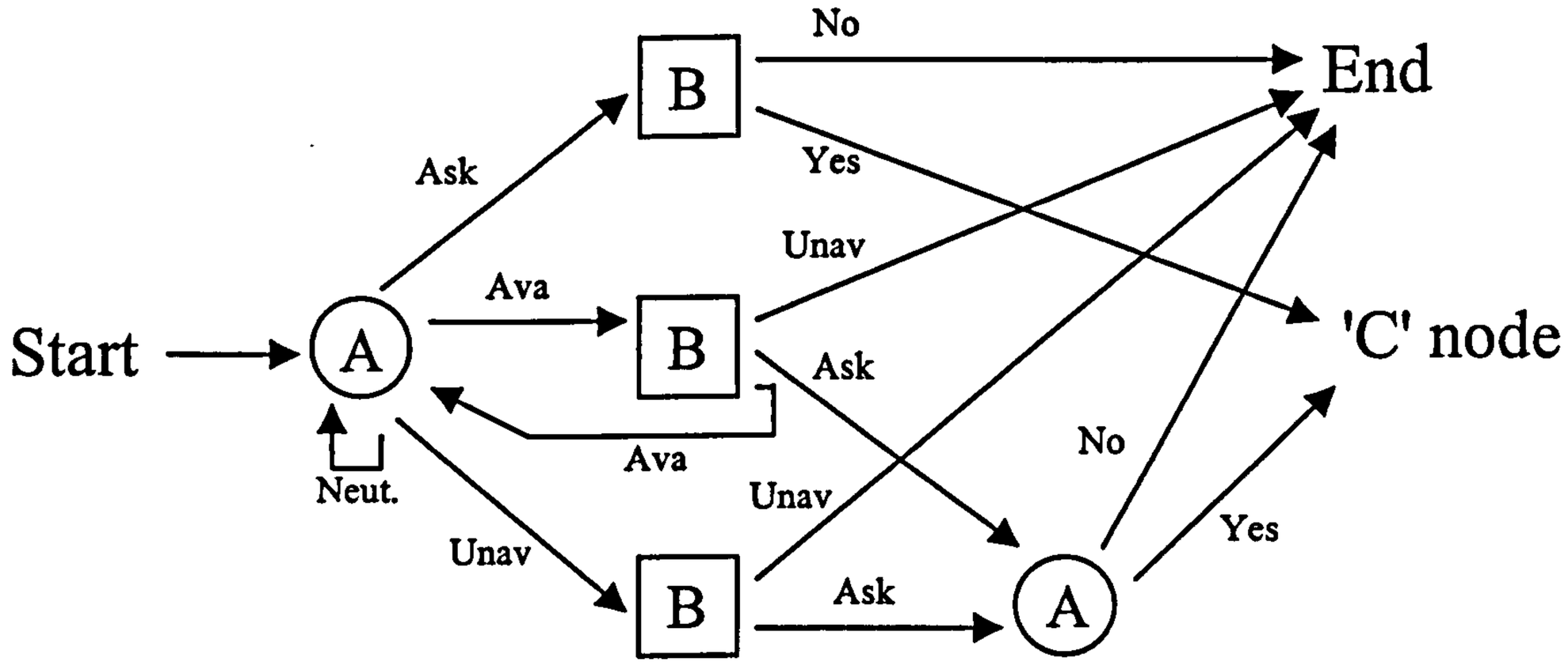


Figure 6-1: Second Level State-space Diagram for Two Participants at the Starting Node (S)

Let us now consider the first course of action that A can perform on the diagram, i.e. to make an offer of conversation. This behaviour forces B into a state where she must decide whether or not to accept the offer. Accepting the offer, shown on the diagram by the arc marked *yes*, leads to the successful continuation on to the conversation node on the *top* level of the model. A rejection, shown on the diagram with the arc marked *no*, leads to the immediate end to the social interaction. Taking no action is also considered to be a sign of rejection.

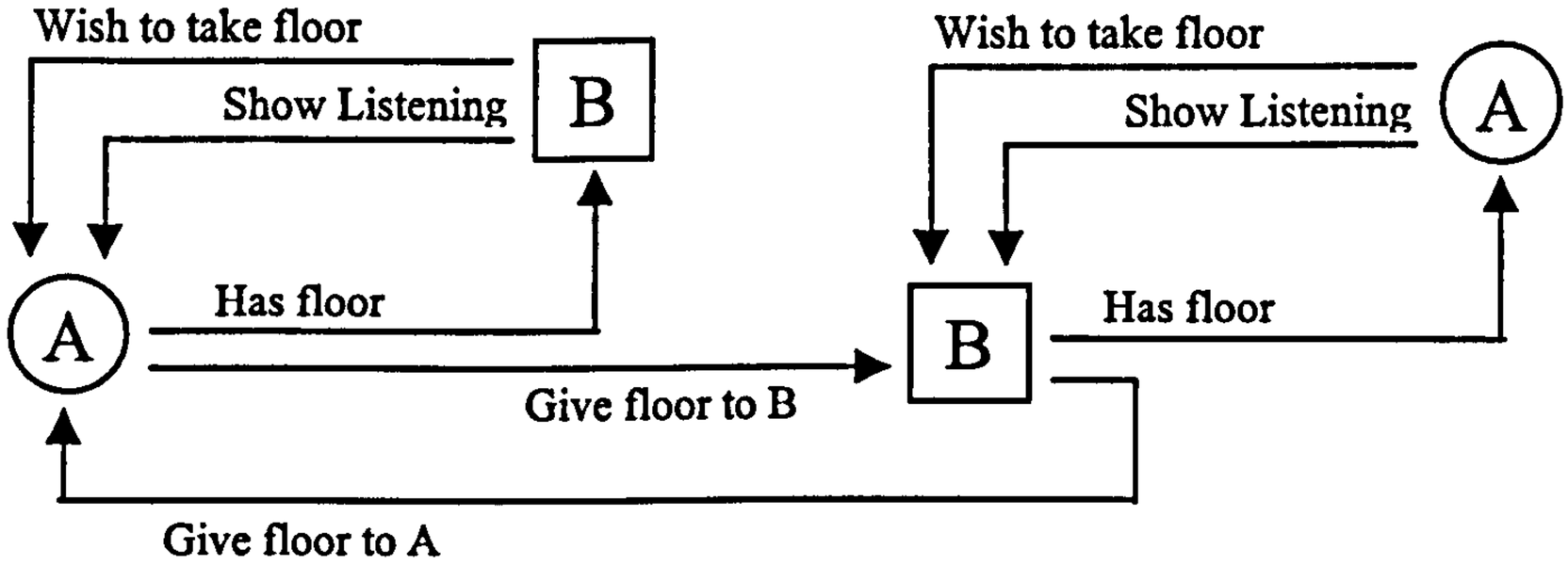
The second course of action is for A at the left-most node to indicate that he is available for conversation. This forces B into a state where there are three possible replies. The first is to indicate that she is unavailable for conversation, therefore ending the interaction. The second is to take the initiative to approach A and ask for conversation. A can then either accept or reject this offer, leading to either the end of the interaction or successful progression onto the *conversation* node. The final arc is B responding with a signal that she is also available for conversation. This is intriguing as it creates a

looped scenario where both participants are showing availability. Eventually human nature will force one of the other behaviours because such a stand off can only be sustained for a limited period of time.

The third course of action is for A at the start node is to indicate that he is unavailable for conversation. B can reply in one of two ways. The first is to do nothing or indicate a state of been unavailable for conversation. This situation will remain until the behaviour of one or both of the participants changes, or until they leave visual contact of each other in which case the interaction will end. The second option is for B to ask for conversation, regardless of the signal given by A that he or she is unavailable. This situation may arise if B has something very important to say to A, or if B feels that A has mistaken her for someone that they do not wish to talk to when this is not the case. A can either accept or reject this offer.

6.22 The Conversation Node: Engaged in Conversation

Figure 6-2 shows the second level diagram for the conversation node. The analysis of the MUD dialogues has shown aspects of social discourse where a NVC can support message content. The diagram however, aims to determine a number of general behaviours that will help regulate the flow of discourse at this stage of the interaction. The diagram is slightly different to the one discussed previously, as it consists of two interconnected loops and there are no outcomes indicated. The objective for participants at this stage of the interaction is to develop a bond and carry the conversation forward to a point where it has meaning. Whether this is achieved or not, one of the participants will at some point initiate a move towards the next stage of the interaction to negotiate an end, i.e. *parting* node.



○ = Person A □ = Person B

Figure 6-2: Second Level State-space Diagram for Two Participants at the Conversation Node (C)

The left-most node represents A who has the floor, i.e. is speaking. There are two behaviours that A can perform at this state. The first is to hold the floor and keep on speaking, shown on the diagram by the arc marked *has floor*. The second is to hand the floor to B, shown on the diagram as *give floor to B*. Following the first option will result in B being in a state where she can reply with one of two behaviours. The first is to show that she is listening, the second is to show a desire to take the floor. Both of these behaviours are feedback information for A who is speaking, therefore arcs are drawn on the diagram back to A at the left most node. The feedback behaviour performed by B, may effect the state of A and subsequently his following behaviour. For example, if B shows a desire to take the floor, then A may surrender the floor prematurely and hand it over to B. Once the floor has been handed over to B, the diagram is repeated but the position of A and B within it is reversed. A can either listen or show a desire to take back the floor, and so on. This situation will hold until B hands the floor back to A.

6.23 The Parting Node: Ending the Interaction

During the analysis of the MUD dialogues, the observation was made that participants offer an, often implicit, suggestion to end conversation sometime before it actually finishes. The analysis also observed that the conversation was not complete until both participants demonstrated mutual agreement that the conversation was over. The diagram, shown in Figure 6-3, is based upon these two observations. It is read from left to right and participants A and B are interchangeable. There are two possible outcomes. The first is that the interaction returns to the *conversation* node. The second is that the interaction moves on to successfully complete the conversation i.e. *win*.

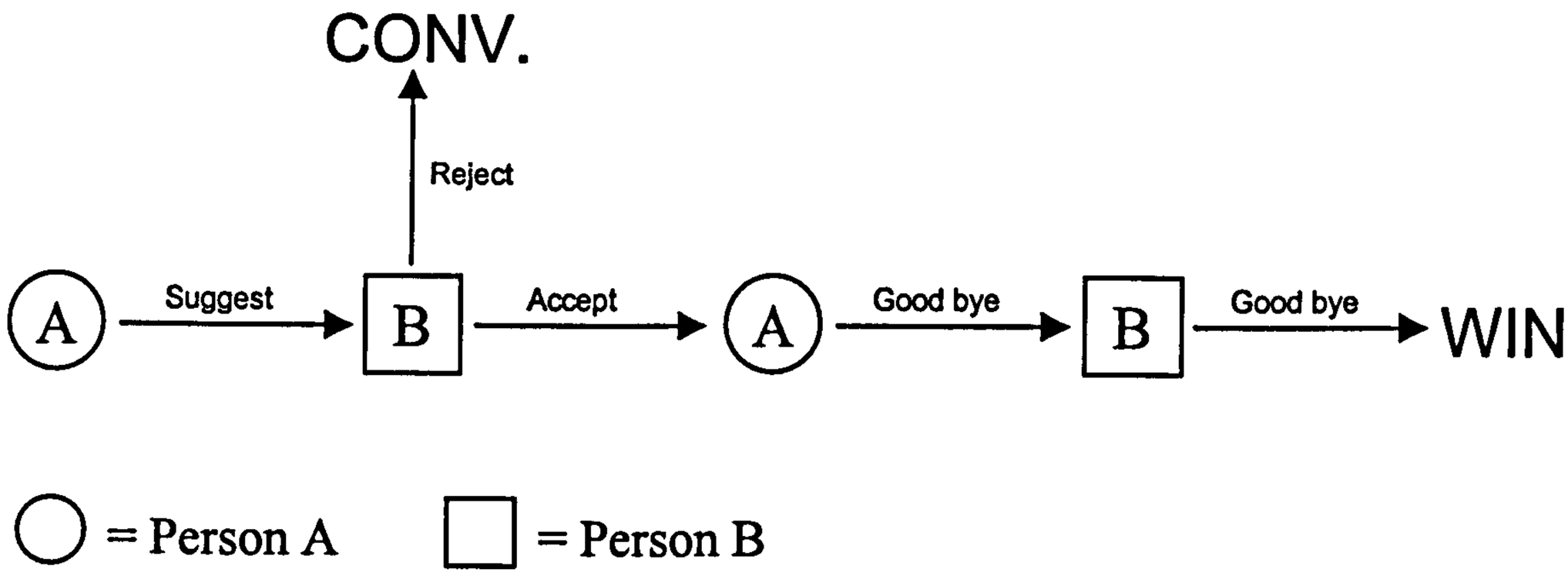


Figure 6-3: Second Level State-space Diagram for Two Participants at the Parting Node (P)

The left most node shows A, who places a suggestion that the conversation should end soon. This is represented with the arc marked *suggest*. B is in a state where she must decide whether to accept or reject the suggestion. If rejected then the conversation will move back to the *conversation* node where, after a time, another more explicit attempt will be made to end conversation. If however, the suggestion is accepted, then after a time A will signal that a mutual agreement has been reached and that the conversation is over. This is represented on the diagram with the arcs marked *Good bye*. If the conversation has come this far then B will almost certainly reply in a similar manner.

The conversation is now over and the participants acknowledge that it is satisfactorily concluded.

6.24 Outcomes

The model constructed for two participants interacting within a social context using state-space diagrammatic technique, has led to a number of behaviours that are likely to occur at different stages of social interaction. These are shown in the table in Figure 6-4.

Single Node (S)	Conversation Node (C)	Parting node (P)
Available for conversation	Show have floor	Wish to end conv. soon
Unavailable for conversation	Show listening	Wish to continue (reject)
Ask for conversation	Show desire to take floor	Agree to end soon (accept)
Neutral	Hand over the floor	Conversation is ended
Accept conversation offer		
Reject conversation offer		

Figure 6-4: Table Showing the Behaviours Necessary to Facilitate Social Interaction Within a Virtual World

One aspect worth documenting at this point is that when it comes to implementing the gestural language, the visual properties may be influenced by a number of factors. For example, the path history within the diagram. If a path needs to be repeatedly followed (e.g. a participant is ignored when showing a desire to take the floor) then it is likely that the accompanying gesture will become more explicit each time. Also, a participants emotional state may effect the gesture properties, e.g. happy or sad. Another factor might be the level of formality of the conversation, e.g. is the conversation partner a friend or a boss at work.

6.30 2D Rapid Prototyping to Explore Model in Practice

In the interests of gaining perspective on aesthetic wholeness, the decision was taken to implement the model developed before attempting to model the third conversant. It was anticipated that reflection on the visual outcome, would enable the identification of additional behaviours. However, rather than move straight to 3D modelling, which is a potentially time consuming task, the model was first explored in 2D. This enables a better understanding of how a 3D version might be constructed without investing considerable time in experimenting with 3D scenes, e.g. what operators are sufficient to convey meaning. The result is a 2D animation that in sequence, plays through each of the possible pathways in the graph diagram.

A simplistic avatar face with limited attributes was created in 2D. Figure 6-5 shows the avatar in a *neutral* state (later in the project we will explore the use of gesture, rather than face expression, i.e. Chapter 8, to convey the same set of behaviours as that discussed here. The pupils, represented with two dots, are located at the centre of the eyes and can be manipulated to achieve a rudimentary gaze direction. Two interchangeable mouth positions are created with the use of a curved line to achieve a neutral and a smile position. The avatar image is then imported into the software *Macromedia Director* and placed onto a simple backdrop that represents a virtual world scene. The avatar is reused to create a second participant, thus the situation modelled in the state-space is recreated graphically. Six facial expressions were created based on the six states and behaviours shown in the leftmost column of the table in Figure 6-4 for this part of the interaction, i.e. available for conversation, unavailable for conversation, ask for conversation, neutral, accept conversation offer and reject conversation offer.

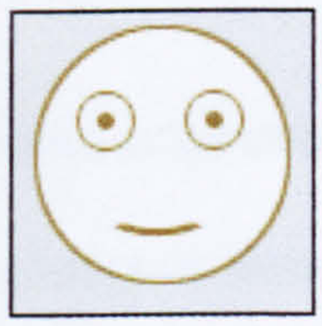


Figure 6-5: Avatar for 2D Rapid Prototype in Neutral State

Each of the ten possible pathways through the second level diagram for the single node are played out using the animated expressions together with spatial changes of each avatar within the space. For example, A moves over to B and smiles to place an offer of conversation. B accepts the offer, and the interaction can successfully move onto the *conversation* node. Next, A moves over to B and smiles to place an offer of conversation. B rejects the offer and the interaction ends, and so on. To aid readability between the diagram and the 2D prototype, text is located above each avatar that indicates their states and the behaviours being performed. The outcome of each path is indicated with flashing text, located below the avatars. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 6-0: 2D Prototype

Description: A 2D animation showing two simplistic avatars, i.e. a circular face with two eyes and a mouth, which represent two AVW participants who attempt to strike up a conversation, i.e. the *start* stage of social interaction.

Demonstrates: The application of the model described Section 6.21 in practice.

Points to note: Notice the use of gaze and distance within the NVC designed.

Instructions: Null

6.31 Relevant Observations

Although the 2D prototype raises no real surprises, there are a few useful outcomes that are worth documenting. Firstly, the designer feels that there enough behaviours implemented to be useful for the *single* part of the conversation. Secondly, in terms of the aesthetic, there appears to be very little difference between the neutral state and conveying that one is available for conversation. It can be assumed that anyone present within the environment is looking to become involved in conversation therefore we can consider these two states as the same. Only five expressions are therefore needed for this part of social interaction.

Thirdly, the designer is convinced that meaning can be adequately conveyed through only a few basic operators, i.e. proxemics, gaze direction and two mouth positions (i.e. neutral and smile). It is suspected that a larger number of operators would enable greater expression to be conveyed however the observation suggests that a 3D model does not need to be complex in its construction.

Fourthly, the observation is made that there may be a difficulty in using gestures and expressions from a predefined library. In this piece, the gaze direction plays an important role in the NVC. Gaze however is dependent upon the spatial relationship between avatars in the world space. A predetermined gesture from a library will not be able to compensate for incorrect orientation of avatars in conversation. Thought needs to be given to this issue if gaze is to be used in the final NVC. One idea might be to give gaze control directly to the participants.

Lastly, when an offer of conversation is made, not only is an appropriate face expression performed, but the avatar making the offer moves into a more intimate

space. Within this prototype the navigation of avatars is control by the animator, enabling the most suitable positioning to be achieved. Within virtual worlds the navigation difficulties make such manoeuvring a difficult task. The question must be addressed, how can participants easily navigate their avatar into an appropriate space to visually make an offer of conversation?

6.40 The 3D Implementation

The 2D prototype provided some useful insights to assist the construction of a 3D implementation of the state-space model for two people interacting at the *single* node. For example, there is no need to develop avatars with lots of operators in order to adequately convey meaning, with proxemics, mouth and eye (gaze) movements likely to be adequate. As with the 2D prototype, all ten possible paths on the diagram (see Section 6.21) are animated using two avatars within a 3D scene. In order to gain an insight into the user's perspective, an element of interactivity is added to give the viewer control of participant A's actions at each node in the graph diagram. The behaviours displayed by participant B are controlled by the computer on a random basis. Particular issues explored in the piece are modelling a 3D avatar head and face, exploring the visual qualities of facial expressions, interface design, navigation, camera positions and the relationship between participant and avatar.

To construct the piece a simplistic 3D scene is created consisting of a ground plane and two female avatar heads that are naturalistic in style. Figure 6-6 shows one of the avatars. To help distinguish between the two avatars there is a difference in hair colour, i.e. red (viewer's) and yellow (computer). They have a number of operators through which facial expressions are achieved, i.e. direction of face, head rotation, raising of eye brows, eye movements and two interchangeable mouth positions (neutral and smile).

To help visually distinguish these features they are coloured. Using the models, each of the ten paths in the diagram are animated and exported in sections as digital movie files, then imported into *Macromedia Director* where they are assembled and the interactive element is added. The visual qualities of the facial animations are informed by Paul Ekman and Wallace Friesen's analysis of facial expressions (see Chapter 2) [Ekman & Friesen 75].

Figure 6-6: Screen Shot of a 3D Avatar Head

The final stage is to design and build a simple interface through which the user controls their avatar. Two field buttons are created that allow the user to switch between the available and unavailable for conversation modes. A text field is added where messages can be compiled, and a send button is included to submit them. Clicking on the computer's avatar at different times also has various functions such as accepting an offer of conversation without verbal confirmation. It is unnecessary to add controls for freely navigating the environment as this feature is not necessary in the piece. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 6-1: Red & Blonde

Description: An interactive 3D animation showing two naturalistic avatar heads which represent participants A and B in the state-space model described in Section 6.21. Control of participant A is given to the reader while participant 'B's behaviour is controlled by the computer on a random basis.

Demonstrates: The theoretical model for the *start* stage of social interaction in practice.

Points to note: There is an information bar running across the top of the piece which displays text that indicates the state and behaviour of each avatar at any given time to help readability between the practical piece and the diagram.

Instructions: Actions can only be taken while the avatars are static, and not while animations are playing. As soon as the piece is launched, the reader can choose between the available or unavailable mode. After a short time B moves into view. The reader has two options at this point which are to do nothing (in which case B may take the initiative to make a conversation offer), or ask for conversation by clicking on B's avatar. B will then ether accept or reject the offer. If B has asked for conversation then the reader must accept or reject the offer. To accept either click the mouse on the computer controlled avatar or type one of the following into the text field and send the message: hi, hello, :) or ok. To reject the offer do nothing or change to unavailable mode.

6.41 Evaluation of 3D Piece

An avatar has been modelled and fitted with a minimal set of operators, that allows a facial expression to be performed for each of the behaviours on the diagram for the *single* stage of social interaction. The piece has begun to develop a NVC, for example to make an offer of conversation: an avatar moves closer, tilts its head backwards, makes eye contact, raises one eye brow and smiles. It is also becoming clear how space will play a role in the NVC developed. For example, in the piece an avatar moves closer into a more intimate position to place an offer of conversation. This leads to an intriguing notion of a *conversation space* which can be understood as a place where avatars gather to converse. This may help to address the problem raised in section 3.20, of scattered avatars in existing AVWs, which results in group members not being visible to one another.

Whilst not fully addressed within the design project, an initial understanding has been gained of the practicality of using text to control an avatar's body language. The technique appears promising and eliminates the need for cumbersome buttons within the

client software interface. Another interface solution to have emerged is the automation of the reader's avatar when navigating into an intimate space to place a conversation offer. The intriguing thing about this is that it occurred almost by accident. As the piece did not allow the reader to freely navigate the avatar about the environment, a method was found to activate the predefined animation, i.e. by clicking on the recipient's avatar. This technique can be applied to groups as well as individuals therefore solving the navigation difficulty. Automation may also be used to address the problem of gaze direction that emerged when evaluating the 2D prototype. As the computer keeps track of avatar positions, it is conceivable that a predefined gesture from a library can be manipulated in real time to achieve correct gaze behaviour.

The participants of an experiment conducted by Anthony Guye-Vuillème (section 2.70), reported that participants felt the need to see their own avatar so that they were aware of what gestures it was performing [Guye-Vuillème 98]. In the interest of developing the relationship between user and avatar, this point has been taken onboard. However, within the piece it has been difficult to find a camera angle that reveals enough of the avatar's faces to show what they are both doing. A different method needs to be found. One idea might be to adopt a similar technique as some home PC computer games. The game *Vampire: The Masquerade* [Nihilistic 00] displays a second, smaller, view of the player's avatar from the front, which is used to indicate the health level. This could be adapted to show the expression being performed within AVWs.

Within Chapter 4, the question was raised about the suitability of facial expression for a screen based NVC. While they are visible enough in this piece it must be appreciated that the piece show only two participants, close up, and no text messages. Their

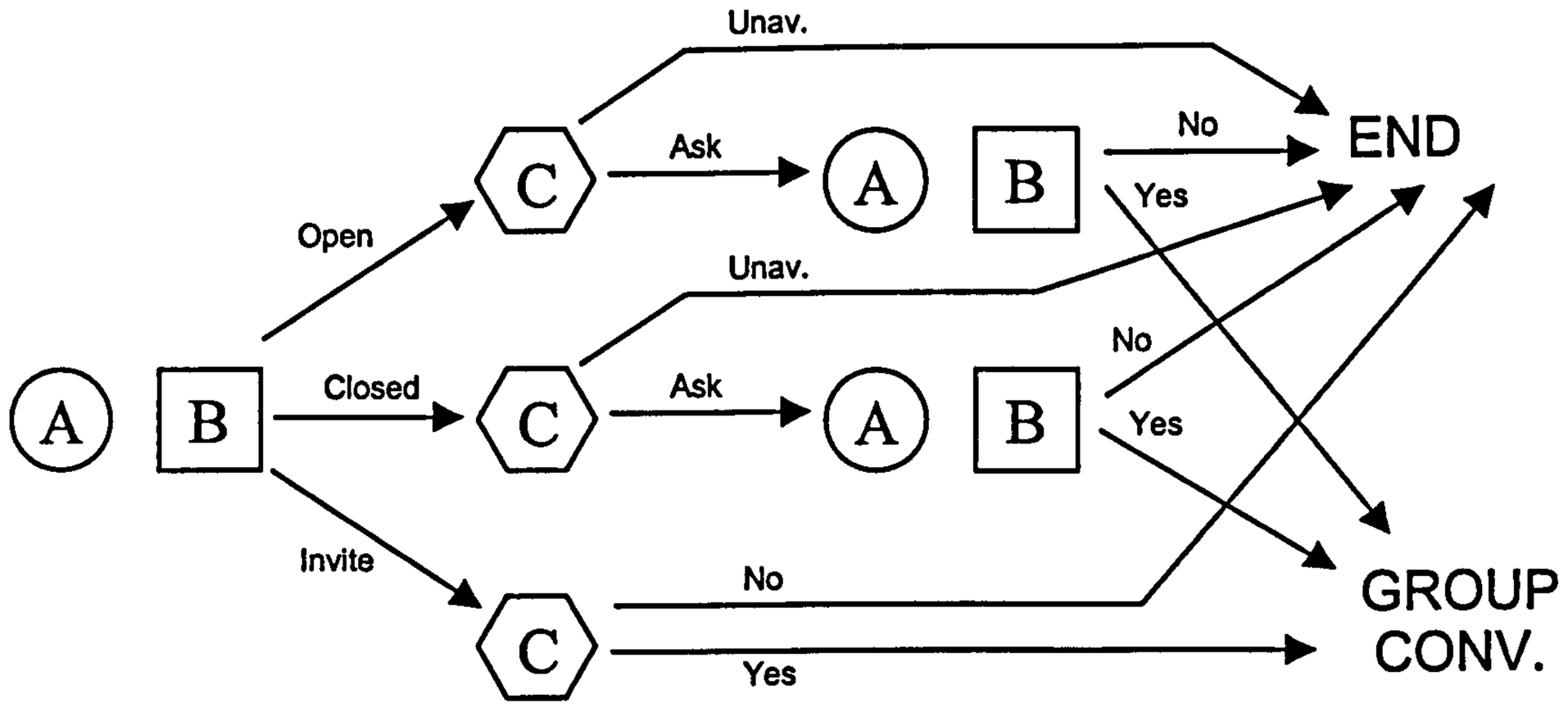
appropriateness for a screen based NVC of this kind therefore remains inconclusive for the time being.

6.50 State-space Diagram for a 3rd Joining a Group of Two

A third participant is introduced into the model in order to help derive some behaviours that may be used during group interaction. The situation is considered where A & B are engaged in conversation at the conversation node and C, the newcomer, has moved into visual proximity. Figure 6-7 shows the state-space diagram. There are two possible outcomes which are 'end' and 'group conversation'.

The left most node represents A & B who are chatting. They are aware of C and can perform one of three behaviours. Firstly, to show that they are open to newcomers which is the group equivalent of being available for conversation. Secondly, to show that they are closed to newcomers which is the group equivalent of being unavailable for conversation. Lastly A & B can invite C to join them.

For the first two behaviours, i.e. open and closed, C can reply in one of two ways. Firstly, to either do nothing or show that he is unavailable which will lead to the *end* outcome. This is shown on the diagram by the arc marked *Unav*. Secondly, he can take the initiative to approach the group and intervene in the conversation. The analysis of MUD dialogue suggests that for a newcomer to be accepted into a group, only one existing member need accept them. Therefore if either or both A and B accept C, then the *yes* arc is followed to the *group conversation* outcome. If neither accept C, then the *end* outcome is reached. It is worth noting that although the diagram repeats itself from the left most node for the two behaviours *open* and *closed*, it is more likely that C will be accepted if he approaches a group that is open.



○ = Person A □ = Person B ⬡ = Person C

Figure 6-7: Second Level State-space Diagram for a Third Participant Joining a Group of Two

The process of drawing the diagram elicited number of behaviours that are likely to be required to enable a third participant to join a group of two. Some of these are similar to those arrived at in two person previous diagrams in this chapter (e.g. ask for conversation). However, it is anticipated that the visual properties of a visual implementation may differ for a group situation, e.g. a newcomer may need to gaze at both existing members in turn. Additionally there needs to be a way for A & B to indicate that they are open or closed to a newcomer and to offer an invitation to join a nearby non group member.

6.51 2D Prototype for a Third Joining a Group of Two

A 2D implementation of the model discussed previously is constructed to further explore it through practice, with the intention of eventually implementing it in 3D. The 2D piece looks and operates in a similar manner to the first prototype except that an overhead view point is used, rather than face on. This is necessary to achieve a circular

formation of avatars so the participants are, spatially, on socially equal terms. The avatar face is reused to create a third participant (newcomer). Each of the possible pathways through the diagram are played out.

As suspected (see 6.50), some of the existing facial expressions created for the first prototype need to be adapted when used within a group scenario. For example when C approaches A & B and asks to be included in conversation, the offer may need to address both of the existing groups members. Even within such a simple implementation, gaze, spatial arrangement and orientation are manipulated accordingly. There are also some additional behaviours that have been created which were not derived by the initial model, for example, C illustrating a desire to become involved in conversation. This behaviour is used before A & B will extend an invitation. While the diagram has provided a way of understanding what behaviours are likely to be useful for a third joining a group of two, this illustrates the benefit of exploring it in practice.

One intriguing visual device used in the prototype is A & B becoming semi-opaque to illustrate that the group is closed to C. The author is not proposing that this is the best solution for such a situation, however the concept of a group language is an intriguing one to consider. From the start of the project is was anticipated that the visual component of communication developed would consist of a set of gestures and expression that could be performed by each avatar. However, it may be useful to extend the visual language designed to include a non verbal group language. This may help to address some of the other shortcomings of existing AVWs discussed in section 3.20 such as, being able to identify what social groups exist and who their members are.

6.52 3D Implementation

A 3D implementation of the model has been attempted based on the 2D version but is unfinished due to encountering several significant practical difficulties. These difficulties are not new but have become serious enough to merit attention before the design can progress. They are related to the spatial arrangement of the group member's avatars and camera language. Due to the absence of a spatial convention, determining certain visual properties such as the most suitable camera angle, the best field of view, distance between and orientation of avatars is a difficult task. This makes the design brief objective of making group members visible to one another a difficult task when using a single camera view.

An early approach was to adopt a first person view. This was limited as too great a field of view had to be used to fit the other group member's avatars on the screen which resulted in an unacceptable amount of image distortion. If any group member's avatar is out of view, it is not possible to see what expression they are performing, or to observe a newcomer's arrival on the blind side. While a third person view is a better option, it is still not possible to view all three avatars faces unless they all stand in a line in front of a camera. In most case this arrangement is most impractical for group conversation. It would seem that a conversation space needs to be designed with definite spatial conventions, and standard view points, so that so that these important practical issues can be addressed and a NVC developed.

6.53 Newcomers to Larger Groups

State-space Diagrams have been attempted for a fourth joining a group of three, and a fifth joining a group of four however, there appear to be no additional behaviours that

can be derived from these. The main factor that will need to be taken into account when considering larger groups is how the visual properties of some of the behaviours derived will be affected by who is addressing whom and the spatial relationship of conversation partners.

6.60 Summary

A way of modelling social interaction within AVWs has been devised using State-space Diagrams. This technique has enabled a number of behaviours to be derived that are likely to be useful during the various stages of one-to-one conversation and for the event of a third joining a group of two. State-space Diagrams have been drawn for newcomers to larger groups, however, no additional behaviours were derived.

Creative pieces of work have been used in order to explore the application of the model. These enabled the designer, to whether or not the set of derived behaviours is appropriate as a means for social interaction. The model was first implemented in 2D in order to gain a better understanding of how a 3D implementations might be constructed. This led to observations such as the need for only five rather than six expressions at the *single* node stage of conversation, which saved time when implementing in 3D.

The 3D work enabled a number of important practical issues to be identified, such as: what is the best field of view to use and how are participants to be made aware of their own avatar's expression? Many of these issues need to be addressed before a body language can be developed as their solutions are likely to effect its design. There are some indications as to how the face may play a role within the body language designed, e.g. raised eyebrows to indicate that a message is a question. The observation was also

made that it would be advantageous for NVC to extend beyond a simple set of gestures for an individual avatar to include group NVC.

When combined with the data gathered from the MUD analysis of the first study, enough requirements have been derived to develop a set of gestures that support various stages of social interaction, and the discourse. The next step is to address the important practical issues raises in this chapter through creative practice, to design a body language.

PART 3: CREATIVE PRACTICE

Chapter 7:

Designing a Space for Multi-modal Communication

Chapter Overview

The work described in this chapter addresses the first of the two tasks set out by the design brief, i.e. in order to create a framework for multi-modal communication, find a solution to make group members visible to one another while integrating the various channels of communication. Through creative practice a concept called the *conversation circle* is developed that gathers together group members within an AVW (avatar virtual world), makes them visible to one another and enables the textual channel to be mapped onto the image. In addition to this, it enables a NVC (non verbal communication) of social groups to be developed based on the notion of place, arrangement of avatars and various scenic properties such as camera position and lighting effects. Firstly, the design task is described in detail, which draws on the design brief and the background research presented in Part 2 of the thesis. Next, some initial ideas and the proposed solution are discussed. Following this, a 3D scene consisting of six avatars is created and used to develop the *conversation circle* concept and design the group NVC. Lastly, the visual properties of the text in the image window are determined.

7.00 Design Task One

The work described in this chapter is about groups, rather than individuals. It addresses the important practical issues that must be resolved before an avatar body language can be developed. These have been defined by the first task of the design brief as the need to (a) make group members visible to one another and, (b) to integrate the various channels of communication, i.e. text, image and audio. In addition to this, the work described aims to develop the concept of a group NVC that was formed during the exploration of the state-space model of social interaction through practical work (see

Chapter 6). The comparative review of AVWs, design brief and background research, suggest a number of aspects of multi-participant interaction that such a language is expected to complement. These are to:

- make social groups visible to external participants,
- show how many groups exist at any one time and where they are located,
- show the number of participants within a group,
- illustrate who the members of each group are,
- make arrivals and departures obvious,
- signal when the group is full (if there is limited capacity)
- signal if the group is closed to newcomers (i.e. private), and
- indicate that a group has dispersed.

The general design method used is to generate and explore a number of potential solutions that achieve the two main goals of the first design task. To help understand the implications of each of their concepts, the ideas are worked through on paper in the form of sketches and short story boards. The idea that meets the greatest number of the above list items is chosen and developed through a series of creative pieces of work. The language developed is independent from design of the individual avatars, and exploits a number of available features such as arrangement of avatars, lighting effects and camera position. While the creative work is primarily concerned with practical issues, the holistic nature of the creative process used deems it necessary to consider technical implications where necessary. This will help to reach a solution that is not lead by the technical aspects, but is feasible within the limitations of existing, and near future, AVW technology.

7.10 Initial Ideas and the Conversation Circle Proposal

In this section the three initial ideas with the most potential are discussed in relation to the requirements of a group NVC (see Section 7.00) as well as the general goals of the design project. The first is to show multiple views of an AVW by dividing the screen into a number of windows that represent each group member, in a similar manner to video conferencing. All group members are simultaneously displayed and text messages appear on all member's screens at the bottom of the author's window. This concept makes group members visible to each other and integrates the text and image channels. In addition to this, arrivals and departures are made obvious by the appearance and disappearance of windows, and it does not require group members to be gathered together therefore the navigation difficulties are avoided. However, these advantages are outweighed by numerous practical problems. For example, it would be difficult to make the group visible as a single entity to external participants, to show who is talking to whom, the 3D space is rendered meaningless during conversation and most significantly, the scattering of text messages about the screen would make following discourse challenging and distract attention form other channels of communication.

The second idea is based on the observation that the 2½D AVWs (avatar virtual worlds) *The Palace* and *Worlds Away*, both break down the environment into smaller locales known as rooms. These rooms can only support a limited number of participants before overcrowding occurs, thus the size of social groups is limited. While this may have originally been designed as a bandwidth solution, it makes sense to create smaller, intimate, and manageable spaces that contain a limited number of users. Based on this, the second idea is to assign areas within a 3D AVW which will be indicated by a shape or pattern on the ground, each one allowing a single group of limited participants to

form within. The avatars would be free to stand anywhere within, and move around, the space. A camera language is proposed that incorporates techniques from traditional film such as the close up, cut and sweeping shot. By automating the camera, the image on each of the group members' monitors, would turn or jump to show an avatar when it speaks, joins or leaves the group. Text messages would be visible above an avatar's head. Group members are therefore made visible when necessary and the text and image channels are integrated. Advantages of this idea are social groups are visible to external participants and it is clear who the members are. The sense of being part of a group for a participant is increased and group identity is strengthened through temporary ownership of the space which it occupies. There are however, disadvantages. It will be difficult to understand who is talking to whom, only one member can speak at a time, participants cannot give implicit feedback while another has the floor, a high frequency of message exchange will result in too much camera movement and participants spatial awareness will be compromised which is likely to result in arrangements of avatars that appear to be unnatural to external participants.

A third idea is the one selected for further development because it can potentially meet all of the criteria set out for a group NVC as listed (see section 7.00), as well as meet the two requirements of the brief for this design task, i.e. to make group members visible to one another and to integrate the various channels of communication. Named the *conversation circle*, it builds on the concept of gathering together group members' avatars but, rather than assigning areas within a world where conversation takes place, groups of limited numbers are able to form anywhere there is room. Only when participants are within a group can they communicate fully with others, thus encouraging their formation. Once formed, a group owns the area it occupies for as long as it exists. The boundaries of this space are defined not by walls and other

physical objects, but by a circular arrangement of avatars facing centre, and by the various social obligations that exist between its members. A circular formation is chosen because it is expandable, facilitates chat and discussion dialogue and, spatially, all members are on socially equal terms. Automation will be used to evenly distribute participants along the circle's circumference onto which they are locked until such time that they wish to leave the group. This retains the circular formation and frees up the keyboard for authoring text messages and controlling avatar body language. In order to increase the sense of ownership of the space, the circle is made off limits to all other participants unless they join. By using a third person perspective, all members are made visible to one another and the textual channel can be integrated by displaying messages around the *conversation circle* which is central on the screen.

The *conversation circle* concept is particularly intriguing because it begins to create a relationship between multi-participant dialogue and the physical space. By creating locales where participants can interact with each other and where social obligations exist, *space* is given social meaning and is therefore transformed into *place*. The concept as a whole also seems to have very few drawbacks, one worth mentioning is that avatar faces are not equally visible. The solution however, contributes much to a group NVC. For example, social groups are visible to external participants who can see where they are located and how many groups exist; the number of members in each group is clear, as is who the individual members are, arrivals and departures are obvious and it is clear when a group has dispersed. In addition to this, the solution does not compromise participant's spatial awareness, the world space integrity is retained, it is possible to show who is talking to whom, feedback can be received while one has the floor, and it is expected that the sense of being within a conversation for individuals will

be enhanced therefore the obligation to contribute to the dialogue will be increased, thus facilitating communication.

The *conversation circle* is developed through a series of creative pieces which specifically address practical issues such as what is the best camera position, what is a comfortable spacing of avatars along the circumference and what are the visual properties of the text. Working creatively allows the designer to get close to the problem from a user perspective and through this, theoretical and interface issues are also addressed such as how are *conversation circles* formed and dispersed?, how do people join and leave?, how do people navigate onto the *conversation circle* parameter?, and how can external participants be discouraged from crossing its boundaries?

7.20 Modelling a Conversation Circle

The first piece of creative work models a *conversation circle* in 3D, with the aim of exploring the following practical questions: what is a comfortable spacing of avatars within a circle?, where should the camera be positioned?, and how can the group's space be further represented without physical objects? The decision is made to limit the number of avatars within a circle to six. If the concept can be proved to work for six, then larger groups can be considered in future work.

A simplistic 3D AVW is created in *3D Studio Max*, containing a number of naturalistic avatars, one of which is referred to as *Red*. Their function within the scene is simply to represent six participants, body language is developed later in the project (see Chapter 8). A camera object is created through which the scene is viewed. In order to find a comfortable spacing of bodies, two avatars are placed onto the circumference of a circle. The diameter is then expanded and reduced until artistic intuition suggests that

PAGE NUMBERS CUT OFF

IN

ORIGINAL

the correct spacing has been achieved. This procedure is repeated for up to six avatars. It is necessary to leave enough room between avatars for a dynamic body language to take place without the risk of contact with neighbouring avatars. In addition to this, enough room needs to be left so that participants' personal space is not invaded, but not so much that intimacy begins to break down. Figure 7-0 shows a screen shot of the spacing for six avatars standing in the *conversation circle*.

Figure 9-0: A Rendered Image of a 3D Scene Containing Six Avatars Within a Conversation Circle

It is decided to use a third person view so that the whole group is visible at a standard FOV, i.e. 50 mm. This also addresses the issue raised by Guye-Vuillème that participants require awareness of what their own avatar is doing (see Section 2.70). To find the best position, the camera was moved about *conversation circles* containing different numbers of participants. The view found allows maximum visibility of the surrounding environment while keeping the group the focus of the image, and at the centre of the screen (Figure 7-0). It also reserves enough space between the group and the edge of the screen for the text messages (the integration of text will be discussed in Section 7.40). As was mentioned in the previous section, one draw back is that not all of the avatar faces are fully visible. In addition to this the faces are relatively small, therefore subtle behaviours such as facial expressions may be hard to make out. This

suggests that the body language that is developed later in the project should focus on hand gesture rather than face expression.

While reviewing rendered images of the scene, the observation was made that the avatar faces were too dark, and this created an image with an impersonal and sinister atmosphere. While experimenting with lighting to give the faces in a *conversation circle* more emphasis, two useful visual effects were inspired. Firstly, during one render, the shadow option with *3D Studio Max* was accidentally turned on. It occurred to the designer that the shadow effect was a good way of helping to visually define the boundaries of the groups space without the use of physical objects. They are therefore included in the conversation circle proposal. While shadows are not yet a feature of existing AVW, they are technically possible. They are computationally demanding so it is proposed that only avatars in social groups will produce them. This will extend their meaning to illustrate that one is part of a group.

The second useful lighting effect is a bright spot on the ground in the centre of the *conversation circle*. This makes it easier to identify groups in the environment, by distinguishing them from other avatars who are not involved in conversation. It also makes social groups a main feature the world, which is appropriate as the primary activity is social interaction. The reader will observe the obvious group in the background of the image of Figure 7-0, and the non-group member who is darkened, indicting absence of social activity. For this lighting effect to work effectively it is necessary to darken the surrounding environment on the monitors of the group members.

7.21 Animating the Model

The scene is animated for the purpose of determining what happens when participants join and leave the group as well as how, and when, the lighting effects appear and disappear. A scenario is imagined and implemented using the scene created. It begins with *Red* as a non-group member, who can see another uninvolved participant in the distance *(Yellow)*. At this point all other avatars are not visible. *Red* approaches *Yellow* and places an offer of conversation which is accepted. This marks the forming of a *conversation circle.* While engaged in message exchange, four newcomers join the group in turn until the *conversation circle* reaches maximum capacity. The participants then leave one at a time, until the *Red* is the only one left at which point the conversation circle ceases to exist. The animation is rendered as a digital movie which shows the scene through *Red's* camera.

It is proposed that a *conversation circle* is formed the moment that the offer of conversation is accepted. It is at this precise moment that the light source fades in while the background is darkened. It has been found that the best camera position for a non group member participant, differs from when a participant is part of a *conversation circle.* The reason for this is that a non group member participant is likely to be seeking conversation and will therefore require a camera angle that is optimised to show the surrounding environment and the social activity within it. Once he or she has joined a *conversation circle* the participant is likely to require a camera angle that is optimised to isolate the group from the general scene. Therefore, in the first situation, a participant's camera is located just above the avatar's eye level, behind the head and facing forwards. When a participants joins a *conversation circle* the camera is automatically raised, pulled further back and angled slightly downwards. Newcomers join the group by moving towards it, when they are near, the circumference expands and the avatars

automatically slide along it to reveal a space, which the newcomer fills. When members depart, their avatar automatically takes a couple of steps backwards, turns around and takes one step forward, at which point the circle circumference decreases and the remaining participants slide along it to achieve even spacing between them. When there are only two participants in the circle and one leaves, the *conversation circle* ceases to exist. The group light fades out, Red's camera returns to its original position and full navigation control is regained. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 7-0: Conversation Circle

Description: A digital animation showing the imagined scenario implemented using a 3D scene containing and six avatars. Viewed through the *Red's* camera, he approaches *Yellow* and places an offer of conversation which is accepted. A further four participants join the group in turn, until the group reaches it maximum capacity of six. The groups members depart one at a time, until the circle is dispersed.

Demonstrates: How the light fades in when a *conversation circle* is formed, how the shadows appear and help define the groups space, camera positions relative to a participants avatar when in both single mode and a *conversation circle*, the spacing of avatars in the conversation circle, how the circle automatically increases and adjusts the avatar positions to allow a newcomer to join, what happens when a group member departs, how the group's light fades out and how a participant camera returns to its original position when the group disperses.

Points to note: On reflection the, conclusion is drawn that the avatars are too close to each one another. More space should be allocated in future implementations. In additional to this the camera positions need to be improved to utilise the compositional potential and reveal as much information about the social activity within the surrounding environment as possible.

Instructions: To begin the animation click the mouse on the avatar in the distance (*yellow*), or click *play*.

7.22 Interface Solutions

Part of the user experience is the interface through which the avatar is controlled. It is therefore necessary to pay attention to such issues as what action a participant takes, to

make an offer of conversation, to accept or reject it, to depart, as well as how a participant navigates into a *conversation circle.* The following actions are proposed to control an avatar and address navigation difficulties. In the same manner as the earlier creative work titled *Red & Blonde* (see Section 6.40), a participant will click the mouse cursor on the avatar of a participant who is a non group member. The computer will automatically navigate their avatar into a position that will form a conversation circle should the recipient accept. The recipient accepts the offer by return clicking on the offering participants avatar, or submitting a message that contains his or her name within it. Rejection is shown by either doing nothing, sending a verbal message, turning or moving away. To join a *conversation circle* that is not full, a participant clicks on any avatar within the group he or she wishes to join. To depart, a participant will press the *end* key board key, appropriate as the conversation is ending for that participant.

7.30 Addition Features

Because a *conversation circle* has a limit of six members, it is helpful to convey to external participants when it is full. A number of ideas are generated through a brainstorming session, and using a combination of *3D Studio Max* and *Adobe Photoshop*, images are created to visualise them. One idea is to encase a group in a transparent dome, another, to make all avatars of a group semi opaque. The solution proposed, is to change the colour of the group's light to red on the monitors of all external participants. This maintains the naturalistic style of the work, extends the use of the light within the group NVC and does not rely on physical objects. The visual result is that the avatars' faces glow red as does the bright spot on the floor. To reinforce this and help novice users learn the meaning of a red light, an audio alert sounds if they attempt to join a full group by clicking on it with the mouse. These

techniques can also be used to indicate that a group is closed to newcomers even though the number of participants is less than six (i.e. private).

It is proposed that an element of camera control is given to participants in *conversation circles* for two reasons. Firstly, to address the issue that not all avatars are equally visible. Secondly, this will enable participants to stay informed about the level of social activity elsewhere in the surrounding environment while remaining part of a social group. This is achieved by observing other *conversation circles* in the background. A creative piece of work is developed that reuses the 3D scene described in previous sections of this chapter. The camera is moved around a group of six avatars to find a range of useful positions. From this an interactive piece of work is created that shows the conversation circle, demonstrates the positions found, explores how the camera moves between them and how it is controlled.

It is necessary to place constraints on the camera's freedom of movement. The relevant group must always be visible on the screen so that the participant's sense of being part of a conversation is never lost. The camera therefore targets the circle centre at all times. The participant is able to rotate it around the group at three height levels. At the lowest, the camera is close to the group and is useful for being able to see avatar faces more clearly. Also, by rotating the camera the participant can decide upon which avatar faces are most visible. The second height is the default, and reveals more of the background therefore increasing participant awareness of other social activity in the surrounding world. At the highest level, a birds eye view is achieved revealing all social activity that is in the group's immediate surroundings. The camera is controlled using the keyboard number pad cursors. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo7-1 : Camera Control

Description: An interactive piece of work that shows a conversation circle containing six avatars through *Red's* camera. The reader is able to move the camera between twelve predefined positions to choose which avatars' faces are most visible, and to increase awareness of surrounding social activity.

Demonstrates: The piece demonstrates the three camera height levels, how the camera rotates around each, how it moves between the different levels and how the user controls it.

Points to note: The piece does not move the camera in real time as it is constructed using pre-rendered digital movies. Because of this, the camera moves between four predetermined positions around the group at each height level. It is proposed that in an actual implementation the camera will rest at any desired point at each height.

Instructions: Use the keyboard cursor keys to move the camera between the predetermined positions.

7.40 Integrating the Text Channel

The following sections of this chapter discuss creative work that explore the visual properties of the text channel when it is integrated with the *conversation circle*. The text channel refers to both the text messages, and the nicknames of participants which the author advocates should be visible at all times. This assists introductions in a similar manner to name tags at a real world convention. It also enables group members to know who is who at a glance, thus contributing to a more flowing dialogue. The final solution found needs to meet the following requirements:

- text must be easy to read, e.g. no overlap, appropriate size, etc.
- it must be obvious who has authored which message,
- text must not dominate the other channels e.g. body language,

Three pieces of work are developed, the first, addresses how participant identity can be indicated as well as how text messages will be displayed. The second, examines the

behaviour of the text when a group member rotates the camera around a *conversation circle*. The third, explores the temporal ordering of text messages. When communicating in AVWs it can sometimes be useful to back track over one or two previous messages, therefore this piece also explores a way of providing access to a message history.

7.41 Displaying Names and Text Messages

Since the conversation circle idea was conceived, it was envisaged that the text messages would be placed around a *conversation circle* in the image window, so space was reserved for this when the camera position was determined. A number of images are created using the software *Adobe Photoshop* together with a screen shot of the *conversation circle*, in order to explore how text messages, and participant nicknames, will be displayed within this space. Early ideas were to display the text in a similar manner to *Active World's* i.e. plain white text floating near relevant avatar (see Chapter 3). It is decided that participant nicknames will be displayed in this manner however, to make it absolutely clear who has authored which message, speech bubbles are proposed which are used in a similar manner to the 2½D world *The Palace*, i.e. elliptical shapes with traditional tails emanating form the message author.

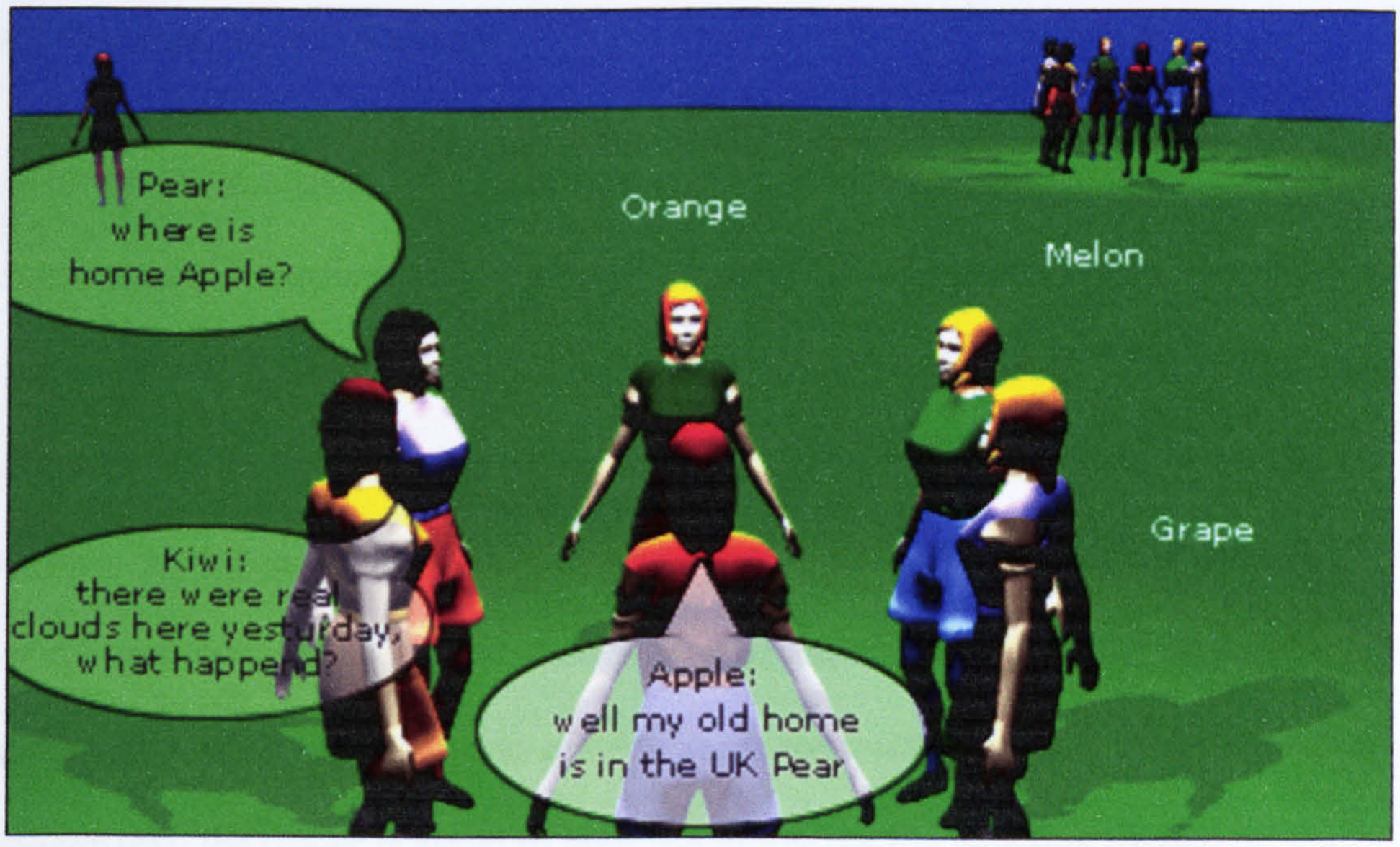


Figure 7-1: Screen Shot of a Conversation Circle Showing How Text Information is Displayed

The best solution is shown in Figure 7-1, in which the reader can observe the following design features. Nicknames are visible at all times, i.e. as plain white text near a participant, or in black and preceding messages when within a speech bubble. The speech bubbles are semi-opaque so not to obscure the surrounding environment. This helps to maintain the sense of being within the world for the participants and allows them to stay informed about background activity. The nearest avatar's speech bubble is more opaque as the text is harder to make out at that point of the screen. A black outline helps to lift the bubbles out of the image. To help readability, only names and text messages of the group a participant belongs to are visible. The reader will note that there is no text displayed for the group in the background of the image, thus preventing message overlap. Lastly, as the text messages are arranged around the *conversation circle*, the eye will move across its centre to follow them. This will continually force participants' attention to glance over the avatars which will perform body language, therefore helping it to be noticed. The reader should note that the image in Figure 7-1 is considerably smaller than it would be on screen in an actual implementation of the system. The speech bubbles are therefore likely to fill less of the surrounding area.

7.42 Rotating Speech Bubbles

It is necessary to consider the behaviour of the text when a participant rotates the camera around a *conversation circle*. It is found that the bubbles and nicknames can be made to follow the avatar as they turn, by rotating in the 2D plane. The mechanism is similar to two mechanical cogs that are interlinked in a perpendicular manner (i.e. 90 degrees). However, the fixed tails no longer point to the message author when the avatars begin to move. They are therefore removed, and replaced with separate arrows that can also rotate so will always point towards the relevant avatar. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 7-2: Rotating Bubbles

Description: A digital movie that shows the view through a participants camera, as it is rotated around a conversation circle containing six avatars each with a visible speech bubble.

Demonstrates: The final bubble design, and how the bubbles behave when a group member rotates his or her camera around a *conversation circle*.

Points to note: The back outline of the speech bubbles has been softened in the final design.

Instructions: Null

7.43 Text Ordering and Message History

The third creative piece aims to explore the following practical questions: how long messages should appear for?, how do they appear and disappear?, how a sense of temporal order can be achieved?, and how can a message history be made accessible? An interactive movie is created to both experiment with potential solutions to these questions, and to demonstrate those proposed. The final piece simulates six participants

in a *conversation circle* exchanging messages. When a bubbles appears an audio alert sounds to help draw participants' attention to it. The speech bubble itself appears in an instant, and remains on screen for a limited duration, or until it is replaced by a new message. Originally it was conceived that the duration would be determined by the number of words in a message, i.e. the greater the number, the longer the duration. However in the interest of achieving a firm sense of temporal ordering there is a standard duration of ten seconds. When the duration is up, a message fades away over five seconds. This is an aesthetically appealing effect which warns others that a message is about to leave the screen. If a message is replaced by a new one, the old one disappears for half a second, before the new one is displayed. The result is a flashing effect which draws attention to the new message.

The piece also experiments with providing a means to recall the last message of any participant. By rolling the mouse onto the speech bubble area, the last message a group member submitted is made visible. To inform a participant that a new message has arrived if they are viewing the previous one, a red dot appears on the edge of the speech bubble. The piece experiments with multiple dots appearing, however it is decided that this is too cumbersome and unnecessary. It would probable hinder the flow of discourse, rather than help it. Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 7-3: Bubbles & Discourse

Description: An interactive digital movie that shows six participants engaged in dialogue within a *conversation circle*. Speech bubbles appear, and then fade out after a limited duration. A level of interactivity is added to the piece so that the reader can try the proposed message recall feature.

Demonstrates: How speech bubbles appear, and fade out, the temporal ordering and the last message recall feature.

Points to note: The duration of the text message on screen is shortened in the piece for viewing purposes. There is no audio alert in this piece and on

reflection, the bubbles should fade away in discrete steps so that the temporal ordering can be recognised more easily.

Instructions: Open the file and watch for approximately twenty seconds to experience the visual properties of text message speech bubbles during verbal discourse. After this time, move the mouse cursor to the area in which the most distant avatar's (one with the dark green shirt) messages appear to recall her previous one. If the previous message is visible long enough, a red dot will appear indicating that a new one has arrived. This is viewed by moving the mouse cursor out of the speech bubble.

7.50 Summary

This chapter has addressed the first task of the design brief through a series of creative pieces of work. A concept named the *conversation circle* has been proposed which represents social groups of up to six members. Only within a *conversation circle* can participants fully communicate with others. *Conversation circles* gather together and automatically arranged group members into a circular formation. With careful positioning of each participant's camera, group members are made visible to one another. The camera position is also such that there is enough space between the outside of the *conversation circle*, and edge of the screen, for text messages which appear within modified, semi-opaque, speech bubbles, which are easy to read, do not overlap and clearly indicate who as authored which message.

The *conversation circle* achieves a relationship between social groups and space by creating a place where meaningful social obligations exist. Lighting and shadow effects help to define this space, and the sense of group ownership of this space is expected to be enhanced by making it off limits to external participants unless they join. In addition to this, a group NVC has been designed which signals what groups exist, where they are located, who the members are, the arrival or departure of group members, whether a group is full and when it has dispersed. In addition to this, feedback can be provided

through body language while one has the floor and it is expected that the personal sense of being in a conversation will be enhanced, increasing the obligation to contribute to the dialogue, thus facilitating communication.

Because the avatar body language has not yet been developed, it is not possible to determine how well balanced the various channels of communication are. This issue is therefore revisited later in the project when the solutions of each of the three design brief tasks are reintegrated (see Chapter 9). Making other group members visible so avatar body language can be seen, and integrating the text with the image channel is however, an essential step towards achieving a multi-modal communication. The next step is to develop an avatar body language which will be used within the *conversation circle*, and that supports both the three stages of social interaction and aspects of the discourse that takes place.

Chapter 8:

Implementing Avatar Body Language

Chapter Overview

The work described in this chapter addresses the second task set out by the design brief in Chapter 4, i.e. develop an avatar capable of performing body language, and a limited set of behaviours for it that supports group interaction and integrates with verbal discourse in a meaningful way. Firstly, knowledge gained from the background research described in *Part Two* of the thesis, is used to arrive at a list of behaviours that support both group social interaction, and aspects of multi-participatory discourse. Next, a naturalistic avatar is designed that has the necessary operators to effectively convey meaning. Using this avatar, a set of gestures that complements the list of behaviours drawn up is designed for use within a *conversation circle*. Measures are taken to achieve gestures whose visual properties are not led by the software used to animate them.

8.00 Design Task Two

The work discussed within this chapter is about the individual rather than social groups. Now that the *conversation circle* has been designed, which enables group members to be visible to one another during discourse, it is the right point in the project to design and implement a set of useful gestures. The set will contain approximately thirty gestures which is a manageable number within the scope the design project, and is well within the limitations of existing AVW technology. Before creative practice can begin, it is necessary to refine knowledge gained from the background research to arrive at a list of behaviours that complement social interaction. This list is informed by two main aspects. Firstly, indicators of behaviour identified during the analysis of the MUD dialogues described in Chapter 5, i.e. acronyms, performative words and punctuation

marks. Secondly, the behaviours derived through the model of social interaction described in Chapter 6.

The next step in the work described is to design and implement a set of gestures through creative practice, that complements the lists of behaviours. It was intended that the naturalistic avatar used in the earlier pieces of work with the *conversation circle* would be reused however, in practice the avatar is found to be inadequate for the task as the particular implementation lacked the ability to represent necessary operators needed to convey meaning. This is because the model was created as a single object which can only be animated using sophisticated techniques such as skinning. This is a dynamic technique where the model is wrapped onto a skeletal biped structure, just like a skin. When the biped is animated and the skin stretches and folds as it does real life. Such techniques are unrealistic for real time, low bandwidth, solutions and are likely to remain so for the foreseeable future. It is therefore necessary to design a new articulated avatar, which is one that comprises separate objects for the various body parts. These parts provide the avatar with the necessary operators to convey meaning.

The last step in the work described is to design a set of gestures for the new avatar to use within a *conversation circle*. A method is developed to design the visual properties of each gesture without being led by the software used to animate them. Five top 3D character animation software titles are evaluated in order to find the one most suitable for the task, i.e. most user friendly, and flexible, allowing sequences of gestures to be constructed from single animations and to be copied and pasted into other avatars within the scene. Ideas for each gesture are generated which are informed by literature, François Delsarte's principles of movement within dance and, the designer's personal knowledge of gesture gained from a life time of, mainly subconscious, observation.

8.10 Supporting the Three Stages of Social Interaction

The table in Figure 8-0 lists the finite set of states, actions and reactions, deemed useful to participants during the three stages of social interaction as determined by the model discussed in *Study Two* (Chapter 6). Each of the behaviours listed can be used in conjunction with, or independently from, verbal discourse.

Start		Middle	End
Form Circle	**Join Circle**		
Neutral	Ask to join conversation	Have floor	Signal a desire to leave soon
Unavailable for conversation	Reject a newcomer	Show listening	Wish to continue
Place an offer of conversation	Accept a newcomer	Wish to take floor	Agree to end soon
Reject a conversation offer		Give floor	Show conversation is finished
Accept a conversation offer			

Figure 8-0: Table Listing the Range of Finite States, Actions and Reactions Needed to Achieve Each of the Three Stages of Social Interaction, Independent From Verbal Discourse.

At the start stage a participant who is involved in conversation can convey one of two states, i.e. neutral, or unavailable for conversation. He or she can become involved in conversation in one of two ways. Firstly, to *form* a *conversation circle* with another non-group member by asking, or being asked and accepting the offer. Rejecting an offer does not necessarily need a gestural complement. It can be shown through the visual channel by doing nothing, turning or moving away. The second way to become involved is to *join* an existing circle. The behaviours necessary are firstly, to ask to join conversation. This is shown by the action of joining a *conversation circle*, and can be reinforced with an explicit verbal message, e.g. "what is everybody talking about?". Secondly, an existing member needs to either accept or reject the newcomer. This may

lead to a gestural complement of its own such as raising both eyebrows if the response is a question, e.g. "anything really….where are you from?".

At the middle stage of conversation, i.e. within a *conversation circle*, the four behaviours listed that are likely to be useful in regulating the flow of message exchange. Each of these will have a gestural complement, however this does not mean that there will be one specific gesture for each. For example, showing one has the floor may be illustrated by series of gestures. In a similar manner, signalling a desire to leave soon in order to end social interaction, may be shown by any one of a number of gestures such as checking one's watch, or shifting direction of face and temporarily breaking gaze.

8.11 Complementing Social Discourse

During the middle stages of social interaction participants are engaged in a process of message exchanges. Section 8.10 lists behaviours that regulate the flow of message exchange, e.g. turn taking. In this section indicators of behaviour are considered that will complement message content. During the analysis of *Study One*, the observation was made that the participants use pre-negotiated techniques to enhance their communication in a text-only medium. Enhancing the communication medium has the potentiality of increasing interactive experience by compensating for the forms of body language which are missed. In a graphical virtual world that features avatars, one would expect to see such meanings conveyed through, or complemented by, virtual body language.

Three techniques were identified. Firstly, the use of performative words which are used by participants to describe actions of the body (e.g. *wave*). Secondly, the use of acronyms (e.g. LOL), many of which also describe body language, although the

observation was made that the use of specialised language such as acronyms, can make novice users feel alienated from the group when they do not understand their meaning. If acronyms can be translated into visual behaviours, novice users will be able to easily interpret their meaning, thus relaxing the social barrier. The third technique identified was the use of punctuation marks from traditional grammar. During the evolution of the specific style of grammar typical of virtual world communication, many of the punctuation marks of traditional grammar have been rendered obsolete. A few remain such as the full stop (although its role has been adapted, i.e. consecutive dots " ..... "). *Study One* showed that the dots are used to implicitly convey meaning that in real life, is carried through a combination of voice tone and body language. If other punctuation marks are used to compensate for the absent body then they need to be singled out and considered for the list.

The next step is to identify other performative words, acronyms and punctuation marks that are used within the textual communication. To do this it is necessary to re-examine the six dialogues used for the discourse analysis of *Study Two*. Each dialogue is read through in turn and any other examples are logged. The result is a relatively large list of examples. In addition to this a fourth device is identified known as *emoticons*. These are a combination of certain text characters, that represent an emotion by bearing similarity to facial features. An example is a colon followed by an closing bracket, which signifies a smile by representing two eyes and mouth rotated ninety degrees counter clock wise, i.e. :) . Emoticons are generally used to convey the spirit that the message was written in, or used on their own to convey an emotional response. As with performative words, they are used to convey body language therefore one would expect to see avatars perform the gesture or facial expression that they describe.

The next step is to examine the meaning of each indicators of behaviour and consider its importance within the context of the social interaction. To ensure that the correct meaning of each example is understood, reference is made to a website that presents a comprehensive listing of these symbols, and describes each one's standard interpretation [14]. The web site is also useful in that it presents other examples that might inform the list. There are however, too many to include them all and this is not necessary for this research, so the decision is made to include only a few from each category. Many are immediately disregarded for any of the following reasons: their meaning is offensive, they are too rarely used, they describe physical contact or simply that a gestural interpretation is not possible. From those that remain, only the ones that appear most frequently within the dialogues are selected. The table in Figure 8-1 shows the acronyms, performative words, emoticons and punctuation marks that are selected to be complemented by body language, mapped onto the main meanings they can convey.

The reader will note that there are three occurrences of laughter in the list. Originally, a single gesture to indicate laughter was to be used, however the three indicators of behaviour s appearing so frequently within the dialogues suggest the importance of representing laughter at three different levels for participants. All three are therefore included and this will increase the number of situations in which laughter can be used. The acronyms AFK and BBS, are used to indicate that a participant has been temporarily distracted in the real world but intends to return soon. It is important to convey this state to conversation partners so that they are aware that a sudden absence of messages is not due to technical breakdown and this contributes to the smooth running of the social interaction. The question mark is important within virtual world textual communication as it indicates a request for feedback, and is therefore often used

[14] see http://www.marshall.edu/isp/ct107/emoticons.html

to hand over the floor. The exclamation mark is used to emphasise a particular message and the high frequency at which it appears at the end of text messages, indicates the importance of this expression for participants.

Indicator of Behaviour (Indicators of behaviour)	Meaning
Wave	Hello, Goodbye, Ask for conversation,
Smile , :) , :-) , :>	Smile, Happy, Ask for conversation, Accept offer, Accept Newcomer to group
:(	Sad face, unhappy, disappointed
Kiss, :-* , :- x	Kiss, Hello, Good bye
Chuckle, hee hee, ha ha,	Giggling
LOL	Laughing Out Loud
ROTFL	Rolling Around On The Floor Laughing
AKF/BBS	Away From the Keyboard / Be Back Soon
?	Request feedback
!	The message is emphasised
.....	The message in unfinished
CU, CUL, CUL8R, B4N	See You, See You Later, See You LATER, Bye FOR Now

Figure 8-1: Table Mapping the Acronyms, Performative Words, Emotion Icons and Punctuation Marks Selected to be Complemented by Body Language, to Their Meanings.

The precise meaning of the consecutive dots has been more difficult to determine. In the MUD analysis, the two participants of the *One-to-One* dialogue used them to indicate that it was the other's turn to create a line in the story that they made up together. This example, however, shows an unusual situation therefore this meaning is disregarded. By cross referencing the dialogues subjected to discourse analysis with a number of other dialogues generated through virtual worlds sessions, a more consistent use of the dots has been determined. Located in the middle of a text message, they indicate a pause in speech. Located at the end of a message the dots imply that the message is unfinished. The latter meaning is one that can further help regulate the flow of conversation so is selected for the set.

8.12 Lifelike Behaviours

In addition to those listed above, a number of lifelike animations will be designed. These will address the issue raised in Chapter 3 (see Section 3.20) concerning the way lifelike animations are used in existing AVWs. While the designer is in favour of a form of body language which increases the sense of life in AVWs, there is a need to rethink how they are used to minimise confliction with message content. One approach is that the computer automatically ceases triggering these animations when a participant is speaking. Thus NVC will indicate that a participant is not speaking and prolonged NVC would emphasis a participant's lack of engagement in the conversation which may gain the attention of another group member who may decide to try to involve and bring the participant into the conversation. Here, the animations will be automatically and randomly triggered when a participant has been idle for a specified period. In this case, it will be that these animations do not to carry obvious meaning. Therefore they will be based on bodily needs such as shifting body weight from foot to foot to relive straining muscles and assist circulation.

8.20 Modelling a New Avatar

In order to determine what operators the new avatar needs to carry meaning, it is useful to draw on the various design considerations discussed within Chapter 4, as well as findings from observations made during other creative work in subsequent chapters. Firstly, the avatar will be naturalistic in style, therefore it will use operators based on those of the real human body. Secondly, the NVC will focus on hand gestures rather than other channels such as facial expressions as they are more visible, particularly when an avatar is viewed from behind. The arms therefore need to be fully functioning in order to mimic real life gestures. Thirdly, if the avatar is given facial features, there

will be an expectancy for the face to play a role in carrying meaning. The decision is made to use basic facial expression such as a smile and eyebrow signals, but mainly in support of the hand gestures. Fourthly, the conversation circle solution raises a further issue which concerns the role of the legs within the body language devised. In real life communication leg position and movement contribute to meaning particularly when dramatisation is required. From the majority of the vantage points of the *conversation circle* available, the avatars' legs are out of view. They will therefore not be used to contribute to the NVC other than to contribute to direction of face, which further shifts the emphasis to the upper body, arms and head. The existence of legs however will raise an expectancy for the avatar to walk in a natural manner. They are therefore made with the minimal operators to achieve a reasonably convincing walk.

The diagram in Figure 8-2 shows the final avatar in two positions. It illustrates the various body parts and gives an impression of the range of movement the model has. The eyes are created in the same manner used for the VRML implementation of Cedrick, i.e. separate eyeballs and irises, as it was anticipated that they would play an important role in the NVC. In practice they are found to be too subtle to be useful but have been left in for the time being as they may be useful in future work, particularly if the camera language is further developed to include close-ups of the face. The mouth has two interchangeable positions which are neutral and smiling. Rather than animate the mouth between the two positions using interpolation techniques, which would significantly increase the file size, two separate mouth objects are created. VRML objects can be made invisible via a transparency property. The visibility of the two jaw pieces is alternated to achieve the two mouth positions. To help visually distinguish certain facial operators they are coloured. The lips are coloured red and white teeth help

PAGE NUMBERS CUT OFF

IN

ORIGINAL

to make the smile more obvious. The eyebrows are darker than the hair and are slightly enlarged.

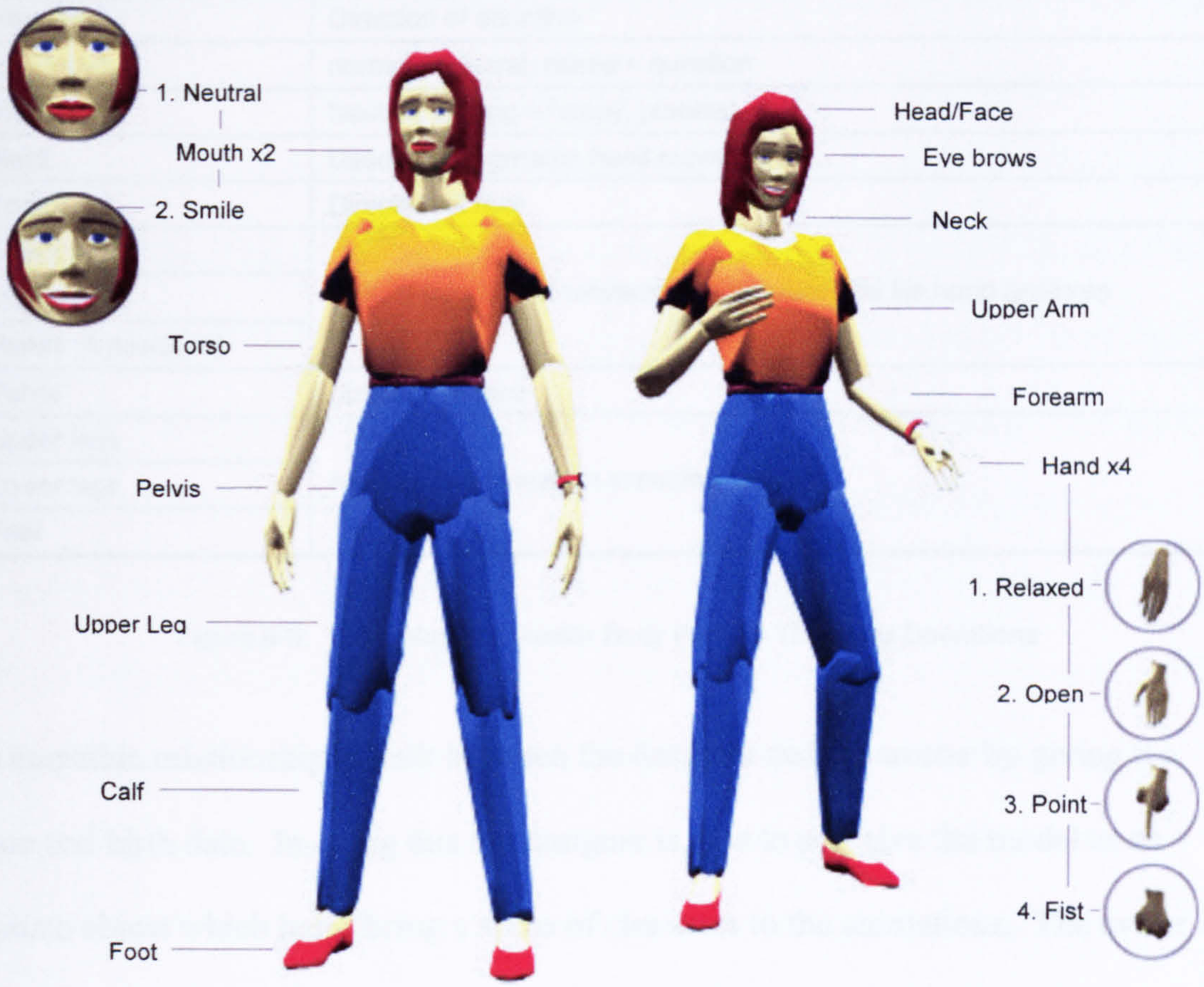


Figure 8-2: Diagram Showing the New Avatar and its Various Operators.

Body parts that have joints are rounded at the end so that they intersect with the connecting segment. This enables limbs to be manipulated without leaving gaps at the joints which can weaken the illusion of life. Rather than animate fingers to achieve different hand signals, four interchangeable hand objects are created to further minimise file size, i.e. relaxed (neutral), open, point and fist. The choice of which hands to use came about during the animating process. Each new hand was created as it was required, resulting in only four types. The new avatar has achieved a lower polygon count then the earlier single mesh model and has the necessary operators to carry

meaning. The table in Figure 8-3 shows the various body parts and their corresponding key operations.

Avatar Body Parts	Operation
Head/face	Direction of attention
Eye brows	normal = neutral, raised = question
Mouth	Neutral, smiling = happy, pleased, etc.
Neck	Used to exaggerates head movements
Torso	Direction of face
Upper arms	Natural freedom of movement to emulate real life hand gestures
Lower arms	
Hands (4 positions)	
Pelvis	Direction of face
Upper legs	Allow avatar to walk in expected fashion
Lower legs	
Feet	

Figure 8-3: Table Mapping Avatar Body Parts to Their Key Operations

An empathic relationship is built between the designer and the avatar by giving it a name and birth date. In doing this the designer is able to perceive the model as an animate object which helps bring a sense of character to the animations. The avatar is named Isabel which is Spanish in origin meaning 'consecrated to god'. Appropriate when we consider that the word 'avatar' originates from Sanskrit and translates as the manifestation of a god on earth in the flesh. Isabel was born on the 9th of July (the day modelling was complete) making her astrological star sign Cancer.

8.30 Gesture Design Method

A method is developed to design the visual properties of the gestures which aims to avoid designs that are led by the software used to animate them. To begin, a search was conducted to find visual sources of reference to assist the design of each gesture, unfortunately little of use was found. One idea was to examine body language in television soap operas such as the BBC's *East Enders*. This idea was dismissed as

specific gestures are sought and locating them within the complex language of the screen-based medium would have been too great a task. The designer therefore turned to literature, however much of the relevant material originates from the social sciences, which tends to examine body language from a psychological perspective. In addition to this, much of this literature focuses on facial expression rather than hand gestures. Four researchers who have discussed visual properties of the face expression in their work are Paul Ekman, Wallace Friesen, Janet Bavelas and Nicole Chovil (see Section 2.60). A number of standardised codes, at least within western cultures, can be drawn from their work. For example, a verbal message can be presented as a question by raising both eyebrows when speaking. Body postures and hand gestures are informed by François Delsarte's principles of movement within dance [Shawn 74]. Delsarte lived a century ago and dedicated his life to understanding the how the body moves under the stimuli of emotion. This material is particularly relevant as it is describes a highly expressive form of communication using the body. Many of his principles are therefore applied to the body language designed, for example; chest held high indicates self respect and pride whereas a protruding abdomen indicates animality, sensuality and lack of bodily pride. This understanding helps to design a neutral pose for *Isabel.*

Knowledge gained from the literature discussed, combined with personal observation of body language, is used to generate a number of ideas for each gesture. Personal observation refers to the designer's own knowledge of gestural communication. From the moment that we are born, we begin to learn both verbal and non verbal codes through experience and observation. In order to animate the model, five 3D character animating systems are considered which are *V-Realm Builder* from *Ligos Technologies*, *Poser* from *Fractal Design*, *Lifeforms* from *Credo Interactive*, *Lightwave* from *NewTek. Inc* and *3D Studio Max* with the *Character Studio* Plug-in from *Kinetix*. The latter is

chosen because it provides the greatest flexibility and versatility. It uses a biped system, which is a hierarchical skeletal structure that demonstrates an understanding of human body mechanics when it moves. *Isabel's* various body parts are attached to the corresponding biped skeletal parts which are then animated. The advantages are that production time is reduced, more realistic movements are achieved and biped animations can be copied and pasted into create sequences and transfer them to other models within a 3D scene. This will make easier the task of animating a number of avatars interacting within a *conversation circle* later in the project.

In order to avoid being led to the end solutions by the 3D animating software used, the gestures are designed without the use of the computer. Using his real body, the designer explores the visual properties and timing of each idea by performing in front of a mirror that is located next to the computer. The movements of the best ideas are then mimicked in the new avatar using *Character Studio* software, therefore achieving an approach where the software works for the designer, rather than the other way round. Each gesture is first designed for a one-to-one scenario. When the set is complete, each gesture is reconsidered in turn and if necessary, adapted to operate within a group of more than two. Rather than create new gestures that are similar but address group members at different spatial position relative to an avatar, some basic rules are established that can automatically modify existing ones in real time. This minimised the number of gestures in the set. These rules are discussed in Section 8.50.

8.31 Practical Issues and Design Features

The sources of knowledge used to inform the visual qualities of the designed gestures inevitably lead to cultural bias. Consideration is given to this issue and where possible, cultural exclusion is avoided. Nether the less, it can be argued that participants in an

AVW have an obligation to respect its culture, and should be prepared to learn the meanings of the gestures appearing in it. This is no different to visiting a foreign country. We should be prepared to learn the language and customs of that culture. In any case, it would be relatively straightforward to create culturally specific gesture sets which could be used in an avatar rather than those provided here. Although the visual properties will be different the underlying behaviours supported by gesture are likely to be the same.

In order to make body language more visible, a key principle is borrowed from a renaissance style of improvised street theatre known as *Commedia Del Arte*. During performances, the actors wore masks to indicate the character they were playing. Because the masks hid facial expression, they had the added effect of encouraging the use of more expressive behaviour via the body. As the NVC will focus on gesture, limb movements are exaggerated to make them more visible and remove attention from other forms that play a lesser role. For further information on *Commedia Del Arte* see [Fisher 92], and for its application to CVE see [Tuomola 99].

Delsarte observed that emotion produces bodily movement and if the movement was correct and true then the end result of the movement would leave the body in a position also expressive of the emotion. However, it has been necessary to make each gesture begin and ends in a neutral pose for the following reasons: to achieve a seamless transition between the various gestures in the library, to ensure that undesired meanings are not accidentally generated by the integration of two animations and, to clearly mark the end of one message and the start of a new one. As a compromise, many of the gestures hold the body in a position for a short duration, then it returns to the neutral pose. The reader will note that this is not true of the available and unavailable states.

They remain set until changed by the participant. Any gesture performed returns to the set state.

One idea worth documenting that may be of use to other researchers, is the effect of distance on the visual properties of a gesture. For example, greeting a participant in the distance might consist of waving an arm in the air. If standing close together, the same meaning may be conveyed through a wave of the hand. This implies that it may be useful to have two versions of each gestures in the set. One for when the recipient is close, and the other for when he or she is distant. The set designed within this design project however, will operate within a *conversation circle* therefore there is a consistent spatial relationship between avatars. This principle is therefore only used within the set designed for the *wave* gesture (see next section)

8.32 The Proposed Set Of States And Gestures

It is found that the list of behaviours and indicators of behaviour s drawn up can be effectively complemented with a set of twenty six states and gestures, which are as follows:

States

01. Neutral
02. Unavailable

Gestures

03. Wave Far
04. Wave Close
05. Smile
06. Sad
07. Nod Head
08. Shake Head
09. Uncertain/So What
10. I am/Myself
11. Sleepy
12. Kiss
13. Chuckle
14. Laughter
15. Laughing Lots
16. Question
17. Emphasis
18. Unfinished
19. Thinking
20. Phone Call
21. Clap
22. Stretch
23. Check Watch
24. Side Shift
25. Arm Swing
26. Tip Toe

Some of the gestures listed are able to convey different meanings depending on the conversation context, for example, *wave close* can be used as a greeting, to place and offer of conversation, or show a mutual agreement that a conversation is over (waving good bye). Others, such as *nod head*, are more specific, i.e. yes, ok. One unexpected outcome is that some gestures can be used to complement other behaviours not on the list, for example, the *uncertain* gesture can also be used to complement the verbal response, "so what!".

The reader will note that some gestures in the list can also be thought of as moods which can be represented through state, e.g. *sad*. However, it is proposed that they are used in the set as gestures that convey an emotional response or convey the spirit a message is written in. It is possible to use them as moods, but they would have to effect the visual properties of all other gestures performed. To create a system that achieves this, it would be possible to combine the work described in this thesis with other projects that have explored and implemented systems where emotion effects an avatar's mannerisms, such as *Comic Chat* which was discussed in Chapter 2 [Kurlander *et al.* 95], the *Public Park* project also discussed in Chapter 2 [Becheiraz & Thalmann 96] and Norman Badler's EMOTE system [Badler *et al.* 99].

8.33 Description of Each State/Gesture

Following is a description of the visual properties of each state and gesture in the set, as it would be seen in a one-to-one situation.

01. Neutral

> The neutral pose aims to convey a sense of being relaxed, confident, alert, honest, approachable and of well being. It is also designed to show that the participant is listening and paying attention. The feet are shoulder width apart and the legs slightly bent. The torso is upright, the neck is straight and the head

is titled slightly upward. Combined with the level shoulders and raised chest, this conveys confidence and that the participant is alert. The arms hang by the sides of the body so when standing head-on, there are reduced psychological barriers between participants generating a sense of approachability (arms folded in front of the body is often received as a defensive stance). The hands are relaxed which Delsarte suggests signifies honesty and good intention. The mouth and both eyebrows are in neutral position. This pose is reminiscent of a soldier standing at ease.

02. Unavailable

Similar to the neutral pose but the avatar becomes semi opaque, and remains so during any movement. Benford has suggested that availability can be linked to degree of presence within an AVW [Benford *et al.* 97]. Although an abstract signal, the ghosting effect works as it reduces the sense of presence of a participant for others in the environment.

03. Wave far

The head tilts backwards to inform any participants close by that the gesture is intended for a recipient some distance away. Both arms rise above the head with hand palms facing forward. Both eyebrows are raised and the mouth is smiling through out the action. The arms are waved four times then the avatar returns to the neutral pose.

04. Wave Close

Open right hand is raised to the front and side. The palm faces forward and is oscillated from left to right at the elbow joint four times. To exaggerate the gesture the right hand emphasis the waving motion. The head is slightly raised as is the left arm. During the action both eyebrows are raised and the avatar smiles.

05. Smile

This is the most difficult expression to implement due to its subtlety, and there is no existing well-known hand gesture to use as a substitute. The proposed solution is to accompany the facial expression with more visible bodily movement that participants will learn to associate with a smile. When the movement is seen, participants will understand that a smile is been performed, even when an avatar's face is not visible. For the movement, the torso rotates forward, the head remains level and the arms move away from the side of the body. The mouth smiles and the body returns to the neutral position. The mouth remains smiling for a time, or until the next gesture is performed.

06. Sad

The head tilts forward and the shoulders drop. This is the universal gesture for sadness, shame and dishonour. It is held for at least four seconds or until the next message so that it is not confused with a bow, which signifies respect. Delsarte observed that a recoil of the body is signifier of a negative emotion. The gesture is therefore exaggerated by drawing the arms into the body.

07. Nod Head

The head nods backward and forward four times. The movement is exaggerated using the neck and torso. To make the gesture more visible the arms move away from the body. The meaning of this gesture is one that is positive therefore the avatar smiles during the gesture.

08. Shake Head

The head shakes from left to right three times. The neck and waist also rotate to exaggerate the movement.

09. Uncertain/so what

Open hands move outwards from the body with palms facing upwards. This signifies an honest declaration of not having an answer or explanation. The hands and shoulders are then quickly raised over a short distance and stop abruptly as if to throw the conversation into the air for another participant to pick up. Both eyebrows are raised during the action.

10. I am/myself

The right hand is placed on the chest. Delsarte explains that this indicates to hold, posses and include. In this context it becomes a reference to oneself. The avatar's left arm is extended forward towards the recipient's avatar, the palm faces slightly forward and upward. This gesture can be used within a positive or negative verbal message therefore the avatar's mouth remains neutral.

11. Sleepy

The palms of both hands are pressed together and placed to left side of the head at an angle of approximately forty five degrees. This symbolises a pillow, onto which the head is rested. The posture is held for a time, then the avatar returns to the neutral pose.

12. Kiss

A kiss usually involves body contact which is not addressed within the scope of this project, therefore the well-known gesture of blowing a kiss is used. The right hand is brought up to the mouth with the palm facing upwards. The head is lowered to the hand onto which a kiss is placed. The torso leans forward as the right hand is thrust towards the intended recipient to signify the kiss is being blown from the hand. The mouth smiles and then the avatar returns to neutral.

13. Chuckle

This is one of the most difficult gestures to design. The solution draws upon a gesture that is often used by children to symbolise laugher, rather then spontaneously performing it. It involves bringing one hand up to the mouth, and dipping the legs three times. The mouth smiles during the action.

14. Laughter

This gesture indicates laughing to a greater degree then that previous. The body spasms to indicate laughter to the point where that the participant is rendered out of control for a period of time. The laughter fades out and the animation finishes with the palms of the hands slapping the thighs. The mouth smiles during the movement.

15. Laughing Lots

Although based on the *Laughter* gesture, it is different as its duration is much longer indicating uncontrollable laughter. The left hand is raised to show the recipient the palm as if to gesture, "please stop I am laughing too much". At the end of the action the laughter fades out and the avatar returns to the neutral pose.

16. Question

The right arm is extended toward the recipient and the palm of the right hand is shown. This represents handing the conversation over to the message recipient. The movement is exaggerated by the torso titling slightly forward. The display is held for a short time or until the participant submits another message. The right eyebrow is raised to punctuate the message.

17. Emphasis

The avatar dips at the knees, then the body straightens to its maximum upright position and the head tilts backwards. The arms move outwards from the side of the body to emphasise the movement. The hands are rotated round so that the palms face forward. This position is held for a maximum of four seconds before the avatar returns to the neutral pose. Both eyebrows are raised to punctuate the message.

18. Unfinished

The recipient is asked to wait by the raising of the right hand and the palm being thrust forward. This resembles the stop gesture of traffic wardens. The head turns away from the hands and is lowered slightly. This gives the impression that the avatar owner is planning the rest of the message. The heads returns to regain eye contact before the arm is lowered. This gesture is held until the next message arrives or until a timeout has occurred.

19. Thinking

The torso tilts backwards and turns toward the left. The left leg is brought forward to counter-balance the leaning torso. This is not a step forward which, as Delsarte points out, is an act of aggression. The head is raised so the gaze appears to travel off into the distance. The left arm is bent across the waist so as to impose a temporary barrier to another avatar while the left hand is brought up toward the chin into a classic thinking pose. The right hand becomes pointed so as not to draw attention to the chin, and to explain that the gaze direction should not be followed by presenting an alternate direction. The gesture is held until the next one is performed.

20. Phone Call

The right hand pulls a mobile phone from right trouser leg pocket. It is brought into clear view of the recipients and a slight hand movement creates the illusion that buttons are being pressed. The phone is then lifted to the right ear. While this takes place the left hand is extended forward with an up-turned palm which in this context is an offer of an apology for the interruption. From this point the head randomly nods as if to be responding to the external conversation. This is repeated until the participant begins to interact with the group again. At this point the mobile phone is put back into the pocket and the rest of the body returns to its neutral position. This gesture clearly indicates that the participants attention has temporarily being removed from the social activity in the AVW.

21. Clap

The hands are raised and held out in front of the body approximately two shoulder widths apart. They are then rapidly drawn together to create the impression of a single clap action. This action occurs only once.

22. Stretch

The arms are raised and then drawn back and straightened to stretch out the chest while the avatar rises onto its toes.

23. Check Watch

Isabel wears a watch on her left wrist which is brought up into view of both Isabel and the recipient. Emphasis is brought to the watch through gaze direction. And by pointing to it with the right hand.

24. Side shift

The first of three lifelike animations. Isabel shifts her weight from the left to the right foot, returns to the neutral pose.

25. Arm swing

The second lifelike animation. Isabel swings her arms backwards, forwards, then back to the neutral pose.

26. Tip Toe

The third lifelike animation. Isabel rises on to her toes then returns to the neutral pose.

8.34 Mapping Gestures to Corresponding Meanings

The table in Figure 8-4 shows the correspondences between the various meanings determined by the list of behaviours and indicators of behaviour s, and the various states and gestures.

Gesture/State	Behaviour
01. Neutral	Well being, alert, paying attention, honesty, approachable, available for conversation.
02. Unavailable	Unavailable for conversation.
03. Wave Far	Greeting, Invitation to chat.
04. Wave Close	Greeting, offer conversation, say good bye (conversation is over), accept an offer of conversation, accept a newcomer to the group.
05. Smile	Accept an offer of conversation, accept a newcomer to the group, message is written in good spirit, say good bye.
06. Sad	Sad, unhappy, disappointed.
07. Nod Head	Yes, agree, confirm, accept conversation offer.
08. Shake Head	No, disagree, reject conversation offer.
09. Uncertain/so what	I don't know, uncertainty, so what, your turn.
10. I am	Reference to self.
11. Sleepy	I am tied and wish to end/leave conversation soon.
12. Kiss	Kiss, affection, accept conversation offer, accept newcomer to group, signal good bye.
13. Chuckle	Happy, amused, indicate that a verbal message is a joke.
14. Laughter	Very amused, indicate a message is a joke.
15. Laughing Lots	Extremely amused.
16. Question	Request feed back, give floor.
17. Emphasis	Message is emphasised.
18. Unfinished	The message is unfinished.
19. Thinking	Regulates conversation flow and indicates a stoppage of message is not due to technical breakdown.
20. Phone Call	Participant is talking to someone in the real world.
21. Clap	Signal desire to end/leave conversation soon, get groups attention.
22. Stretch	Signal desire to end/leave conversation soon, get groups attention
23. Check Watch	Signal desire to end/leave conversation soon.
24. Side shift	Participant has not contributed to verbal discourse for sometime, participant is bored.
25. Arm swing	Participant has not contributed to verbal discourse for sometime, participant is bored.
26. Tip Toe	Participant has not contributed to verbal discourse for sometime, participant is bored.

Figure 8-4: Table Showing Correspondences Between Meanings and Gestures/States

8.35 Demonstration of Gestures

Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM.

Demo 8-0: Gestures Demonstration

Description: An interactive digital movie showing each of the states and gestures designed one at a time. The reader is given control to advance the movie or view the previous state/gesture.

Demonstrates: How the avatar operators are used to achieve a set of gestures that complement the list of behaviours and indicators of behaviour s, for a one-to-one situation. Also, the visual properties of each of the states /gestures.

Points to note: The avatar begins and finishes each behaviour in the neutral pose. In an actual implementation, some of the gestures such as *Thinking*, will be held until the next gesture is performed or the avatar changes state. All animations in the piece have been slightly slowed to help the reader observe what is going on.

Instructions: To view each animation, click *Next*. To view the previous animation, click *Previous*.

8.40 Additional Animations

Two animations that it is necessary to develop, but which are not part of the described set, are *walk* and *turn*. Animating a walk is a relatively easy task using *Character Studio* however, getting *Isabel* to walk in a manner that is appropriate is an entirely different matter. Delsarte points out that a walk can convey considerable information about one's mood, character, identity and intentions. The final solution therefore aims to convey a sense of calm, self confidence and well being. This will visually add to the overall atmosphere of the AVW, and more importantly, to how a participant will be received by others. The torso and neck are straight, and the head tilts slightly upward. Shoulders are kept fairly level and the chest is raised indicating confidence and self pride. The speed and stride length are that of a relatively brisk walk until the avatar

approaches a *conversation circle*, where they exponentially decrease in velocity and length over the last eight steps to allow enough time for the circle to expand. The final step places the avatar along the *conversation circles* circumference.

Creating convincing foot movements for an avatar turning on the spot proves to be extremely difficult. This is particularly true for a real time animation because the degree a participant will turn is unknown to the computer. An animation was conducted to gain an impression, of how convincing rotating an avatar in the *y* axis, i.e., without any movement of the legs or feet. This was judged to be satisfactory and was therefore taken forward as a solution to the turning action.

8.50 Rules for Modifying the Set for Group Conversation

The states and gestures described are designed to operate within a one-to-one scenario. In a group of more than two, many of them (such as the *Question*) will need modifying to show who the message is addressed to, i.e. which individual or the whole group. Rather than create new gestures for the set, some rules are proposed to automatically modify the existing ones in real time so that they address the correct group member(s). This is technically easy to achieve as the computer is able keep track of who is whom and where they are located in relation to one another.

The rules were determined by experimentation with a 3D scene containing a number of avatars in a *conversation circle*. Starting with the neutral state, it is was found that an avatar can give attention to a specific member by rotating the avatar's head to gaze at them. To relax the body, particularly when the head needs to be considerably rotated, the waist is also turned towards the recipient, but not to the same degree as the head. The feet remain directed towards the circle centre to maintain the sense of group

membership. The whole group can be addressed by a sweeping gaze, achieved by rotating the head and waist over time.

The next step is to reconsider each gesture in the set and determine which ones may need to be directed towards certain members, and whether turning the head and waist is sufficient. It is found that most in most cases it is, however there are some that need further modification. For example, the *Kiss* gesture also needs the right arm to be thrust towards the intended group member. The *Question* gesture is also effected by the spatial relationship to the recipient. If the recipient is located on the left side of the circle, then it is better to extend the left arm toward them when performing the action, rather than the right as was originally designed. If the right arm is used when addressing someone on the left, then a barrier is created that excludes the rest of the group.

To address the whole group, all relevant gestures can be made into sweeping actions in support of the gaze behaviour. The *Question*, also requires additional modifications to complement the sweeping action (this particular action supports the *open question* technique discussed in Chapter 5,). There are also some gestures that do not need modification. For example, the *Laughter* can start by retuning the avatar to the unmodified neutral pose, performing the action then return to the modified neutral pose so he or she is gazing at the correct recipient. The table in Figure 8-5 shows how each gesture/state is modified to operate within a group situation. Those not needing any are marked as null. Those simply requiring gaze adjustment are marked as 'head and waist'. The other body parts to be manipulated are stated for those that need further modification. The rules developed will be demonstrated in the final piece which is discussed within Chapter 9 (see Section 9.20).

Gesture/State	To Address Individual	To Address Whole Group
01. Neutral	Head & waist rotate	NA
02. Unavailable	Null	NA
03. Wave Far	Null	NA
04. Wave Close	Head, waist & right arm	Sweeping head, waist & right arm
05. Smile	Head & waist	Sweeping head & waist
06. Sad/Unhappy	Null (gaze recoiled)	Null
07. Nod Head	Head & waist	Sweeping head &waist
08. Shake Head	Head & waist	Sweeping head & waist
09. Uncertain/so what	Head & waist	Sweeping head & waist
10. I am/myself	Head & waist	Sweeping head & waist
11. Sleepy	Head, waist & hands	Sweeping head, waist & hands
12. Kiss	Head, waist & right arm	Sweeping head & waist, right arm thrust repeated once for each recipient
13. Chuckle	Null	Null
14. Laughter	Null	Null
15. Laughing Lots	Null	Null
16. Question	Head & wais. Arm closest to recipient is extended towards them	Sweeping head & waist. Left arm extends then retracts as other takes over (visa versa)
17. Emphasis	Head & waist	Sweeping head & waist
18. Unfinished	Head & waist	Sweeping head & waist
19. Thinking	Null	Null
20. Phone Call	Null	Null
21. Clap	Head & waist	Sweeping head & waist after clap action.
22. Stretch	Null	Null
23. Check Watch	Null	Null
24. Side shift	Null	Null
25. Arm swing	Null	Null
26. Tip Toe	Null	Null

Figure 8-5: Table Showing How Each Gesture/State is Modified to Operate Within A Group Situation

8.60 Chapter Summary

An avatar has been designed with the minimal operators necessary to carry meaning, and a set of gestures has been implemented for it to be used within a *conversation circle*. In order to determine what gestures to design, knowledge has been drawn from the two studies conducted as background research. Firstly, techniques that participants

use to compensate for the absent body in the text-only medium are translated by gesture, i.e. acronyms, performative words, emoticons and punctuation marks. Translating this specialised language also helps to relax social barriers for novice users who can be made to feel excluded when they do not understand meaning. Secondly, the behaviours necessary to achieve the three stages of social interaction are included in the set, which were identified using the model of social interaction described in *Study Two*. Lastly, some additional general purpose behaviours have been added that were not derived through the formal methods, but are useful within the context of AVW social interaction.

In order to carry meaning, it has been necessary to design a new, naturalistic, articulated, hierarchal avatar that comprises of separate body parts (i.e. head, neck, torso, upper arms, forearms, hands, pelvis, upper legs, calves and feet). These parts act as operators that have the same freedom of movement as they do in real life. The avatar is primarily designed to exploit the hand gesture form of NVC, however there are some facial operators, i.e. eyebrows which are used to punctuate messages, and two mouth positions which are neutral and smiling.

A method has been devised to design the gesture's visual properties without being led by the 3D animating software used. Firstly, ideas are generated for gestures that complement the list of behaviours and indicators of behaviour, which are informed by literature and the designer's personal knowledge. These are explored away from the 3D software, by using the designer's own body to perform them in front of a mirror located next to a computer. The 3D software is then used to mimic the designer's movements in the new avatar. This achieves gestures that are closer to the designer vision.

A set of twenty six states and gestures has been designed that can be used to complement the behaviours derived. Some of these directly represent a specific behaviour, e.g. *nod head* represents yes, agree, confirm, etc. Others can be used in a number of different ways depending on the context, for example, the *smile* can indicate that the participant is happy, pleased to see another participant, show acceptance of a conversation offer, to say good bye, etc. It has been found that it is helpful to make the gestures more visible by exaggerating movements. Lastly, rules have been proposed to manipulate relevant gestures in real time, so that they address specific members or the whole group. In most case this involves rotating the head and waist to gaze at the intended recipient, although a few gestures need additional manipulation of the arms and hands.

PART 4: EVALUATION & CONCLUSIONS

Chapter 9:

Evaluation

Chapter Overview

Evaluating the final outcome is an integral part of a creative design process. While evaluation and feedback has been an on-going theme within the project, this chapter aims specifically to evaluate the larger design activity. Firstly the general evaluation method is discussed. Next, a study is presented that shows each of the gestures developed to a group of subjects who are asked to write down what they think each gesture is trying to convey. The aim is to gain an understanding of how well the set of gestures convey intended meaning. Lastly, the development of a final piece of creative work and its outcomes are discussed. This is used to gain an understanding of how effectively the various parts of the design problem, that have been hitherto handled separately, work together, particularly with respect to proxemics, body language, group NVC (non verbal communication), etc.

9.00 General Evaluation Method

There are two aspects of the work that can be evaluated without the need to integrate the design solutions with advanced technical systems, which, is beyond the resources of the design project. First, we can assess how well the implemented individual gestures convey their intended meaning. For this purpose we used the *Gesture Demonstration* which was discussed in Chapter 8. The creative work, which shows an avatar perform each of the gestures in turn, was shown to a group of subjects who had no prior knowledge of the work but are informed of the situation in which the gestures would be seen. The subjects were asked to write down what they thought each gesture meant to convey. Analysis of the data produced indicates that the set is effective in conveying intended meaning (see 9.10).

We can also assess how well the different parts of the design solution work together, e.g. proxemics, body language, group NVC, etc. A new piece of work was produced called *Final Piece*. A representational textual dialogue from an actual 3D AVW session was selected from a number that had been gathered. A digital movie was created using this dialogue as the script, complemented by the full repertoire of NVC developed in the project, i.e. set of gestures, *conversation circle*, etc. The movie initially shows three participants engaged in dialogue within a *conversation circle*. Two newcomers join it in turn, and the piece ends when one of the group members departs. The movie demonstrated that the NVC developed is effective although its construction brought into question the use of text as a means of linguistic dialogue.

9.10 Evaluation Of Individual Gestures

Following a presentation of the design work at a research seminar that took place in November 2000 at Exeter School of Art & Design, the creative piece *Gesture Demonstration* (see Demo 9-0) was shown to a group of seven research students and academics. The digital movie shows the avatar *Isabel* standing face on and in full view. *Isabel* performs each of the gestures developed in turn for a one-to-one scenario, and in no specific order. The movie was modified slightly so that each gesture was assigned a number, which was displayed at the bottom left of the screen. The subjects, who had either little or no prior knowledge of the research, were told the type of situation within which each gesture would be experienced. For example they might be within an AVW viewed from a first person perspective, watching another person's avatar perform a number of visual behaviours within the context of chat type conversation. On questionnaires that were handed out, each subject indicated their level of experience with AVWs, and against numbers that corresponded to those on screen, wrote down what they thought each gesture was trying to convey. From the data gathered it is

possible to gain an impression of how well the intended meaning of each gesture was understood by the subjects. A gesture scores one point for each correct interpretation, therefore seven is the highest possible as this was the number of subjects, and zero the lowest. The subjects interpretations and the resultant scores are presented in Appendix E.

9.11 Outcome

Eleven out of the twenty six gestures scored six or seven points, indicating that they are effective in conveying intended meaning. Ten gestures scored four or five points, indicating that they are reasonably effective. The remaining five gestures scored between zero and three points. On closer examination of the gestures in each category, it is clear that those that score highly are generally those based on better known gestures in the real world such as *Wave close* to say hello, *Nod Head* for yes, and blowing a *Kiss*. They are easily understood even though some of them can be used within a variety of situations. In addition to these the meaning of the *neutral* pose was also interpreted correctly by all seven subjects.

The gestures that fall into the second category are *sad, clap*, *question*, *stretch* and the three laughing behaviours and the three lifelike animations. It is believed these did not score so well as they are dependent upon the conversation context, and have no clear or singular meaning outside of this. For example, the *laugh* will be easier to interpret if it is accompanied by, or is a response to, a joke in the verbal channel. Others such as the *sad* gesture were interpreted quite differently by the various subjects. In addition to its intended meaning other interpretations of the *sad* gesture, which is based on a bowing action, include showing respect, apologising and being depressed. This is actually viewed as a positive result because each interpretation is reasonable in the absence of a

conversation context (i.e. shows an understanding of the action), thus demonstrating the flexibility of the gesture for different situations.

Gestures in the lowest scoring category are *unavailable*, *smile*, *emphasis*, and *unfinished*. The *unavailable* for conversation state was only interpreted correctly by Subject 4, who was the only one who had experience of interacting within AVWs. This suggests that its slightly more abstract language (i.e. ghosting effect) makes sense to experienced users. It is suspected that *Smile* is not understood as it requires investment of time to learn its associated meanings. It is believed that results would be improved if it were experienced a number of times in various conversation contexts. *Emphasis* and *unfinished* however, appear to be of poor design and are unclear in their meaning.

An intriguing outcome is that new interpretations that have not been previously considered by the designer, were given for many of the lower scoring behaviours. Many of these make sense, at least within a certain context. For example, *tip toe* was interpreted by one subject as expressing delight or excitement. This is significant as it demonstrates that people will attach meaning to gestures when there is none obvious rather than simply render it useless. It is decided that rather than take the ineffective gestures like *tip toe* and *unfinished* and attempt to improve upon them, it is prudent to leave them as they are and let social groups negotiate their own meanings. In this way an open language may be achieved that has the potential to be flexible, to change with the various social relationships and help to define group identity through the precise meaning of certain body language codes. This implies that over time the meaning of gestures will be determined by the users. Whilst this may not effect the initial design, the criteria of success will be related to the extent to which the gesture is used. Overall however, the set conveyed intended meaning effectively.

9.20 The Final Piece

The aim of the *Final Piece* is to test the larger concept and gain an impression of how well the different elements of the visual language work together when integrated, e.g. avatar, conversation circle, set of gestures, group language, etc. This is necessary as, within the design project, these elements have been developed independently from one another. The method used is to collect a number of dialogues from 3D AVW sessions. Of these one is selected for its typical form, and the clearly identifiable social group within it which reaches a maximum of five members (a suitable number for the conversation circle which holds a maximum of six). In addition to this there are two occurrences of a newcomer joining the group, and one occurrence of a group member departing, which enables an impression to be gained of how well the new visual language supports these protocols. Based on the dialogue, a digital movie is created that reconstructs the graphical dimension using the various design solutions arrived at within the project.

The movie begins at a point where a social group has formed consisting of three members engaged in chat type conversation. Three avatars are therefore arranged in a conversation circle within a 3D Studio Max scene, which is viewed through the camera of one of the group members. The avatars' hair and clothing are varied in colour to create a sense of individuality. For confidentiality, participant nicknames have been substituted by colour names relating the colour of the avatars' shirts. In creating the animation, each text message was considered in turn and a gesture chosen to accompany it. The decision as to which gesture to use was informed by a scheme for an autonomous gesture activation that detects of keywords within verbal messages. This idea, which is also emphasised in the *Blaxxun* system described in section 3.14, begins to address the issue raised in Section 3.20 concerning the need for a better method of

controlling avatar body language. The idea here is that when a predetermined keyword is identified, an associated gesture is automatically triggered as the text is displayed. The idea employed here is different to that of *Blaxxun* in two respects (a) it is designed for the set of gestures developed in the design project and (b) it is designed for an individual who will act within a social group, i.e. a *conversation circle*. The scheme and the rules it would employ are presented in the form of a user guide in Appendix F.

The appropriate animation biped files are loaded into the corresponding avatar to create sequences of gestures in each model. When this procedure is complete, each gesture is slightly modified using the operators presented in Chapter 8 (section 8.50), so that it visually addresses the correct recipient. For the final step, text messages are added to the movie using the speech bubble solution. Through this piece of work it is possible to gain an impression of the balance of the various channels of communication on screen from the view point of one of the participants. The reader should note that the assessment is subjective of the designer and is not made purely on the completed final piece, but is ongoing throughout the process of its construction.

During construction a number of important practical issues emerged which have not previously been considered. Firstly, there was a need for an additional gesture/state which will automatically load after a participant has submitted a message addressed to the whole group such as an open question. In this situation, no individual has been specifically addressed therefore it is inappropriate for an avatar to end up gazing at a single group member. The solution is to create a new animation based on the neutral pose, but with the head and waist subtly oscillating from side to side to distribute the gaze around the group, thus removing attention from any one member. This has the added effect of making the neutral state an *attentive* pose. Secondly, it has also been

necessary to create an additional rule: while in the *open* state, the avatars head will automatically rotate towards anyone who says their name. This rule holds until the participant explicitly names another group member at which point his or her avatar will turn to gaze toward the recipient and the usual rules are reinstated. Lastly, it is necessary to consider a rule to determine what happens to the avatars' gaze when a participant leaves the group. In the *Final Piece* all other group members are looking at *Orange* when he departs. When he has left their gaze is fixed on empty space which breaks the flow of social interaction. The solution proposed is for the remaining avatars to switch into the new *open* state. From this point participants are able to pick up existing, or begin new, conversations.

When adding the text messages to the movie, the designer became concerned that there was too much information on the screen to take in, particularly when the *conversation circle* nears it maximum capacity. In the interest of design elegance, it is necessary to find a way of reducing the clutter, which will help to create a more relaxed experience for the users. At this point in the project audio became a practical possibility. It also became clear that the best way to obtain this goal is to use audio rather than text to achieve verbal discourse. This, however is a considerable step to make as audio presents an entirely different set of challenges, but makes sense as the channel is currently uninvolved in the multi-modal communication. The text messages are therefore removed from the piece and replaced with audio ones to improve the visual experience and to gain an understanding of the issues involved. So that it is clear who is speaking, the surrounding environment is darkened and a spot light illuminates their avatar. This solution is appropriate for the work described within the design project as it extends the communicative role of lighting which already plays a part within the visual language. An additional feature borrowed from *Onlive Traveller* is proposed,

which is to use volume levels and the stereo channels to create a sense of spatial positioning.

Before continuing with the text, the reader is asked to view the following file which is located on the accompanying CD ROM. The reader should note however, that if running the CD ROM browser from the CD ROM drive and not the computer hard drive, the reader should refer to the VHS video cassette version instead.

Demo9-0: Final Piece

Description: The piece shows three group members who are engaged in dialogue within a *conversation circle*. The scene is viewed through the camera of one of the participants so the reader is able to experience what a participant would see. Two newcomers join the group in turn and participate in dialogue. After a duration of message exchange one group member departs at which point the movie ends.

Demonstrates: The integration of the various components of the visual language designed, i.e. the avatar, conversation circle, set of gestures and group language, etc. It also shows some new language elements such as the spotlight effect to show who is speaking and the *open* state.

Points to note: As the piece was originally designed for use with text messages, the movie appears to run slowly when audio is used. There is also no message overlap which one would expect to see in an audio system. This will however, help the reader to follow what is happening. Also, audio is mono in this particular piece and nicknames are not visible so refer to shirt colour to identify participants.

Instructions: Null

9.21 Final Piece Outcomes

While constructing the *Final Piece* it became apparent that the various channels of communication were not well balanced. The textual information cluttered the screen and distracted too much attention from the others, making the switch to audio discourse a necessary step. This has raised two issues. Firstly, there needs to be a way of illustrating who has uttered which message. This particular issue is addressed within

the *Final Piece* using a spotlight effect but other methods may need to be considered for other applications, e.g. nodding heads or lip synchronisation as used by *Onlive Traveller*. Secondly, in the piece there is no message overlap, however it raises the question, what happens if everybody keeps speaking at the same time? In this situation it is likely that messages will become difficult to hear. One idea to address this problem is to locally process each message's volume so that a recipient only clearly hears those addressed to him or her, while others are quietened.

There are a number of additional practical observations that are worth documenting. Firstly, the designer feels that the spotlight effect needs to be made more visible. This can be achieved by reducing the background lighting and increasing the spotlight's intensity so that greater emphasis is put on the speaker. Secondly, although the movements of the gestures were exaggerated to make them more obvious, they still appear to be too subtle in a group situation. It may be advantageous to further exaggerate movements in order to make them more visible. Another good idea, although somewhat problematic, is to provide a clue that something is about to happen. This will help attract the other group members' attention and induce a level of anticipation. One way to achieve this would be to programme an avatar's spotlight to switch on a short time before a message is displayed on screen. The problem with this is how it is implemented, if the server delays a message to provide enough time for a warning, then lag will occur inhibiting conversation flow. It is therefore necessary to come up with a method of detecting a message prior to its arrival which is equally challenging.

9.30 Summary

Two aspect of the design work have been evaluated. The first aspect is to gain an understanding of how well the meaning of each gesture is understood. This was achieved by showing the gestures in turn to a group of subjects who wrote down what they thought each one was intending to convey. While not all gestures were interpreted correctly by the participants, overall the gestures proved to be effective in conveying intended meaning. We also evaluated was how well the different elements of the visual language work together when integrated (e.g. avatar, conversation circle, set of gestures, group language, etc.). To achieve this, the *Final Piece* movie was constructed around an actual AVW dialogue. The graphical dimension was created using the various design solutions arrived at within the project during its construction. It was found that the textual channel attracted too much attention from the other channels, therefore a move was made to audio verbal discourse. This generated new design requirements, such as how to indicate who is speaking and what happens if everyone speaks at once. Overall, however, it is conceived the visual outcome of the final piece is sufficiently satisfactory and it is anticipated that participants would find that the visualisations enhance the experience of interacting within an AVW.

Chapter 10:

Conclusions

Chapter Overview

This chapter concludes the thesis. Firstly, the main outcomes of the work contributing to knowledge are summarised. This includes the research techniques and findings relevant to the field of study. Secondly, there is discussion of the original contribution in relation to the initial research question. Lastly, suggestions are made for future research projects to carry the work forward.

10.00 Summary of Outcomes

The initial rationale for the research was based on the observation that despite the graphical dimension of 3D AVWs (avatar virtual worlds), communication within them is almost entirely limited to the exchange of verbal messages. It was argued that if the graphical dimension is to become useful within the context of social communication we must address the poor design of visual channels such as body language. To investigate the potential of NVC (non verbal communication) in AVWs, a design project was embarked upon to develop an avatar that exploits the communicative power of body language within the context of social interaction. Thus, the research was concerned with the loosely defined question: how can multi-modal social communication that exploits the rich potential of body language, be achieved in 3D AVWs?

Contribution to knowledge has been made in a number of ways. The research has increased understanding of social interaction within virtual worlds. This was achieved through two studies conducted as preliminary research. The first study was designed to increase understanding of the patterns of behaviour in MUD environments. As the only medium of communication in a MUD environment is text messaging, the investigation of social interaction involved the analysis of social conversation. Two dialogues were

selected and analysed at both the word and message level. To aid the message level analysis, a linear diagrammatic technique was developed to represent the structure of conversation. This allowed the identification of where each conversation began and ended, who was talking to whom, how many conversations were taking place at any one time and how the conversations were intertwined. This enabled the identification of patterns of behaviour which subsequently led to an understanding of a number of AVW social conventions, e.g. what happens when a participant wishes to leave a conversation.

The analysis suggested two main aspects of virtual world social communication which would benefit from the development of NVC. Firstly, the devices by which participants augment discourse to convey actions and expressions, i.e. acronyms, punctuation marks and performative words. For example, gestures to supplement performative words (e.g. *wave*) would be used. Secondly, the social conventions at the three stages of social interaction (i.e. start, middle and end) would be supported. For example, NVC would be used to signal an a wish to end the conversation.

In the second study social interaction was modelled in order to establish a finite set of behaviours that would be adequate for engaging in social interaction. The model was developed using a diagrammatic technique known as *state-space* which was borrowed from the field of expert systems design. First a model of one-to-one scenario was developed and operated at two levels. The top level represented conversation as three consecutive stages, i.e. start, middle and end, and a conversation was conceived of as a game in which the participants must get from the first stage through to the last in order to achieve a meaningful experience. At the second level, each stage was broken down into separate diagrams where the possible states of each player were drawn out as a node tree in a move for move type order. The nodes were connected by arcs

representing the behaviours that change the state of the players. Later, a third participant was introduced to model group interaction. Unlike the work of the first study, the model was developed with reference for dialogue and adopted a scenario based approach taking a user perspective. Additionally, the study led to the identification of a number of specific behaviours that are likely to be useful to participants during social interaction. In contrast, the first study provided a more general understanding of the social dynamics of virtual words.

In order to explore the practical value of the model, design work was undertaken in which each of the possible pathways were played out. This enabled an impression to be gained of whether or not the set of derived behaviours were adequate in practice for social communication. Following an assessment of the necessary resources to complete this task, it was decided to implement the model in 2D first. The benefit of doing this was that time was saved when constructing the 3D versions. For example, an understanding was achieved of the minimal set of facial operators necessary for an avatar to convey the required meaning. It was argued that this would have taken much longer to determine in 3D because of the greater time required to model and experiment in 3D. The subsequent 3D versions suggested that the set of behaviours was complete for conversation purposes and additionally, a number of important practical issues were identified, such as the need to consider scenic properties such as camera position and the potential of group NVC.

Informed by the knowledge gained from these studies, a series of creative works addressed the two tasks defined by the design brief, i.e. (a) to create a framework for multi-modal communication that integrates the various channels of communication, i.e. text, image and audio, in which group members are visible to one another, (b) to design

an avatar capable of performing body language and a set of behaviours supporting group social interaction and verbal discourse.

To address the first task, a concept called the *conversation circle* was proposed. Under this concept, avatars are gathered together in a social group and are automatically arranged in a circular formation in which each participant faces the centre. The goals of the solution were initially two-fold. Firstly, to make group members visible to one another so that body language can be seen. Secondly, to create a predictable and consistent screen image which is integrated with the textual channel, thus eliminating the need for a separate dialogue window. The *conversation circle* satisfied these goals. In addition to this, it contributed a number of other useful features. Firstly, it enabled group NVC signalling what groups exist, where they are located, who the members are, when there is an arrival or departure, whether a group is full and when a group has dispersed. Secondly, it enabled gaze behaviour (developed later in the project) to show who was talking to whom. Lastly, it made it possible for body language to be used to provide feedback to other participants. It was also argued that the sense of being within a conversation for individuals will be enhanced thereby increasing the obligation to contribute to the dialogue, thus facilitating communication.

An initial library of two states and twenty-four gestures was proposed derived from the preliminary studies and implemented through creative practice. The set is intended to complement verbal discourse and support the three stages of group interaction, i.e. start, middle and end. In order to implement the gestures an avatar that mimics the human form (but does not strive for photo-realism) was constructed with a set of operators necessary to carry meaning. These were informed by a number of design considerations drawn up early in the project. For example, the decision was reached that the avatar

would be naturalistic in style, and therefore the operators would be based on those of the real human body. In this way the body language should be intuitive to anyone familiar with Western culture, whatever their level of experience with AVWs. A method was developed to design the visual properties of each state and gesture. Using the science of François Delsarte and the designer's personal knowledge of body language, ideas for possible solutions were first performed using a mirror, and then implemented in the avatar, *Isabel*. Development focused on the hand gesture and gaze forms of body language, although there is some facial indications, i.e. smile and eyebrow signals. Basic rules were designed to manipulate the gestures for use within a group situation, i.e. targeted at a specific member, or the whole group. The gestures were designed to work specifically within the *conversation circle*, and the language as a whole strives to be as culturally unbiased as possible.

The outcomes were evaluated in relation to two issues. First, an understanding of how effective the gestures are in conveying intended meaning was determined by showing each gesture in turn to a group of subjects who reported its meaning to them. The results were that, generally speaking, the gestures convey intended meaning, even though they were not presented in context of specific conversation. Secondly, an understanding has been achieved of how well balanced the various elements of communication are on a screen through the construction of a *Final Piece* of work that is built around a dialogue produced in an AVW session. During the construction of the piece, it became clear that the text messages would distract too much attention from the body language in an actual implementation. Consequently, audio was used to convey the dialogue rather than text. Overall, the NVC in the AVW appears well integrated leading us to conclude that AVW participants should find the proposed solution useful.

10.10 Discussion

Two aspects of the research presented contribute to the field of avatar design. The first is the new way in which virtual space is used. In existing AVWs, participants can freely navigate any part of the virtual environment and converse with others regardless of their relative locations, but space carries little social meaning. However, space plays an important role in the NVC mechanisms developed proposed here. By creating a locale where avatars have to be arranged in circular formations before they can communicate fully with others, the various social groups that exist are represented distinguishable by virtue of their location. It is expected that this will enhance the sense of being part of, or outside, a conversation, thereby heightening user's experience. The use of spatial arrangements also provides the precondition for NVC because it makes it possible for one's conversational partners to see it, which is not the case for most of the time when interacting in existing AVWs.

The second major contribution is that a category of gestures has been derived and implemented which supports verbal communication by supplementing devices used to augment discourse such as acronyms, performative words, emoticons and punctuation marks, and by complementing the various social protocols, such as those engaged in order to leave a conversation. Unlike the work of others in the field (e.g. Anthony Guye-Vuillème *et al* [Guye-Vuillème *et al.* 98]) the body language was designed for use within a social group rather than just a one-to-one situation. For example, an open question addressed to the whole group, which is a mechanism identified to start new conversations (section 5.53), can be accompanied by the *question* gesture which is manipulated by a set of predefined rules in real-time so that the avatar looks at all the other group members in turn while performing the action.

Two aspects of the design work have been assessed, i.e. (a) how well the implemented individual gestures convey their intended meaning, (b) how well the different parts of the design solution work together, e.g. proxemics, body language, group NVC, etc. However, while the outcomes of the evaluation conducted have value, one limitation of the design project is that that it will only be possible to fully validate the design solutions when they are implemented in a complete AVW system. If such a task was undertaken, a group of twenty to thirty subjects might be arranged consisting of some experienced virtual world users, people who know each other before hand and others who are total strangers. Using the AVW created, the subjects would be free to form social groups (conversation circles) and engage each other in chat type dialogue over a number of sessions. By observing these sessions, interviewing the subjects afterwards and analysing digital video recordings made, it would be possible to assess a broader range of aspects such as: do participants find conversation circles and body language useful and do they enhance their overall experience; have enough or the right behaviours been derived to support group social interaction and complement verbal discourse; how well do the participants understand the meanings of the NVC designed; and how quickly do they learn non intuitive design solutions (e.g. clicking on an individual participant's avatar to initiate conversation)? Conducting this broader evaluation is a task of considerable size. It would however, make for a valuable future project that could be considered to be a second phase of the research described here.

10.20 Future Directions

In the following, suggestions are made for both modest and large scale future research projects that would take the work described within this thesis forward. These challenges are of interested to not only those working within the field of Art & Design, but also computer scientists, sociologists, game developers and educators. Modest

projects which would be direct extensions of this work are as follows. Firstly, scaling up the set of gestures which has been kept to a manageable size within the scope of the design project. This can be achieved by identifying more acronyms, emotion icons, performative words and punctuation marks, as well as extending the discourse analysis to identify further mechanisms at work during social interaction. There is a point however, where these methods will reach their limits and new ones will need to be found. Secondly, the set could be developed to include a haptics channel (body contact). In order to address this a better understanding of user's perception of personal space and body contact within AVWs will be required. Also, if designing for a world containing different forms of avatar, then there will be visual challenges such as how does a naturalistic humanoid avatar shake hands with one that is symbolic such as a saxophone. Lastly, the evaluation described within the previous section that aims to assess a broader range of aspects of the work could be carried out. This in itself would no doubt lead to a number of additional and important future directions. Suggestions for four large scale projects follow.

The design work described within the thesis has focused on developing a NVC for chat type conversation, however, other kinds of dialogue may require a different set of gestures. If this could be determined, an important future project would therefore be to investigate how the set designed could be adapted for other AVW applications such as CSCW, conferencing, religious ceremonies, learning and cognitive disabilities where the social relations between participants may be significantly different from those studied in Part 2 of the thesis. It is believed that this goal would be obtainable through different sets of gestures, configurations of avatars, ways of representing arrivals and departures, etc. The ultimate goal would be to create an avatar that is able to contribute

to communication within a variety of applications by switching between different palettes of gestures, which are specifically designed for the conversation context.

The gestural language developed within the design project is based on that of the real world, and has therefore been implemented in a humanoid avatar. An intriguing project would be to see how the list of necessary behaviours drawn up (see Chapter 8) can be implemented in other, non humanoid, avatars such as an animal, or an inanimate object such as a musical instrument. Any one who undertakes such a project needs to approach the design from a very different perspective, and it is suggested that a good place to start would be to examine the language of animated cartoons. As Norman Badler points out, cartoon characters stretch and bend in ways that defy the physical laws of our universe yet we correctly interpret their intended message [Badler *et al.* 93]. An understanding of this grammar is likely to prove valuable and a NVC based on it would have the power to be highly expressive through abstract symbolism. Research into this area will be of interest to creators of worlds whose purpose is entertainment, educating young children or role playing games.

An important future project would be to address the concern raised in Section 3.20. It is believed that one of the reasons why gestures are rarely used in existing systems, such as *Active Worlds*, is the way in which participants are required to manually activate them from tool bars and menus. This activity is very different to that of authoring verbal messages and participants therefore forget, or found it too difficult, to use the gestures available. It is likely that this problem would still exist in an AVW that uses the design solutions arrived at here. Therefore, there is a need to find a better method of controlling avatar body language if the set of gestures designed is to be used by participants. It is suggested that anyone undertaking such a project might consider

developing the automatic gesture triggering systems based on the detection of keywords within verbal messages similar to that implemented in *Blaxxun's* latest world *CyberTown* and in Hannes Vilhjálmsson's *BodyChat* [Vilhjálmsson 97]. Both of these systems are currently limited to one-to-one conversation and support a relatively small set of gestures compared to that presented here. Appendix F presents ideas, in the form of a user guide, for developing the keyword principle to support the various solutions arrived at.

Lastly, if one wishes for something more subtle, a good project would be to create an avatar that can keep track of its owner's interpersonal relationships with other users, and automatically adjust the visual properties of a gesture to reflect the particular relationship with the current conversation partner. Interpersonal relationships are affected by a number of factors such as the level of formality (e.g. one behaves differently when talking to a boss at work and to a friend at a party) and the social history of conversation partners (e.g. an argument between friends is likely to effect social behaviour on their next encounter). If a method can be found to store interpersonal relationship profiles, track social history and automatically manipulate gestures when performed in accordance to this, then user experience can be significantly enhanced through increased communication and personal expression.

10.30 Summary

The chapter has shown how the work described in the thesis contributes to the field of avatar design by: increasing knowledge and understanding of social interaction in AVWs; deriving a set of gestures supporting social interaction; designing a naturalistic avatar that implements these; and by deriving a number of scenic properties that support

the comprehension of the NVC (e.g. spatial arrangement, camera position, lighting effects, etc.).

A number of starting points for both modest and large scale future projects have been discussed, i.e. scaling up the set of gestures, implementing an AVW based on the design work to conduct a more broad ranging evaluation, adapting the NVC for other kinds of dialogue, mapping the NVC to non humanoid avatars, finding a better method of controlling body language, and creating an avatar that automatically adjusts the visual properties of gestures to reflect its owners interpersonal relationship with a conversation partner.

Appendix A: One-to-One Dialogue

One-to-One dialogue discussed within Chapter 5. The reader will note that the messages have been numbered consecutively to assist reference to them within the thesis text.

01 Pear: Apple!
02 Apple: *wave* hello
03 Pear: blinking at screen.....
04 Pear: tell a story
05 Pear: please!
06 Apple: what do you meanblinking at screen
07 Pear: that's what i was doing....
08 Apple: are U tired
09 Pear: until i figured something to say
10 Pear: maybe
11 Apple: yeh it takes me a minute or two to warm up in these caht rooms
12 Pear: ever been in here when they're jamming?
13 Apple: it takes even longer for me to spell things right
14 Pear: lol
15 Apple: no I've never been here before
16 Apple: can you actually here sound then
17 Pear: it rocks
18 Pear: i love the beta lounge
19 Apple: oh exciting!
20 Apple: so yous a reguklar visitor in these parts then?
21 Apple: by the way, interesting name
22 Apple: yo little doddy
23 Apple: sorry, doggy
24 Pear: [good to wash dishes too to]
25 Pear. [hold on my too and to are mixed wrong way round...]
26 Pear: unfortunately yes
27 Pear: i really should find something better to do with my life
28 Pear: is it extremely late where you are?
29 Pear: *bows* thank you
30 Pear: oh look. a female is wanted for anal
31 Pear: makes you feel special eh?
32 Pear: lol
33 Pear: so... no story?
34 Apple: very special, it's not late here
35 Apple: mid afternoon
36 Apple: it definately not time for bedtime stories yet
37 Apple: so do yous have one
38 Apple: or are you thinking about that female at the top of the page again
39 Pear: don't you hate those days you wake up groggy and stiff and get in the shower but it's too hot, only makes you want to spew and pass out and so you get out

40 Pear: and suffer until five and then you can eat dinner
41 Pear: do i ahve a what? dog? Cow tail...
42 Pear: ha ha ha
43 Pear: i hope no one goes
44 Pear: i should make a room
45 Apple: no I mean a story
46 Pear: "don't go to wanted f-ass for anal'
47 Pear: but then i'd have to sit in it lonely
48 Apple: but I guess you just told me one
49 Pear: oh a story....probably
50 Pear: somewhere. why?
51 Pear: *Pow* shoots her self in the shoulder. shut up Pear
52 Apple: we could make a story up ourselves
53 Apple: we take it in turns and make up one line a time
54 Pear: once i was a little boy [yourturn]
55 Apple: and you had a lovely little..
56 Pear: blackberry bush…
57 Apple: and everyday you would visit that blackbury bust to...
58 Pear: talk to all the little blackberry elves that sat on the branches and waited for the sun to....
59 Apple: rise, as thats when all the little girl sun fairies came out to play, but one day...
60 Pear: was walking to the blackberry bush singing a little song 'i'm just a little boy running in the dew' and the elves said to me...
61 Apple: hey, why don't you eat some of this magic dust and shrink to be the same size as us so that you can play with us all day, so you said...
62 Pear: "hold on. i'm not a boy! i'm a girl" and as I realized this some magic dust fell into my mouth and I shrunk...
63 Apple: thats when those lovely little elves turned out to be not so lovely, they took you to the top of the bush and tied you down to sacrifce you to the great elve god, oh no you shout, who will save me, just then....
64 Pear: I said 'god damnit. i'm not going to play this stupid game anymore. just because i'm a girl doesn't mean I have to be killed or rescued by some stud herol' and I slide down the branch, dodging the thorns
65 Apple: as you desend to the ground you see other inocent children who have been shrunk and tied up, so you set them all free by breaking of a torn to cut through the vine ties…
66 Pear: I think' i always wondered what happened to suzy and albert and colleen and amarjat' when we reach the ground we realize we're still shrunk...and our houses would take an eternity to get to..
67 Apple: the only thing to do is to bravely renter the bush to raid the elves stash of enlarging powder, so armed with a torn you galently make your way up the branches fighting the elves as you go…
68 Pear: but then a big dog named Pear comes and eats all of them up, squashing all the berries at the same time and make rivers of blackberry wine because columbo is on and Pear is making her way to her tv to watch the beloved cigar smoking detective
69 Apple: so without hesitating you jump onto to Pear's back and ride the fine beast all the way back to the house, you get home just in time to watch columbo and you live happily ever after, just a little bit small but hey, no one is perfect right

70 Pear: the end. hee hee....i gotta run Apple, this lady just attacked him or something. think it was the butler....
71 Apple: hey what a great story, as much as I am enjoying it I have to log out and get on with something but thanks for the chat,
72 Pear: yup the butler type guy, he's the one
73 Apple: B4N
74 Pear: anyways.....i'II see you around. thanks for the story

Appendix B: First Steps Dialogue

First Steps dialogue discussed within Chapter 5. The reader will note that the messages have been numbered consecutively to assist reference to them within the thesis text.

001 Kiwi: anything that comes to mind really!
002 Kiwi: Right now diet pills
003 Orange: get some gensing
004 Apple: thats something I dont know much about
005 Orange: pills
006 Kiwi: um well i stole these
007 Kiwi. neither do i Apple
008 Kiwi: its all good
009 Apple: in fact i love to eat
010 Orange: they give alot of energy but don't make you hyper
011 Orange: and are very good for ya
012 Kiwi: so do I! I only ate 1 time today or yesterday and that was like 3 pierogies
013 Orange: you can get um at health stores
014 Kiwi: thats cool! i will have to by that!
015 Orange: i got some they are cool
016 Apple: whats pierogies???
017 Kiwi: I am not fat though! you understand???
018 Orange: yes
019 Apple: so what if you are fat
020 Kiwi: they are potatoe pockets filled with like mashed potatoes! they are really good
021 Kiwi: im not! though
022 Apple: sound tasty
023 Kiwi: they are!
024 Kiwi: i love them!
025 Kiwi: Apple you got ICQ?
026 Apple: whats ICQ?
027 Kiwi: oh guess not! its like a net pager that lets you know when your friends are on! you can talk through there and send files and everything! ya want the addy?
028 Kiwi: Mr i was sendin ya a message and ya came back on! LOL
029 Apple: I only started using these chat lines yesturday
030 Apple: but I use Email
031 Kiwi: oh! welcome to my room! i am dust
032 Kiwi: would you like the addy for ICQ?
033 Orange: I got disconnected from my server
034 Apple: cool nice to meet you
035 Kiwi: i noticed!
036 Kiwi: WB
037 Kiwi: same here! age sex?
038 Kiwi: 17/f
039 Orange: thanx

040 Kiwi: welcome
041 Kiwi: Mr is my bitch! LOL
042 Apple: yopu have to read prety fast to keep up with the conversations in these rooms don't you
043 Orange: i is her bitch
044 Kiwi: not really! LOL
045 Kiwi: like in the lounges you do!
046 Kiwi: but usually not here!
047 Orange: but a dude at the same time
048 Cherry: hi Apple *wave*
049 Apple: hi Cherry
050 Kiwi: shit its like 7:05a.m here on Wed
051 Kiwi: hey Coconut
052 Coconut: hi Kiwi
053 Apple: how are you?
054 Kiwi: ROTFLMAO
055 Orange: blah blahblah
056 Kiwi: whats up?
057 Orange: rtaoh
058 Coconut: what are you talking about
059 Orange: YTOP
060 Kiwi: anything that comes to mind!
061 Kiwi: *hugs Orange*
062 Orange: TTPO
063 Cherry: what's ROTFLMAO?
064 Kiwi: 0 U 8 1 2?
065 Coconut: nothing in mind now
066 Kiwi: rolling on the floor laughing my ass off
067 Cherry: sorry I need to learn the lingo!
068 Kiwi: its cool
069 Orange: TIF
070 Apple: hi Coconut I have been in this room for 5mins and Im not sure what we are talking about!
071 Kiwi: you will learn it!
072 Cherry: LOL
073 Kiwi: TIF??
074 Orange: guess
075 Kiwi: um
076 Kiwi: ????
077 Cherry: yes, mum=mom, remeber, I'm an aussie!
078 Apple: because you are really happy
079 Orange: movie???
080 Kiwi: I LOVE YOU YOU LOVE ME
081 Coconut: film
082 Kiwi: cool! thanx Cherry
083 Cherry: Kiwi you don't get out much, do you?
084 Kiwi: film??
085 Kiwi: out?
086 Orange: i is what i is it is how it is that is

087 Kiwi: there is an outside?
088 Coconut: what?
089 Kiwi: LOL
090 Kiwi: i havent seen a good movie in like 2 weeks
091 Orange: outside??
092 Kiwi: whats that?
093 Apple: wheres outside??
094 Orange: the stuff with that big light bulb you mean?
095 Cherry: 'fraid so Kiwi! you should introduce yourself to it once in a while, it can actually be quite fun!
096 Kiwi: who knows! know that if i go outside a big crab will kill me and eat me!
097 Orange: light sux
098 Cherry: Ah yes, but you do insist on being depressed!
099 Apple: What do you do Cherry?
100 Coconut: what with social life now
101 Cherry: college, what about you Apple
102 Apple: same
103 Cherry: where do you live, anyway?
104 Cherry: what are you studying?
105 Apple: do u know englad
106 Cherry: not too well
107 Apple: I live in the south
108 Kiwi: Ok im out! itwas nice meeting all of you!!! BYE BYE!!!!!
109 Cherry: Bye
110 Kiwi: Orange i will talk to you later babe!
111 Coconut: bye Kiwi
112 Orange: seeya hun
113 Apple: art & design
114 Kiwi: see ya!!!!!
115 Kiwi: bye love!
116 Orange: byebye babe
117 Apple: B4N Kiwi
118 Cherry: cool. bye Kiwi
119 Apple: I just finished a degree...
120 Apple: my main interest was multi-media interface design
121 Apple: I am just starting a research degree
122 Cherry: ooh ahhl I study media and psychology. Does anyone know anything about statistics?
123 Cherry: shit, I'm having trouble with the typing skills, aren't I?
124 Apple: supposed to, but wouldn't like to be quizzed
125 Cherry: so you couldn't help?
126 Apple: sorry one moment AFK
127 Cherry: how you going, Coconut, you've disappeared
128 Cherry: hi Melon
129 Melon: hi all
130 Melon: hi Cherry
131 Cherry: who are you?
132 Melon: your up late
133 Melon: who me?

134 Cherry: yep!
135 Coconut: scuba age and sex
136 Apple: soory Cherry, had to stop and talk to someone on the phone
137 Melon: m 35 living in london
138 Melon: u?
139 Cherry: never mind. The outside world seeps in at times, doesn't it?
140 Coconut: woow
141 Cherry: woow what?
142 Coconut: been to the wimbledon
143 Apple: wimbledon, its done nothing but rain here for a month (UK)
144 Apple: bye Cherry
145 Coconut: LOL ! that is true but still it was nice
146 Cherry: doesn't anyone else care?
147 Apple: do you play tennis yourself Coconut
148 Coconut: Ok bye Cherry
149 Melon: you in The UK Apple?
150 Apple: attention seeking are we Cherry
151 Coconut: i wish to but no
152 Apple: yes I' in uk, devon
153 Coconut: I just watch it among other things
154 Melon: so where is everybody else from?
155 Apple: my dad was a ball boy in wimbledom when he was a lad
156 Coconut: i am from jordan
157 Coconut: nice Apple
158 Apple: what the weather like inJordan
159 Coconut: it is rather hot
160 Melon: where in Jordan
161 Apple: may be I should come and liv ewith you then
162 Coconut: amman
163 Coconut: you like the sun then
164 Apple yesy indeedy
165 Melon: near the coast?
166 Coconut: we don't have coast in amman
167 Apple: definately, i grew up on the south cost of england
168 Coconut: only in aqaba
169 Apple: spent my teenaqes on the beaches

----------Point at which all participants are booted from server -----

170 Apple: hey Coconut
171 Coconut: what happend
172 Apple: my connection went down
173 Coconut: also mine
174 Apple: how did u cal for Apple in the list of rooms window?
175 Apple: i couldn't find the dust room either where did that go?
176 Mango: whats up everybody?
177 Coconut: i created a room and named calin for Apple
178 Coconut: by the command /n
179 Apple: hey Mango

180 Mango: me too
181 Coconut: ithink it was died
182 Apple: is it easy to create a room ??
183 Coconut: I will creat one now
184 Apple: wicked
185 Apple: that a nice tip if you loss some one
186 Mango: that kicks!
187 Apple: you must be well practiced at this chat stuff
188 Coconut: that was easy no
189 Apple: what other MUDs do you use
190 Mango: this is only my second time
191 Coconut: just this one, mostly!
192 Apple: whys htat?
193 Apple: sorry, whys that?
194 Mango: it is fun
195 Coconut: i am connecting via telnet
196 Apple: whats the difference…
197 Coconut: you cant conect to some systems with, telnet, and this one rocks
198 Apple: I agree – hee hee
199 Coconut: cool
200 Berry: hello all
201 Mango: how old are you all?
202 Apple: hey Berry
203 Coconut: hi Berry
204 Berry:13
205 Apple: 24 male before you ask as every one seams to
206 Grape: Have you ever noticed that reaching our goals is never as satisfying as we had imagined? As humans) we are way to "success oriented". We put all the pressure on whether we win or lose in life. The sad part of success is that it comes with the immediate pressure of repeating your accomplishment and/or bettering yourself. The instant that we succeed in whatever we are doing is not as memorable as the struggle to reach that goal. We should learn to take pride in the struggle or the effort we put into life rather than the rewards that we give it. Anyway - wouldn't you rather be someone who did their best and failed than someone who "lucked" their way into success?
207 Berry: hey man
208 Coconut: for how don't know I am 21 m
209 Berry: yeah
210 Apple: where did that lot come from?
211 Coconut: hi again Melon
212 Berry: my teacher tells me about
213 Melon: hi guys sorry got booted
214 Apple: hey Melon a thought we lost you
215 Coconut: i don't know
216 Melon: you did!
217 Mango: Berry m\f??
218 Apple: we all got booted in dust room
219 Berry: m
220 Coconut: someone gave us a book

221 Melon: give up Mango its a sausage party in here!
222 Apple: our cool conversation must have been to much for the server to handle
223 Coconut: that what i think
224 Berry: later of to a differant room
225 Apple: Mango are you a sausge to
226 Coconut: Mango has left
227 Apple: Berrys a sasage to
228 Apple: i wish I could spell
229 Coconut: yeh
230 Berry: huh?
231 Melon: well quy's I'm gonna go walk about catch you later ok?

Appendix C: Textual Devices

Techniques used by the participants within the *One-to-One* dialogue that have the potential for increasing social interaction by compensating for the forms of non verbal communication which are missed.

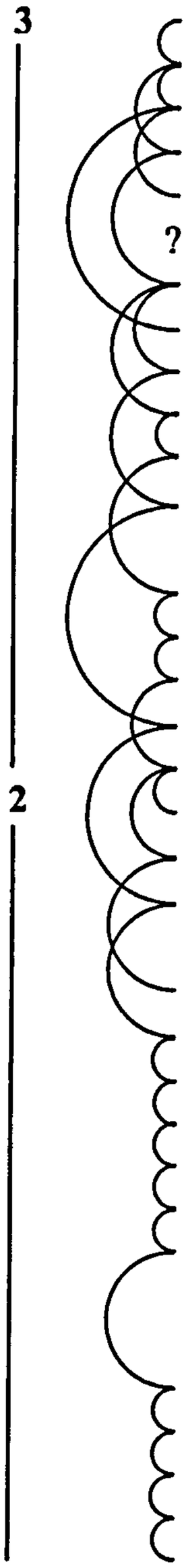
Word/Symbol	Apparent Meaning
!	Used often to indicate that a message is emphasised
wave	Describes the physical action of waving the hand
.....	Used on a number of occasions: in the middle of a message they usually represent some hesitation or delay such as might occur in speech, at the end of a message they usually imply that the message is unfinished. Within this dialogue they are also used to hand over the floor.
LOL	Acronym meaning 'Laughing Out Loud'
?	Used often, usually to request feedback
[ ]	Squares brackets used to comment implicitly
bows	Describes the physical action of bowing
Pow	Describes the sound of a gunshot
B4N	Acronym meaning 'Bye FOR Now'

Additional techniques observable in *First Steps* dialogue

Word/Symbol	Apparent Meaning
ROTFLMAO	Acronym for 'Rolling Around On The Floor Laughing My Ass Off'
CAPITAL LETTERS	Used to indicate that a verbal message is pronounced loudly
AFK	Acronym meaning 'Away From Keyboard'. Used to illustrate that a stoppage of messages is not due to technical breakdown and that the author intends to return to conversation soon.
hee hee	A representation of laughter

Appendix D: Directional Links

Illustration showing the decision made for the directional link of the first thirty two messages of the *First Steps* dialogue discussed within Chapter 5. The numbers on the left side indicate the number of participants present at any point in the dialogue.



001 Kiwi: anything that comes to mind really!
002 Kiwi: Right now diet pills
003 Orange: get some gensing
004 Apple: thats something I dont know much about
005 Orange: pills
? 006 Kiwi: um well i stole these
007 Kiwi. neither do i Apple
008 Kiwi: its all good
009 Apple: in fact i love to eat
010 Orange: they give alot of energy but don't make you hyper
011 Orange: and are very good for ya
012 Kiwi: so do I! I only ate 1 time today or yesterday and that was like 3 pierogies
013 Orange: you can get um at health stores
014 Kiwi: thats cool! i will have to by that!
015 Orange: i got some they are cool
016 Apple: whats pierogies???
017 Kiwi: I am not fat though! you understand???
018 Orange: yes
019 Apple: so what if you are fat
020 Kiwi: they are potatoe pockets filled with like mashed potatoes! they are really good
021 Kiwi: im not! though
022 Apple: sound tasty
023 Kiwi: they are!
024 Kiwi: i love them!
025 Kiwi: Apple you got ICQ?
026 Apple: whats ICQ?
027 Kiwi: oh guess not! its like a net pager that lets you know when your friends are on! you can talk through there and send files and everything! ya want the addy?
028 Kiwi: Mr i was sendin ya a message and ya came back on! LOL
029 Apple: I only started using these chat lines yesturday
030 Apple: but I use Email
031 Kiwi: oh! welcome to my room! i am dust
032 Kiwi: would you like the addy for ICQ?

Appendix E: Questionnaire Results

The table on the following page presents the subjects' interpretations from the questionnaires used to help evaluate how well the individual gestures in the set designed convey their intended meaning (see Chapter 10). Those considered to be correct, or are a reasonable interpretation, are displayed in table cells with a shaded background. The number of participants that correctly interpreted each gesture is shown as a numeric score in the right-most column.

The subjects have presented their interpretations by, a verbal translation (e.g. so what, who cares), a verbal description of the physical action (e.g. clapping), describing what they think each gesture or state is being used for (e.g, getting attention) or a combination of these. When the subjects wished to indicate that they are unsure of or have no interpretation, they either leave the space blank or place a message in brackets, e.g. (no idea).

Figure E-0: Table Showing the Interpretations of Seven Subjects Who Were Asked to Write Down What They Thought Each of the Designed Behaviours is Trying to Convey

Behaviour	Subjects Interpretation of the Intended Meaning of Each Behaviour							Score
	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	
01. Neutral	I'm listening	paying attention	giving me attention	neutral pose	ready for conversation	listening	ready	7
02. Unavailable		embarrassment		going for a coffee, busy for a minute	I've lost interest	I sleep now	I feel sick, let me out of here	1
03. Wave Far	wave hi	Ohoy there!	hello	get attention	hi / gaining attention	hello there	hello over there	7
04. Wave Close	hello	hello or good bye	wave hello	bye bye	hello - wave	hello or good bye	hello, good bye	7
05. Smile		(no idea)		taking the mick	rudeness	La peur (fear)	Maybe	0
06. Sad/Unhappy		(don't know)	sorry	sad	Oh no - depressed	respect for a person	who cares	5
07. Nod Head	agree, confirm	yes	yes, affirmative	yes	agreement - nodding	yes	yes	7
08. Shake Head	no	no	no	nope!	disagreement (no)	no	no	7
09. Uncertain/so what	Unsure or 'I don't know'	shrug, I don't know	means 'I don't know'	I dunno!	saying I've no idea	I don't know what you mean	so what, who cares	7
10. I am/myself		me, or mine	me	me	I am	myself	me, I	6
11. Sleepy	I wish to sleep soon	Time for bed, or boring	I'm tired and need to sleep	falling asleep, bored	bed	I'm off to bed, see ya	night night	7
12. Kiss	blowing a kiss	blowing a kiss	blowing a kiss	kiss you!	blowing a kiss	I give you a kiss	Blows a kiss	7
13. Chuckle	laugh	laugh a bit	laugh	Laughing	Uncertainty	do you want to go to the toilet?	"Ooh! I don't know"	4
14. Laughter	laugh more	(no idea)	laugh	LOL	very physical laugh	that a game	no, that's how I feel	4
15. Laughing Lots	laugh lots	(no idea)	more laughing	even more LOL	just jiggling	lire	laughing	4
16. Question	What do you mean?	(don't know)	(not sure)	join us, welcome	asking question	what do you think?	how about you?	4
17. Emphasis		(no idea)		so what!	body shrug	?	What me	0
18. Unfinished	Wait, hang on	some kind of emphasis		that point?	(figure trick) you		now, you just listen	1
19. Thinking	thinking	hmmmm	oh I see (in thought)	thinking	contemplation	thinking	let me think	7
20. Phone Call	on the phone in real world	(don't know)	on the phone!	talking on phone for a minute	telephoning	I listen to someone else on the mobile	talking to someone not in virtual world	6
21. Clap	clapping hands	clapping	anger	(no idea)	clap – seeking attention	applaud	listen up!	5
22. Stretch	get attention	(no idea)	I'm tired, yarning	stretching, waking up	stretch – I'm about to start moving	I want to make gym, exercise	I like exercise	4
23. Check Watch	I will be leaving soon	looking at the time – gotta go	Its time to go	time to go	Its time, look at the time	I must leave now	time to go	7
24. Side shift	feels uninvolved	(no idea)	bored	waiting/bored	whey?	I'm bored	uneasy	4
25. Arm swing	bored	(no idea)	bored	bored as well	I'm bored	It's not possible	Well!	4
26. Tip Toe	bored	(no idea)	bored	yupee!	I'm still bored	Bored again	Woo! Woops!	4

Appendix F: User Guide

A draft user guide for an avatar virtual world that uses the solutions arrived at in the design project as well as an autonomous gesture activation systems based on the detection of keywords within text messages.

Contents

Overview

The AVW you are about to enter is a 3D digital space accessible through the Internet, where you can meet and communicate with others in real time through the exchange of text messages and virtual body language. Once logged in, you are able to move around to find someone, or a group of people, to talk with. However, to be able to fully communicate with others, you must form, or join, a conversation circle. These are circular formations of avatars which allow communication to take place. They are formed when one visitor approaches another who agrees to start a conversation, and they form at the location of the recipient. Other visitors can join the conversation circle which automatically expands until it reaches its maximum capacity of six. They are dispersed when the last two members decide to end the conversation.

Conversation circles are recognisable by the exact circular arrangement of the avatars, lit avatar faces, bright spot on the floor and shadows which are only visible for group members avatars (Figure F-0). Note that if a conversation circle is full, the lit faces and spot on the floor turn red in colour to indicate that it is closed to newcomers.

PAGE NUMBERS CUT OFF

IN

ORIGINAL

Figure F-0: Screen Shot of Five Avatars Within a Conversation circle

When interacting within a conversation circle, there are twenty three gestures that can be performed to complement verbal discourse. Unless instructed otherwise, the computer will attempt to determine one appropriate gestural complement for each message you write, which is triggered by the detection of keywords within it, for example, "hello" will activate a wave gesture (see Controlling Body Language). If there is no appropriate gesture available then none will be performed. It is possible to turn this feature on and off, customise the keywords that trigger each gesture and override the computer's decision for individual messages. Each of these gestures, which are specifically designed to support chat-type dialogue, can be manually activated at any time using a graphical tool bar. In addition to this the computer will track who is talking to whom through the detection of participant names within messages and turn your avatar's head to face who you are speaking to.

Moving Around

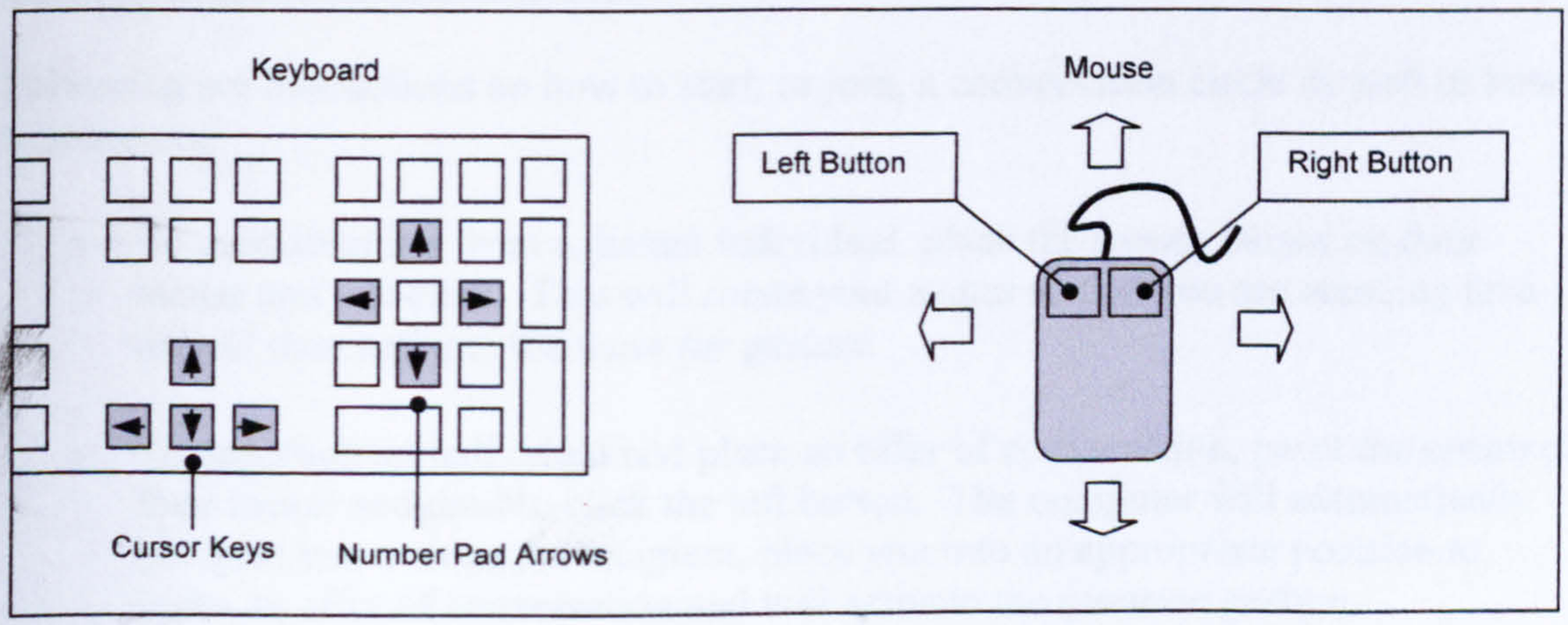


Figure F-1: Diagram of Key Board Cursors and Mouse Input Devices

Navigate the world by using the keyboard cursor keys or the mouse, to walk forwards, backward and turn left or right. These controls are shown in Figure F-1 and will be

PAGE NUMBERS CUT OFF

IN

ORIGINAL

referred to throughout the guide. Within the text, left-click will refer to clicking the left mouse button, and right-click refers to clicking the right.

Changing States

When you are not part of a conversation circle you can choose between two general states which are *available* and *unavailable* for conversation. The latter may be used to reject a conversation offer or to indicate to others that, for the time being, you do not wish to talk. It may also be used within a conversation circle to indicate to others that you have been temporarily distracted in the real world but wish to remain part of the group and will return soon. To change state, right-click on your own avatar to call up a contextual menu and select the desired state (Figure F-2). Note that the menu has a number of other functions which will be discussed later (i.e. keywords, scanning and leave circle).

Figure F-2: Screen Shot of Contextual Menu

Getting In And Out Of Conversation

Following are instructions on how to start, or join, a conversation circle as well as how to leave one:

- To gain attention from a distant individual, place the mouse cursor on their avatar and left-click. This will rotate your avatar so that you are standing face on and then activate the *wave far* gesture.

- To approach an individual and place an offer of conversation, point the cursor at their avatar and double-click the left button. The computer will automatically navigate you over to the recipient, place you into an appropriate position to make an offer of conversation and will activate the *question* gesture.

- To accept an offer of conversation, point and left-click on their avatar. To reject it either do nothing, tell them verbally, change to *unavailable* for conversation state, turn or move away.

- To join a *conversation circle*, point the cursor over the desired group and double-left-click. The computer will automatically navigate you over to the *conversation circle* and place you along its circumference providing it is not full. To welcome a newcomer to the group, either greet them verbally, e.g. "hello Paul", ask a question, e.g. "where are you from Paul?" or manually activate a gesture, e.g. smile or wave (*see Controlling Body Language*).

- To leave a *conversation circle*, select *leave circle* from the contextual menu shown in Figure F-2 or hold the downwards keyboard cursor (walk back key when in navigation mode) for approximately two seconds. This will instruct the computer to automatically make your avatar step backwards two paces, and rotate it away from the group. Note that the other cursor keys have no effect when interacting within a *conversation circle* as navigation is not possible. However they will become fully operational when you leave it.

Camera Control

You can change the camera position when you are within a conversation circle. This enables you to decide which avatars are most visible, as well as to scan the surrounding environment for other social activity without leaving the group. The camera can move between three height levels which are eyelevel, default and birds eye. Eye level reveals more of the avatars faces and also the distant environment. Default shows the whole group and some of the surrounding environment. Birds eye shows the whole group and all of the immediate environment. The camera can be rotated about the group at any of these levels but will always face the centre of the circle. To control the camera when within a conversation circle use the keyboard number pad arrows (Figure F-1).

- Press the up-arrow to move the camera closer to the group and lower to the ground.

- Press the down-arrow to move the camera away from the group and further from the ground.

- Press the left- or right-arrows to rotate the camera around the group.

Controlling Avatar Body Language

There are three main types of body language which your avatar can exploit which are lifelike actions, hand gestures and gaze behaviour. The lifelike actions are animations that will occur randomly only when you are not participating in a conversation circle or if you are but have not contributed to verbal discourse for some time. Their purpose is to help to increase the sense of life in the AVW and they may encourage other members to include you in conversation when part of a group. It is not possible to control these gestures, however, they will not be performed when you are actively participating in discourse to avoid conflicting with your message content.

The twenty three available hand gestures are specifically designed to complement verbal discourse and unless specified, are activated automatically on the detection of keywords and text symbols within messages. The table shown in Figure F-3 shows the available gestures that can be used within a conversation circle and their corresponding

default triggers. The last three gestures in the table are designed to indicate that you intend to leave the conversation soon, therefore you need to consciously decide when it is time to use them. There is one trigger text symbol that will activate any of the three actions. When it is detected the computer will randomly decide which to use. This helps avoid monotony. A priority system is in place to decide which gesture will be performed when there is more than one trigger word or symbol within a message. The priority of a gesture can also be changed in the customise trigger keyword window (see Keyword Window).

Gesture		Trigger Word/Text Symbol
	***Wave Close***	hello, hi, *wave*, bye, good bye, CU, CUL8R, B4N
	***Smile***	*Smile*, :)
	***Sad***	:(, *sad*
	***Nod Head***	yes, yeh, ok, agree, why not
	***Shake Head***	no, disagree
	***Uncertain/so what***	I don't know, maybe, so what
	***I am/myself***	I, I am, me, my, myself
	***Sleepy***	Tired, sleep, sleepy
	***Kiss***	*Kiss*, kisses, :-x
	***Chuckle***	Hee hee, ha ha, *laughing*
	***Laughter***	LOL
	***Laughing Lots***	LOTFL
	***Question***	what, why, when, where, would, who, how, ?
	***Emphasis***	!, wow, (CAPTIAL LETTERS)
	***Unfinished***	.....
	***Thinking***	think, thinking, um
	***Phone Call***	AFK, BBS
	***Clap***	> (type at start of message), excellent, well done
	***Stretch***	> (type at start of message)
	***Check Watch***	> (type at start of message), the time is

Figure F-3: Table Showing the Available Gestures When Within a Conversation Circle and Their Corresponding Default Trigger Words and Text Symbols Separated by Commas

PAGE NUMBERS CUT OFF

IN

ORIGINAL

Gaze behaviour is used to indicate who is talking to whom. To determine this, the computer scans the text of your messages for the nicknames of other group members (or the first three letters of them). On detection of a name, the computer will automatically turn your avatar's head to look at the recipient, unless you are already gazing at them, before performing an appropriate gesture. If a message contains more than one nickname, then the avatar will address each of them in turn. To save typing everybody's name, the computer also looks for symbols to know when you are addressing the whole group. If no name is detected then your avatar's gaze remains fixed on the last recipient and any gestures are performed for them. To control your avatar's body language, use the following techniques:

- to trigger a *hand gesture*, include any of the text symbols shown in the table of Figure F-3 within a text message, e.g. "**hello**"

- to address a message to an individual, include their nickname (or first three letters of) within a text message, e.g. "**hello Paul**".

- to address a message to the whole group, type either *everyone*, *everybody*, *all* within a message or ' / ' at the end, e.g. "**hello all**".

- to activate a gesture without typing a message, move the cursor to the top of the screen. This will make a tool bar visible that contains icon buttons for each of the twenty three available gestures (Figure F-4). This bar fades away when you have selected a gesture from it. Alternatively, type the corresponding trigger word and the symbol that instructs the computer to perform the gesture, but not display the text (see next section).

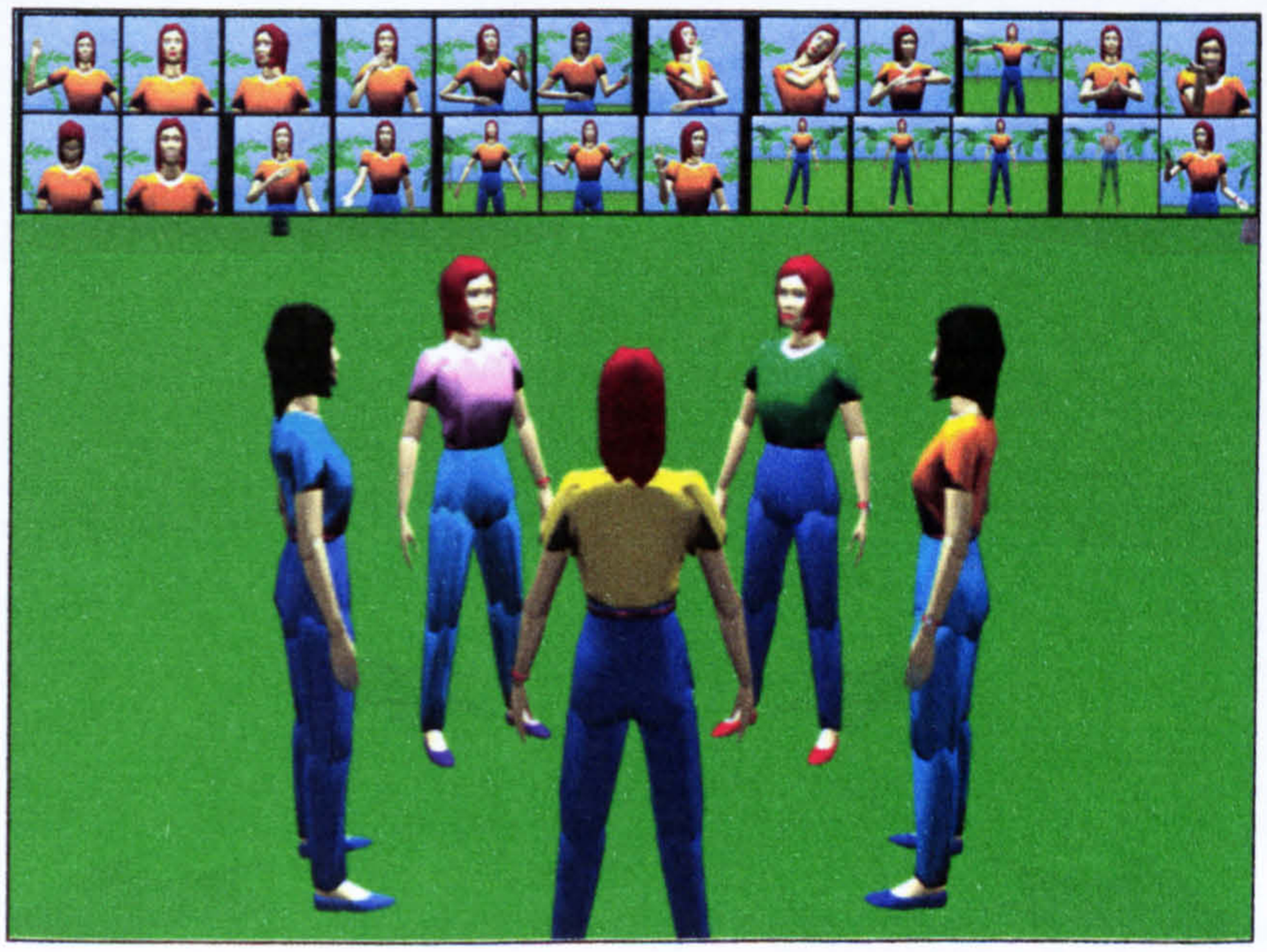


Figure F-4: Screen Shot of Gesture Tool Bar Made Visible When Mouse is Rolled to the Top of the Screen

- Use the *store* and *load* commands to save keyword trigger lists for use in different situations, i.e. formal or chat type conversation.

- Use the *priority* command to call up a list of all keywords and assign either *high*, *medium* or *low* settings for each.

- Click the *default* button to restore the original keyword settings.

- To close the keyword window, click the  in the top right of the window.

References

[Anderson & McGrath 97]

Anderson, B., and McGrath, A. (1997) Strategies for Mutability in Virtual Environments. *Proceedings of BCS Conference: Virtual Environments on Internet, WWW, and Networks*, National Museum of Photography, Film & Television, Bradford, UK, April 14th-17th.

[Argyle *et al.* 73]

Argyle, M., Ingham, R., Alkena, F., and McCallin, M. (1973). The Different Functions of Gaze. *Semiotica*, Vol. 7, pp. 10-32.

[Badler 99]

Badler, N. (1999) Virtual Humans. *Course Notes One: International Conference Virtual Humans: Towards Interactive Actors*, National Museum of Photography, Film & Television, Bradford, UK, 13th-15th April.

[Badler *et al.* 93]

Badler, N., Philips, C., and Webber, B. (1993) *Simulating Humans: Computer Graphics Animation & Control.* Oxford University Press, New York.

[Badler *et al.* 99]

Badler, N., Chi, D., and Chopra, S. (1999) Virtual Humans Animation Based on Movement Observation and Cognitive Behaviour Models. Course Notes: *Virtual Humans: Towards Interactive Actors,* International Conference: From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks, National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Ball *et al.* 97]

Ball, G., Ling, D., Kurlander, D., Miller, J., Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Dantzich, M., and Wax, T. (1997) Lifelike Computer Characters: The Persona Project at Microsoft. In J.M. Bradshaw (Ed.), *Software Agents*, AAAI/MIT Press, Menlo Park, CA, pp. 191-222.

[Barr & Feigenbaum 86]

Barr, A., and Feigenbaum, E (1986) *The Handbook of Artificial Intelligence.* Addison-Wesley, Reading, MA, Vol. 1, pp. 32-45.

[Bates 94]

Bates, J. (1993) *The Nature of Character in Interactive Worlds and The Oz Project.* Technical Report CMU-CS-92-200, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

[Bavelas & Chovil 97]

Bavelas, J. B., and Chovil, N. (1997) Faces in dialogue. In J. Russell & J.-M. Fernandez-Dols (Eds.), *The Psychology of Facial Expression*, Cambridge, U.K., Cambridge University Press, pp. 334-346.

[Becheiraz & Thalmann 96]

Becheiraz, P., and Thalmann, D. (1996) The Use of Nonverbal Communication Elements and Dynamic Interpersonal Relationship for Virtual Actors. *Proceedings of Computer Animation '96*, IEEE Computer Society Press, June 1996, pp.58-67.

[Becker & Mark 98]

Becker, B., and Mark, G. (1998) Social Conventions in Collaborative Virtual Environments. *Proceedings of CVE 98*, Manchester, UK, 17th-19th June, pp. 47-55.

[Becker 97]

Becker, B. (1997) Virtual Identities: The Imaginary Self. *Proceedings of the 6th Cyber Conference*, Oslo, Norway, June 5th-8th.

[Benford *et al.* 97]

Benford, S. D., Bowers, J. M., Fahlen, L. E., Greenhalgh, C. M., and Snowdon, D. N. (1997) Embodiments, Avatars, Clones and Agents for Multi-user, Multi-sensory Virtual Worlds. *Multimedia Systems*, Vol. 5(2) Springer-Verlag, pp. 93-104.

[Blumberg & Galyean 95]

Blumberg, B., and Galyean, T. (1995) Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. *Proceedings of SIGGRAPH 95*, Annual Conference Series, Los Angeles, California, August 6th-11th, pp. 47-54.

[Bowers *et al.* 96]

Bowers, J., Pycock, J., and O'Brien, J. (1996) Talk and Embodiment in Collaborative Virtual Environments, *Proceedings of CHI 96*, Vancouver, Canada, ACM Press, April 13th-18th, pp. 58 -65

[Boyd-Davis *et al.* 96]

Boyd-Davis, S., Huxor, A., and Lansdown, J. (1996) *The Design of Virtual Environments with particular reference to VRML.* Technical Report released by the Centre for Electronics Arts, Middlesex University.

[Brown 98]

Brown, J. (1998) Punching into Life Online. *Wired Magazine*, October 1998, Accessed July 2001, http://www.wired.com/news/print/0,1294,7951,00.html.

[Bruckman 93]

Bruckman, A. (1993). Gender swapping on the Internet. *Proceedings of INET'93*, Reston, VA: The Internet Society, San Francisco, California, August 1993.

[Bruckman 94]

Bruckman, A. (1994). Programming for Fun: MUDs as a Context for Collaborative Learning. *Proceedings of National Educational Computing Conference*, Boston, MA, June 1994.

[Çapin *et al.* 97]

Çapin, T. K., Pandzic, I. S., Magnenat-Thalmann, M., and Thalmann, D. (1997) Realistic Avatars and Autonomous Virtual Humans in VLNET Networked Virtual Environments. *Proceedings of International Conference: From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks*, National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Carlsson & Hagsand 93]

Carlsson, C., and Hagsand, O. (1993) DIVE: A Platform for Multi-User Virtual Environments, *Computers and Graphics*, Vol. 17(6), pp. 663-669.

[Chovil 92]

Chovil, N. (1992) Discourse-Oriented Facial Displays in Conversation. Research on Language and Social Interaction, Vol. 25., pp. 163-194.

[Clark & Clark 97]

Clark, C., and Clark, A.F. (1997) VRML Interfaces to Information Systems. *Proceedings of From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks International Conference*, National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Curtis 92]

Curtis, P. (1992). Mudding: Social Phenomena in Text-Based Virtual Realities. *Proceedings of Conference on Directions and Implications of Advanced Computing*, Berkeley, May 1992.

[Damer 96]

Damer, B. (1996) Inhabited Virtual Worlds. *ACM Interactions*, Sept-Oct, Vol. 3(5), pp. 27.

[Ekman & Friesen 75]

Ekman, P., and Friesen, W. (1975) *Unmasking the Face.* Prentice-Hall, INC., New Jersey.

[Fisher 92]

Fisher, J. (1992) *The Theatre of Yesterday and Tomorrow: Commedia Dell'Arte on the Modern Stage.* The Edwin Mellen Press, Lewiston/Queenston/Lampeter.

[Fujii *et al.* 95]

Fujii, T., Imamura, K., Yasuda, T., and Yokoi, S. (1995) A Virtual Scene Simulation For City Planning. In Earnshaw, R. A., and Vince, J. A. (Eds.), *Computer Graphics, Developments in Virtual Environments*, Academic Press, London, pp. 483-496.

[Gibson 84]

Gibson, W. (1984) *Neuromancer.* Harper Collins Publishers, Glasgow.

[Guye-Vuillème *et al.* 98]

Vuillème, A., Capin, T., Pandzic, I., Magnenat-Thalmann, N., and Thalmann, D. (1998) *Nonverbal Communication Interface for Collaborative Virtual Environments.* Proceedings of CVE 98, Manchester, 17th-19th June 1998.

[Hafner 97]

Hafner, K. (1997) The Epic Saga of The Well. *Wired Magazine*, May 1997, pp. 98-142.

[Hayes-Roth *et al.* 99]

Hayes-Roth, B., Johnson, V., Van Gent, R., and Wescourt, K. (1999) Staffing the Web With Interactive Characters: Intelligent Agents in the Form of Personable Characters Interact With and Serve Human Customers. *Communications Of The ACM*, Vol. 43(3), pp. 103-105.

[Heiskanen *et al.* 98]

Heiskanen, K., Jäkälä, M., and Pekkola, S. (1998) Communicating Through Virtual Reality: Visualisation in Collaborative Virtual Environments. In Rasmussen, L.B., Beardon, C., and Munari, S. (Eds.), *Computers & Networks In the Age of Globalisation*, Kluwer Academic Publishers, Boston, pp. 231-244.

[Hertz 92]

Hertz, K. (1992) A Coherent Description Of The Process Of Design. *Design Studies*, Vol. 12(4), pp. 392-410.

[Huxor 98]

Huxor, A. (1998) An Active Worlds Interface to BSCW, to Enhance Chance Encounters. *Proceedings of CVE 98*, Manchester, UK, pp 87-92, 17th-19th June.

[id 99]

id Software (1999) *Quake 3.*

[Jackson & Fagan 00]

Jackson, R., and Fagan, E. (2000) Collaboration and Learning Within Immersive Virtual Reality. *Proceedings of CVE 2000*, San Francisco, USA, September 10th-12th, pp. 83-92.

[Jackson 86]

Jackson, P. (1986) *Introduction to Expert Systems*. Addison-Wesley Publishing Company, Wokingham, England.

[Jones 00]

Jones, M. (2000) Collaborative Virtual Conferences: Using Exemplars to Shape Future Research Questions. *Proceedings of CVE 2000*, San Francisco, USA, September 10th-12th, pp. 19-27.

[Kalra *et al.* 99]

Kalra, P., Magnenat-Thalmann, N., Moccozet, L., Sannier, G., Aubel, A., and Thalmann, D. (1999) Real-Time Animation of Realistic Humans. Course Notes: *Virtual Humans: Towards Interactive Actors*, International Conference: From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks, National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Kurlander *et al.* 96]

Kurlander, D., Skelly, T., and Salesin, D. (1996) Comic Chat. *Proceedings of Computer Graphics '96*, Annual Conference Series, SIGGRAPH, New Orleans, Louisiana, August 4th-9th.

[McCarthy *et al.* 98]

McCarthy, L., Stiles, R., Johnson, L., and Rickel, J. (1998) Enabling Team Training in Virtual Environment. *Proceedings of CVE 98*, Manchester, UK, 17th-19th June 1998.

[Lawson 90]

Lawson, B. (1990) *How Designers Think: The Design Process Demystified*. 2nd Edition, Butterworth Architecture, London, pp. 22-29.

[Machado *et al.* 00]

Machado, I., Prada, R., and Paiva, A. (2000) Bringing Drama into a Virtual Stage. *Proceedings of CVE 2000*, San Francisco, USA, September 10th-12th, pp. 111-117.

[Maes 94]

Maes, P. (1994) Agents that Reduce Work and Information Overload, *Communications of the ACM*, Vol.37(7), pp. 23-29.

[Magnenat-Thalmann & Kalra 95]

Magnenat-Thalmann, N., and Kalra, P. (1995) *The Simulation of a Virtual TV Presenter*. Accessed July 2001, http://www.miralab.unige.ch/ARTICLES/PG95.html.

[Magnenat-Thalmann *et al.* 95]

Magnenat-Thalman, N., Pandizic, I., Moussaly, J., Thalmann, D., Huang, Z., and Shen, J. (1995) The Making of the Xian Terra-Cotta Soldiers. In Earnshaw, R. A., and Vince, J. A. (Eds.), *Computer Graphics, Developments in Virtual Environments*, Academic Press, London, pp. 281-296.

[Magnenat-Thalmann *et al.* 96]

Magnenat-Thalmann, N., Carion, S., Courchesne, M., Volino, P., and Wu, Y (1996) Virtual Clothes, Hair and Skin for Beautiful Top Models. *Proceedings of Computer Graphics International '96*, Pohang, Korea, pp.132-141.

[Mateas 97]

Mateas, M. (1997) *An Oz-Centric Review of Interactive Drama and Believable Agents*. Technical Report CMU-CS-97-156, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

[Mitchell 97]

Mitchell, W.L. (1997) Moving the Museum onto the Internet: The Use of Virtual Environments in Education About Ancient Egypt. *From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks,* National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Morningstar & Farmer 92]

Morningstar, C., and Farmer, R. (1992) The Lessons of Lucasfilm's Habitat. In Benedikt, M. (Ed.) *Cyberspace: First Steps*, MIT Press, Cambridge, Mass.

[Morris 67]

Morris, D. (1967). *The Naked Ape.* Vintage, UK.

[Nielsen 1999]

Nielson, J. (1999) *Designing Web Usability*. New Riders Publishing, USA.

[Nihilistic 00]

Nihilistic Software, Inc. (2000) *Vampire: The Masquerade – Redemption*.

[Norman 94]

Norman, D. (1994) How People Might Interact with Agents. *Communications of the ACM*, Vol. 37(7), pp. 68-71.

[Noser *et al.* 96]

Noser, H., Çapin, T., Pandzic, I.., Magnenat-Thalmann, N., and Thalmann, D. (1996) Playing Games Through the Virtual Life Network.. *Proceedings Alife 5 '96*, Nara, Japan, May 16th-18th, MIT Press, 1996, pp. 114-121.

[Pandzic *et al.* 94]

Pandzic, I.., Kalra, P., Magnenat-Thalmann, N., and Thalmann, D. (1994) Real-time Facial Interaction. *Displays*, Vol. 15(3), pp. 56-62.

[Parise *et al.* 96]

Parise, S., Kiesler, S., Sproull, L., and Waters, K. (1996) My Partner is a Real Dog: Cooperation with Social Agents. *Proceedings of CSCW 96*, Boston, MA, USA, 16th-20th November, pp. 399-408.

[Pausch *et al.* 96]

Pausch, R., Snoddy, J., Taylor, R., Watson, S., and Haseltine, E. (1996) Disney's Aladdin: First Steps Toward Story Telling in Virtual Reality. Proceedings of SIGGRAPH 96, Computer Graphics Proceedings, Annual Conference Series, New Orleans, Louisiana, August 4th-9th, pp. 193-203.

[Rheingold 94]

Rheingold, H. (1994) *The Virtual Community*, Secker & Warburg, London.

[Riecken 94]

Riecken, D. (1994) A Conversation with Marvin Minsky About Agents. *Communications of the ACM*, Vol. 37(7), pp. 23-29.

[Rosenberg 92]

Rosenberg, M. (1992) *Virtual Reality: Reflections on Life, Dreams, and Technology; An Ethnography of a Computer Society*. Accessed July 2001, ftp://ftp.lambda.moo.mud.org/pub/MOO/papers/ethnography.txt.

[Rowe 87]

Rowe, P. (1987) *Design Thinking*. The MIT Press, London, pp. 46-50.

[Salem & Earle 00]

Salem, B., and Earle, N. (2000) Designing A Non Verbal language For Expressive Avatars. *Proceedings of CVE 2000*, San Francisco, USA, September 10th-12th, pp. 93-102.

[Sawyer 97]

Sawyer, M. (1997) Lara Croft. *'The Face' Magazine*, Conde Nast Publishing, June, pp. 62-70.

[Selker 94]

Selker, T. (1994) Coach: A Teaching Agent that Learns. *Communications of the ACM*, Vol. 37(7), pp. 93-99.

[Shawn 74]

Shawn, T. (1974) *Every Little Movement: A Book About Delsarte*. Dance Horizons, New York.

[Slater *et al.* 00]

Slater, M., Howell, J., Steed, A., Pertaub, D-P., and Garay, M. (2000) Acting in Virtual Reality. *Proceedings of CVE 2000*, San Francisco, USA, September 10th-12th, pp. 103-110.

[Stephenson 92]

Stephenson (1993) *Snowcrash*, Roc Penguin Books, St. Ives.

[Suler 97]

Suler, J. (1997) *The Psychology of Avatars and Graphical Space in Visual Chat Environments*. Accessed December 1997, http://www.preferred.com/~eschuler/wcpublic.htm.

[Thalmann *et al.* 95]

Thalmann, N., Kalra, P., and Pandzic, I. S. (1995) Direct Face-to-Face Communication Between Real and Virtual Humans. *International Journal of Information Technology*, Vol.1(2), pp.145-157.

[Thomas & Johnston 81]

Thomas, F., and Johnston, O. (1981) *Disney Animation: The Illusion of Life.* Abbeville Press, New York.

[Thórisson 94]

Thórisson, K. (1994) Face-to-Face Communication with Computer Agents. Working Notes, *AAAI Spring Symposium on Believable Agents*, Stanford University, California, August 13th-16th, pp. 86-90.

[Tuomola 99]

Tuomola, M. (1999) Drama in the Digital Domain: Commedia dell'Arte, Characterisation, Collaboration and Computers. *Digital Creativity*, Vol. 10(3), pp. 167-179.

[Turkle 95]

Turkle, S. (1995) *Life on the Screen: Identity in the Age of the Internet.* Simon and Schuster, New York.

[Turkle 96]

Turkle, S. (1996) Who Am We? *Wired Magazine (4, 1)*. Accessed July 2001, http://www.wirednews.com/wired/archive//4.01/turkle_pr.html.

[Turkle 97]

Turkle, S. (1997). Constructions and Reconstructions of Self in Virtual Reality: Playing in the MUDs. In Sara Kiesler (Ed.), *Culture of the Internet*, Lawrence Erlbaum Associates, NJ, pp.143-155.

[Vesna 98]

Vesna, V. (1998) Marketplace: from Agents and Avatars to the Information Personae. *Digital Creativity*, Vol. 9(3), pp. 129-136.

[Vilhjálmsson 97]

Vilhjálmsson, H. (1997) Autonomous Communicative Behaviours in Avatars. MA Thesis, MIT, Accessed July 2001, http:/hannes.www.media.mit.edu/people/hannes/msthesis.

[Volino & Magnenat-Thalmann 99]

Volino, P., and Magnenat-Thalmann, N. (1999) Interactive Cloth Simulation: Problems and Solutions. Course Notes, *Virtual Humans: Towards Interactive Actors,* International Conference: From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks, National Museum of Photography, Film & Television, Bradford, UK, 14th-17th April 1997.

[Guye-Vuillème 98]

Guye-Vuillème, A., Capin, T., Pandzic, I., Thalmann, N., and Magnenat-Thalmann, D. (1998) Nonverbal Communication Interface for Collaborative Virtual Environments. *Proceeding of CVE 98*, Manchester, 17th-19th June 1998.

[Waters 96]

Waters, C. (1996) *Web Design Concept*. New Riders publishing, Indianapolis.

[Wilcox 98]

Wilcox, S. (1998) *Web Developer.com Guide to 3D Avatars*. Wiley Computers Publishing, New York.

[Wooldridge & Jennings 95]

Wooldridge, M., and Jennings, R. (1995) Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, Vol. 10(2), pp. 115-152.

Publications

The Role of Obligation within Virtual Encounters

Nic Earle & Colin Beardon
NVRCAD Research Centre, Exeter Faculty of Art & Design,
University of Plymouth, Earl Richards Road North,
Exeter, EX2 6AS, ENGLAND
+44 (0)1392 475126
nic@hal.nvrcad.plym.ac.uk
http://www.esad.nvrcad.plym.ac.uk

Abstract

This paper examines aspects of the social behaviour of people when participating in chat oriented Multi-User Dungeons (MUDs). The research focuses on aspects of the discourse that take place between people in these environments with the aim of applying the knowledge to visual multi-participatory environments. The research is part of a project to inform art and design practice, the eventual result of which will be a multi-user experimental environment with a high visual content.

The paper includes a brief introduction to MUDs and a description of present research before describing the research methods used in this project and analysing two dialogues that took place in a chat oriented MUD. The analysis suggests that the dialogues within such MUDs typically break down into three stages and that they can be further described through the use of a concept of 'obligation'. The processes whereby new participants join a group, a particular activity is decided upon, and individuals leave a group are each examined in this light. There are additional observations on larger groups.

An understanding of the social dynamics of interaction in textual environments provides a firm grounding to develop visual environments in which different modes of communication are complementary.

Keywords

Multi-User Dungeon/Dimension (MUD), networked virtual environments, conversation analysis, social obligation, virtual art

1: Introduction

The long-term aim of this project is to develop interactive three-dimensional multi-user environments that are more 'sociable'. We believe that to achieve this the different media involved (text, sound, image, etc.) must work together to positively encourage social interaction, in much the same way as they do in real social environments such as a reception or a party. As part of our design research we need to examine various media, how they are used in practice, and how this knowledge might inform the larger design activity.

MUDs are text based, interactive, multi-participatory virtual reality environments accessible through the Internet. In MUDs people can meet and interact with each other in almost real time, typically through the use of text. Each participant is identified by a self-chosen name and will be able to see a list of the names of the other people present in that room which is updated each time there is an arrival or departure. Participants can make and name a room of their own which will operate in the same way as a room provided by the multi-user server software. The only difference is that these rooms cease to exist when the room becomes empty.

The medium has certain specific features which we will refer to later. Firstly, in these environments the real participant usually chooses to be anonymous (i.e. only their MUD name will be known to other participants). Secondly, communicating through a textural interface removes many of the additional channels of communication that usually accompany discourse such as facial expressions, lip movements, body language, body posture, eye contact and voice tone (Çapin 1997). Thirdly, a common event in these environments is that a technical breakdown occurs and participants lose their connection to the multi-user server without warning (also known as getting 'booted').

Considerable research into MUDs has taken place including investigations into the social phenomenon (Curtis 1992) and the development of a specific culture (Rosenburg 1992), exploring questions of identity

(Rheingold 1993) and gender swapping (Bruckman 1993), the roles of spatial perception and navigation for interface design (Tromp 1995) and their use as environments for collaborative learning (Bruckman 1994). The linguistic structure of such conversations have also been covered in speech act (Grosz and Sidner 1986) and discourse theory (Maier and Hovy 1991) within linguistics. From the point of view of design, however, there is still a need to investigate the use of MUDs in order to analyse aspects of behaviour that are both necessary and are capable of being implemented in various other media.

2: Terminology

The term "participant" refers to any person present in the environment. In order to preserve the anonymity of participants we have given each participant an alias (the name of a fruit) within this paper. We use italicised text when referring to a participant in the text (e.g. *Apple* said...). To avoid cumbersome text, we vary our use of the pronouns "him" or "her". (Many people in MUDs present themselves as the opposite gender so it would not be possible to know from our data the gender group of each participant.)

We use the term "message" to refer to a single continuous piece of text that is sent as one unit to the MUD address by a participant and is subsequently received by all other participants. It is important to note that the structure of messages is not based upon linguistic units, such as sentences. A message may contain several sentences, or a single sentence may be split over two or more messages (Fig. 1).

Apple: in fact i love to eat
Orange: they give a lot of energy but don't make you hyper
Orange: and are very good for ya
Kiwi: so do i! i only ate 1 time today or yesturday and that was like 3 pierogies
Orange: you can get um at health stores
Kiwi: thats cool! i will have to try that!

Figure 1. A typical MUD conversation

When a message is read it is usually (though not always) seen by a recipient as referring to one or more previous messages, in which case there is said to be a (directional) link from the later message to the earlier one. We use the term "conversation" to describe a sequence of messages linked to each other in this way. The start of a new conversation is a message that does not refer to any previous message. A "group" consists of the participants in a particular conversation, so when a participant has contributed to a conversation we may refer to that participant as a "group member".

3: Research Method

The chat environment used for this research is one of the more popular ones operating from the US. The environment is mainly used as a social meeting place and has several rooms which a participant can choose to enter. Because participants are aware of anyone who logs into the environment it is not possible for a researcher to be present as an uninvolved observer. This requires a researcher to decide how to behave when observing the dialogue. The decision was taken, at the initial stage of the research, to try to act as naturally as possible: to enter the environment as someone who has recently discovered the MUD and has returned for another chat as this was the real-life position of the researcher. There was therefore no particular objective in mind other than to chat and no statements or questions were prepared in advance.

The primary interest of the research is to increase understanding of the patterns of behaviour and social interactions that are taking place in the MUD environment and, because the only medium of communication is through the exchange of typed messages, the key to this is an analysis of conversations that have taken place. After participating in and observing six sessions, two dialogues were selected for further analysis. The first is named *One-to-One* because it consists of a single conversation between two people with no other participants entering the room. The duration of this conversation is thirty five minutes. An analysis of the text of *One-to-One* was conducted at both word and message level. It was hoped that this would provide a base-line understanding of the mechanisms at work when two people in a MUD environment interact with each other. It should be noted that this example describes an unusual conversation and is used in this research as an illustration of fundamental mechanisms, rather than as a demonstration of their general importance within MUDs.

The second dialogue is named *First Steps*. The duration of this session is one hour and twenty minutes and it involves a total of nine participants, with a minimum of three at any given time and a maximum of five. Unlike the *One-to-One* dialogue, it does not represent one continuous conversation from beginning to end. Instead it represents several conversations that develop and, to some extent, intertwine. The method involves seeing how the observations that were made in *One-to-One* might apply when more than two people are involved. To achieve this a text analysis of the dialogue was conducted. To further aid the

identification and understanding of the patterns of conversation a linear diagram was constructed which is explained in Section 6.

4: Analysis of the *One-to-One* Dialogue

The *One-to-One* dialogue is, in some senses, very straightforward and provides an opportunity to develop a simple analysis which can be applied to more complex situations. Although it raises no real surprises, it is worth documenting carefully as it provides a base line from which complexity can be measured. The dialogue forms a narrative which divides naturally into three parts, which are labelled *Start*, *Middle* and *End*. In the *Start* section, the participants at first engage, try to get to know each other and negotiate a meaningful topic of conversation. In the *Middle* section the participants undertake a specific activity. In the *End* section the participants negotiate a way to disengage from the conversation.

4.1 Word Level Analysis

From an analysis at the word level of the *One-to-One* dialogue it is clear that the participants have pre-negotiated techniques to enhance their communication in a text-only medium. Enhancing the communication medium has the potentiality of increasing interactive experience by compensating for the forms of non-verbal communication which are missed. Text-only modes of communication can prove frustrating for the authors of messages as some of the added layers of meaning that might be expressed incidentally have to be carried explicitly.

For example, two asterisk symbols (**) are used to contain a word or sentence to show that it describes a performative action or sound, such as "*hug*". Laughter too has textual representations such as "hee hee" or "lol" (an abbreviation for "laughing out loud" — such abbreviations are becoming standardised, see URL http://www.cs.columbia.edu/~hgs/acronyms/m.html). Participants will also often use between two and five consecutive dots (i.e. ".....") either in the middle or at the end of a message. In the middle of a message they usually represent some hesitation or delay such as might occur in speech. At the end of a message they usually imply that the message is unfinished. We will discuss some implications of this in Section 4.2.

4.2 Message Level Analysis

Analysis at the message level is aimed at investigating the following questions: how do people get into conversation?, how do people set up parameters for the conversation? and how do people get out of a conversation? After several attempts at analysis we found that a meaningful way of describing the interactions that took place is in terms of a general sense of obligation towards other participants which can be refined to take on specific forms at different points in the dialogue. We became interested in what kinds of obligation exist, between which participants they occur, and how they are manifest at various stages of the conversation.

When the two participants first get into conversation they typically acknowledge each other's presence in the room by way of a greeting. Their mere presence, it might be assumed, means that they feel an obligation to find something to talk about. Within the *Start* section smaller sections could be identified that explore potential topics of conversation. These explorations often reference the few basic assumptions that it is reasonable to make about other participants. For example, it is reasonable to assume that the other person is in front of a computer monitor, so one of the first comments is "blinking at the screen...".

After exploring several avenues, the conversation returns to one of the initial topics, the idea that the two participants should create a story. This is explicitly suggested and can be seen as fulfilling an initial obligation to find a topic of conversation. Its acceptance by the other participant creates a new context, a kind of nested obligation within which the participants have agreed to co-operate in the construction of a narrative. From this point, the structure of the dialogue changes radically in line with the new obligation. The messages become much longer (i.e. contain more characters), there is a very regular pattern of exchange of messages and as a result the dialogue itself becomes fluent and easier to interpret. Why is this? and how are the rules, parameters and conventions for such game-like behaviour established?

Within the dialogue *Apple* suggests that they construct a story together and proposes two basic rules: (a) to make up one line at a time, and (b) to take it in turns. This is a clear and explicit offer, which *Pear* accepts but not explicitly. Rather, he responds with what is clearly an instance within this new obligation framework: that is to say, he produces what is the first line of a story, "once I was a little boy" and adds "[your turn]" at the end of the message. This message, through its content, is a clear acceptance of the offer to go into story-telling mode, and the indication that it is unfinished underlines the obligation for the other participant to respond. The next time *Pear* sends a message, "[your turn]" is replaced by the

consecutive dots symbol that is identified above and which is used consistently throughout this middle section of the conversation.

The obligation the participants have to each other is complicated further by the nature of the activity (i.e. constructing a narrative). Narrative is normally thought of as having a beginning, a middle and an end and because of this the participants have an obligation to complete each message in a way that both continues the narrative and drives it forward to reach a point where the story can be completed.

The rules and conventions for playing this game are firmly established. The developed sense of obligation creates a sub-environment in which the participants can compromise the usual convention of using short sentences, allowing them to become more descriptive in their use of language. The messages are semantically coherent and arrive in a fixed structure (i.e. in turn). Greater fluency is achieved in the conversation as the pair develop and work within a committed environment.

The game contains its own obligation upon the participants to complete the narrative. Also, it would not seem appropriate to end a conversation abruptly, say by simply informing the other that one has to go. In the example these two distinct obligations begin to run together.

So how do the participants indicate that they wish to end the game while fulfilling their obligation to complete the story? Several lines before the conclusion of the story one of the participants fails to put consecutive dots at the end of his message, and we suggest that this might be a (possibly subconscious) signal from *Pear* to *Apple* that he wishes to end the conversation soon. Several messages later *Pear* offers an ending to the story. *Pear's* method for ending the story is ingenious: he brings a character into the story who is also named "Pear" and who intends to leave soon ("... to watch Columbo on television"). *Pear's* construction means that, in one move, he addresses the nested obligation to complete the narrative, and the encompassing obligation to end the conversation.

Having completed the construction of the narrative, the participants still do not immediately leave but conclude their conversation by saying "thank you" and "good-bye" to each other. This is not just a signal that the conversation is over, but also an acknowledgement that the mutual obligations have been met and the conversation is satisfactorily concluded.

4.3 Summary of main observations on *One-to-One*

- The participants begin with a greeting to acknowledge each other and initiate the conversation.
- There is a process of negotiation about potential topics, based upon common knowledge and shared aspects of their circumstances.
- In middle section the participants settle on an activity for which a set of conversational rules are negotiated. Here the level of interaction is at its highest.
- The participants build various obligations to each other and thereby form a working relationship with commitment.
- The obligations can be clearly illustrated when it comes to ending the conversation.

5: Analysis of the *First Steps* Dialogue

The first dialogue represents an uninterrupted conversation between two people, it is now necessary to see how the analysis works in an environment in which more than two participants are involved. In this context there is a need to describe where each conversation begins and ends, who is talking to whom, how many conversations are taking place at any one time, and how the conversations are intertwined. A method is developed to represent the structure of a conversation as a linear diagram which also enables a more rigorous definition of some central concepts.

5.1 The Diagram

The diagram is produced by creating a node for each message and arranging these in a vertical linear path corresponding to the order they appear on the screen. Each participant is given a letter, and a letter is attached to each node representing the author of the message (Fig. 2).

Starting at the beginning of the dialogue each message is considered in turn. Where a message is understood as a direct response to some previous message (or messages) a link is created between the nodes. (Colouring the link according to the author improves the visual readability of the diagram, as does numbering the messages consecutively.) It is not always possible to link a message into a conversation. Some messages start a new topic, some are an open question to all participants and some are difficult to understand at all in the context of the ongoing dialogue. The judgements made here are subjective by the

researchers involved but we think that they would be confirmed by other people with experience of similar environments.

There are several occasions where two or more messages in succession have the same author. If the message is a semantic and syntactic continuation of the previous message then it is linked to it (i.e. the semantic boundary of what is being said extends over two or more messages). If, however, the second message is a response to some other message then it will be linked to that in the normal way. There are many examples of both in the dialogues studied.

Other information in the diagram includes a number to indicate how many people are in the room at any particular time, a horizontal line to indicate the point at which the whole group is 'booted' from the multi-user server and some written observations and comments.

By using this diagramming technique it is possible to readily identify how many people are involved in a particular conversation, how long a conversation has lasted, how many conversations are going on simultaneously, which participants are not in a group, and many other concepts which otherwise might have remained vague.

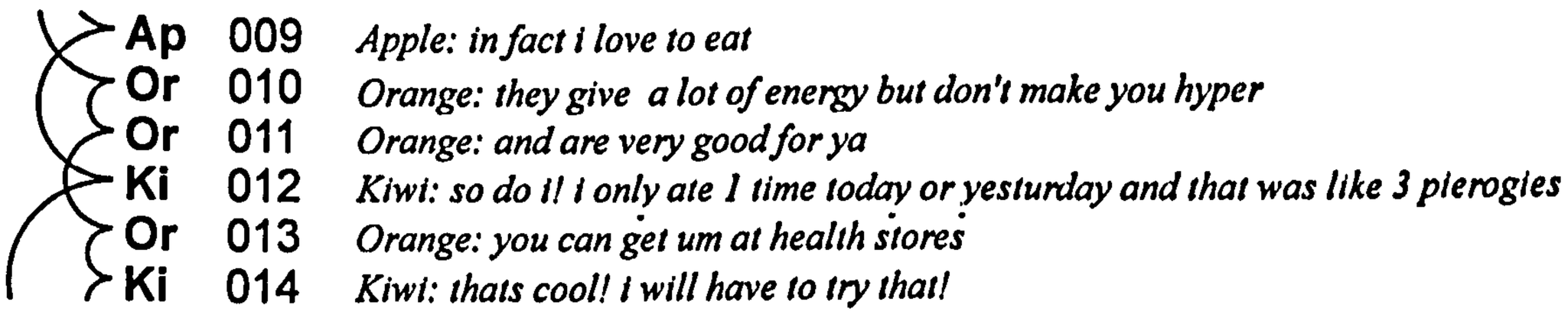


Figure 2 A comparison of dialogue text and diagram

5.2 A Third Participant Joins a Group of Two

When there are only two people in the MUD room, the links are easy to follow. When there are three or more participants present, the patterns of conversation are very different. From our first analysis we have some understanding of how two people interact and this raises the question of what happens when a third person enters the room and wishes to join a group of two?

Two main ways in which a participant may join a group are identified. She may either wait until a present member directs a message at her, such as a greeting, or she may intervene in the present group's discussion with a message such as "what's up everybody?" and await a response. There is no identifiable initiation ceremony to join these social groups (as in Morris, 1967); it is our observation that the newcomer usually has to be brought into the group by one or more of the present members in order to be accepted. If the newcomer does not receive an invitation into the group then she is likely to leave the room completely.

Once accepted into a group, the newcomer is expected to participate in the conversation. Two occasions in our data when a third person joins a group of two will be considered. Figure 3 shows that the flow of conversation is very different in the two examples. Why is this?

In the first example, *Apple* tries to join *Kiwi* and *Orange*. The latter two are already engaged in conversation when *Apple* arrives and he is brought into the group by *Kiwi* who continues to include him in the conversation. In the next three consecutive messages *Kiwi* continues her previous conversation, responding to *Orange* and then responding to *Apple* in consecutive messages. This pattern continues for the next eighteen messages. The diagram shows us that when *Apple* enters the room he responds only to *Kiwi*'s messages, not *Orange's*. If we follow the paths of conversation at this point in the diagram we see that *Kiwi* is really having two parallel conversations, one with *Orange* and one with *Apple*.

In the second example, *Mango* tries to join *Apple* and *Coconut*. *Mango* takes the initiative in this case and intervenes, inviting the other two to include her in the group. She is given an initial greeting by the other two group members but they immediately return to their conversation and she finds no major role to play. She submits messages at a much lower frequency then the other two participants and, despite her persistence, does not mange to attract conversation her way. This is a very different model. So why are the models different?

One possible factor is that, about three quarters of the way through the dialogue, all participants are 'booted' from the server. Before that event *Melon* is trying to get involved in a conversation. He receives little response from *Apple* but *Coconut* addresses *Melon* on several occasions (the situation is, thus, similar to our first example). This would seem to eliminate the possibility that *Apple* and *Coconut* were unwilling to include a newcomer in their group.

Apple and *Coconut* are engaged in a one-to-one conversation when they are booted from the MUD server. What happens next seems very significant. *Apple* and *Coconut* not only log back in immediately but they also seek each other in order to continue the conversation. This is initially frustrated because the room that they were in had been created by *Kiwi* and when the group were booted the room disappeared.

Coconut therefore creates a room and names it 'Calling for Apple' with the hope that *Apple* will log back in, see and enter the room, and find *Coconut* in order to resume the conversation. *Coconut's* plan is successful. *Coconut* and *Apple* continue their interaction in the new room. This experience serves to confirm and enhance the bond between them. Their obligation to each other becomes strong so that by the time *Mango* enters the room she finds it difficult to become accepted within their conversation.

The sense of obligation that is used to describe the participants' behaviour in the first dialogue plays an even stronger role in describing behaviour in multi-participant environments. This obligation is a bond that exists at various levels of commitment within a larger group. Furthermore, if the level and type of obligation between two participants is sufficiently high it may provide a barrier to a newcomer when trying to join a conversation.

The role of obligation as an analytical tool is further strengthened when we consider descriptions of the situation where a participant leaves a group of more than two (see Section 5.4).

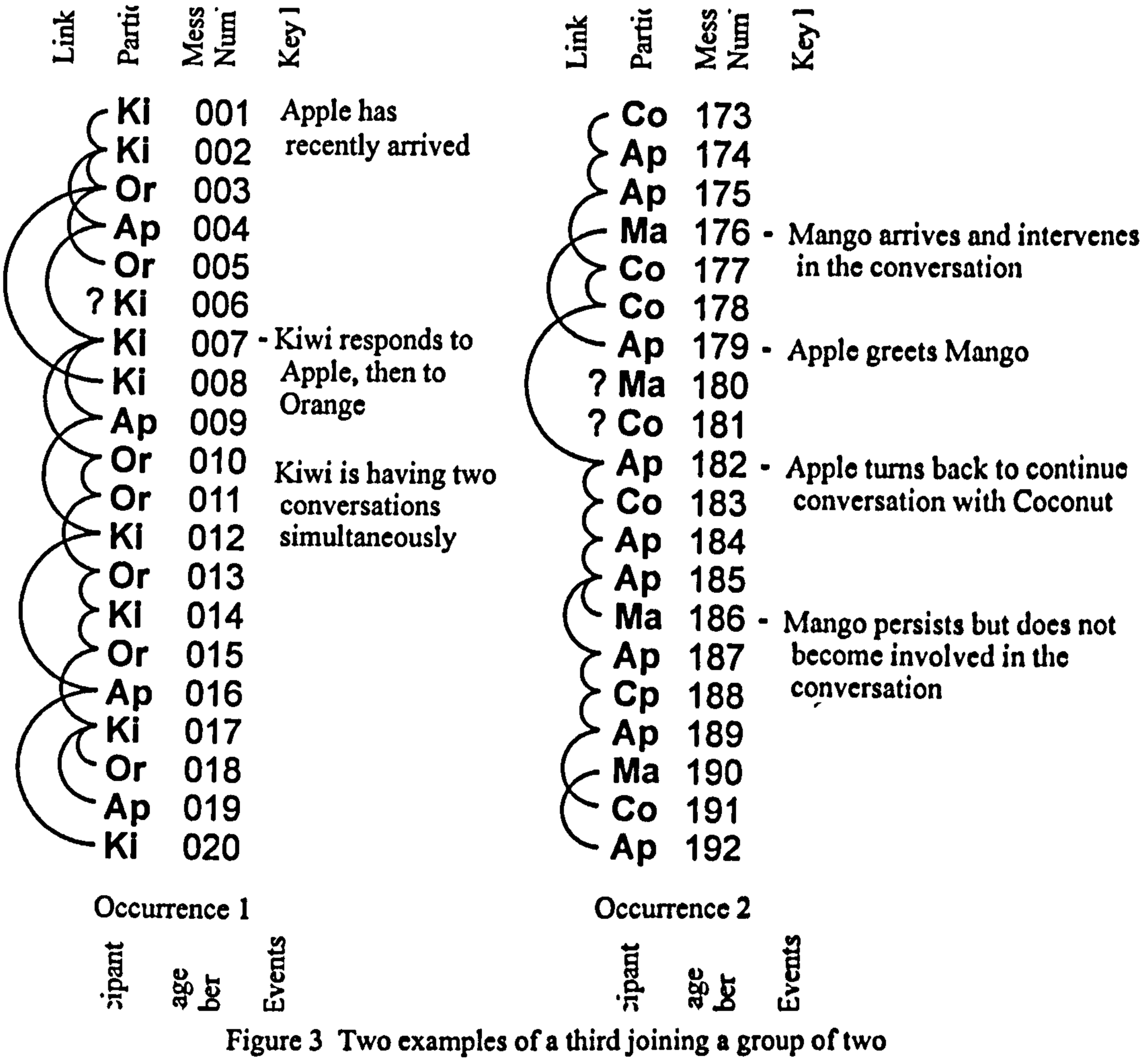


Figure 3 Two examples of a third joining a group of two

5.3: Newcomers to a Group of Three or More

The major aspects of the situation when a newcomer joins a group of more than two have been covered in previous examples, but a few useful observations can be made. In groups of four or more there is sometimes a different mechanism to initiate a conversation. As well as exploring a number of potential topics, participants will occasionally ask a question that is clearly directed to all group members, such as "where are you all from?" This is similar to the *Start* dialogue in *One-to-One*, but addresses a wider audience. If all the participants respond, this can serve to bring together a group as a whole and initiate a conversation which, in turn, creates a new group bond. From there on the number of conversations that are taking place will vary and the number of people involved in those conversations will also vary. This

model will tend to hold until the next open question or until the number of people in the room reduces to three.

Specialised language is also used to enhance group bonding and can create a further barrier to a newcomer. When *Cherry* enters the room in the *First Steps* dialogue, two of the group members are using a lot of acronyms in their messages. She indicates that she feels exterior to the group with the message, "I need to learn the lingo". The use of expressions such as "lol" causes *Cherry* to feel (temporarily) excluded from the rest of the group.

5.4: Leaving the Group

How does the process of negotiating an exit work in a group of more than two? Figure 4 illustrates the typical situation. The general pattern is that a participant gets to a position where she is not directly involved in the conversation and then signals to the group an intention to leave. This is not sufficient, however, and is rarely the last message that she submits. The person leaving needs the same sense of completion that was identified in the *One-to-One* dialogue. Half way through the *First-Steps* dialogue, *Kiwi* signals that he intends to leave. All of the participants, except *Apple*, respond by saying "good-bye". *Kiwi* does not leave at this point and specifically waits for *Apple* to respond. This suggests that, unlike the arrival of a new member to the group where she normally has to be invited in by one person, the departure from a group normally requires the recognition of all group members. While the general background conversation continues, it is interleaved with every group member saying good bye to *Kiwi*.

To underline the significance of this leaving protocol this we can look at the departure of *Cherry*. When *Cherry* signals that he is leaving, there is a response from only one of the three members present. *Cherry* clearly does not feel that the other two have fulfilled their obligation to the group and specifically requests a response ("doesn't anyone else care?"). The others then respond and the obligation is satisfied. Only at this point does *Cherry* leave.

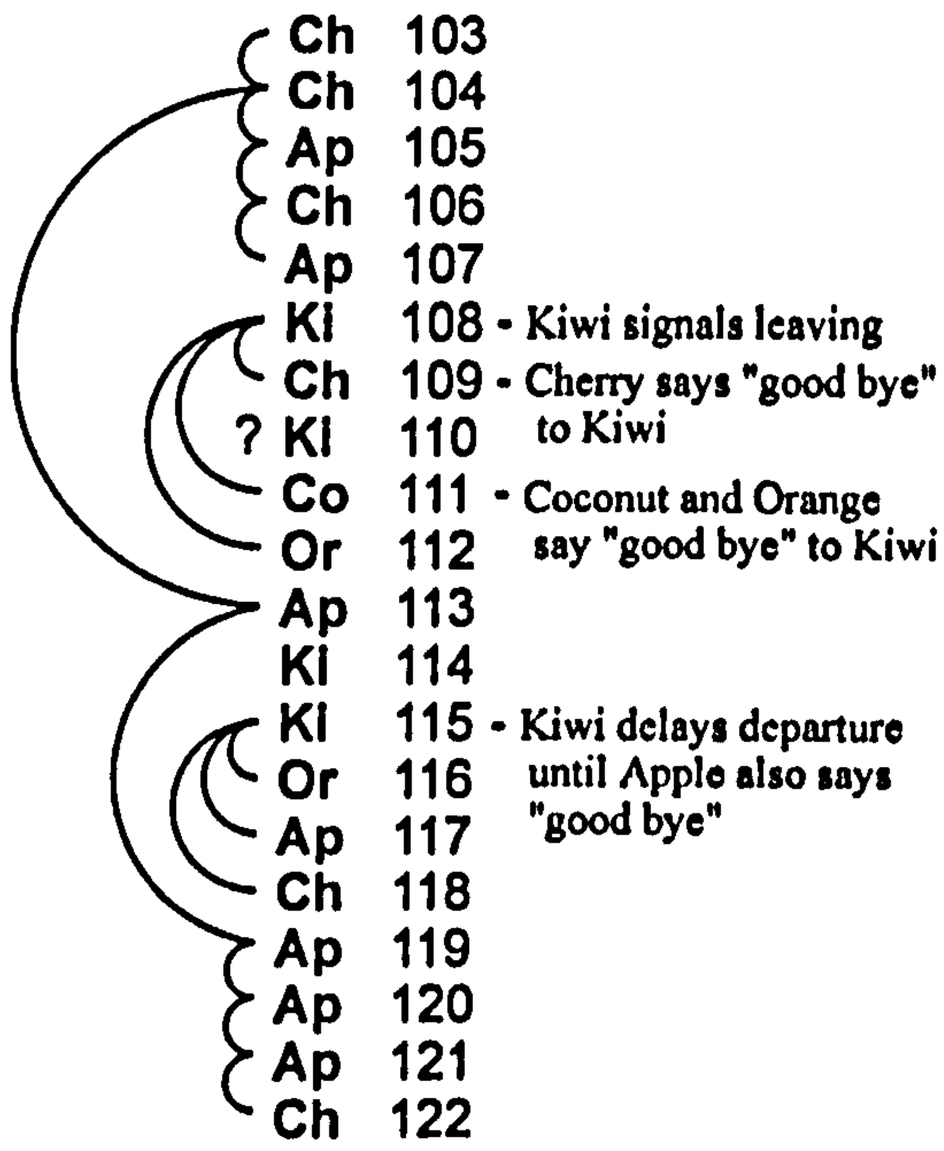


Figure 4 Leaving a group

6: Conclusion and Future Work

Having investigated several dialogues in such domains, we believe that there are distinct types of communication, which can be changed by the participants through the use of text only communication. To understand these we believe that it is necessary to develop the concept of obligation and its relationship to dialogue analysis.

One particular mechanism, which may not have been adequately covered by this research, concerns the transition from one-to-one conversations into genuine multi-participant conversations. We saw how some three participant conversations were, in effect, two one-to-one conversations in parallel but not all conversations can be described in such terms. One mechanism that is identified is the open question ("where are you all from?") but it is suspected that other mechanisms are possible. More data is required before this line of exploration can be pursued.

There is also a need to reflect on research methodology. The researcher did not declare his intention to observe and record what participants say to each other. If one of the essential elements of these environments is a sense of trust and obligation, then keeping this hidden means that the researcher is immediately acting in bad faith unlike (we assume) the other participants. On the other hand, if the researcher declares his true intention then this is likely to greatly affect the dialogue that follows.

The work reported here is preliminary research that is undertaken as part of a particular artistic practice, aimed at the creation of multi-user environments that work visually rather than textually. Typically, each participant will be represented by an avatar which can have a limited number of attributes from which their external appearances may be partly derived. The aim of the research has been to inform us about the possible nature of those attributes and appearances and the way they may interact with the dialogue in a purely formal sense (i.e. unrelated to the content of what is said).

We could imagine, for example, a multi-user environment in which conversational groups are identifiable, where a newcomer can clearly indicate their interest in a particular conversation, where even the degree of bonding can be indicated through the orientation of speakers to each other, and so on. However, the methods of artistic practice are not always as explicit as other disciplines and there is other research that we may wish to complete before embarking on an implementation. We are, however, convinced that an understanding of the social dynamics of interaction in virtual worlds should inform design practices and we believe that this research is a contribution to that process.

Bibliography

Bruckman, A. (1993). Gender swapping on the Internet, *Proceedings of the INET'93,* http://www.cc.gatech.edu/fac/asb/papers/index.html

Bruckman, A. (1994). Programming for fun: MUDs as a context for collaborative learning. *National Educational Computing Conference* in Boston, MA, June 1994, http://www.cc.gatech.edu/fac/asb/papers/index.html

Çapin, T.K. Pandzic, I.S. Thalmann, N.M. Thalmann, D. (14th - 17th April 1997). Realistic Avatars and Autonomous Virtual Humans in VLNET Networked Virtual Environments. *From Desktop to Webtop: Virtual Environments on the Internet, WWW and Networks,* International Conference, National Museum of Photography, Film & Television, Bradford, UK.

Curtis, P. (1992). Mudding: Social phenomena in text-based virtual realities. *Proceedings of DIAC' 92,* Berkeley ftp://ftp.lambda.moo.mud.org/pub/MOO/papers/DIAC92.txt

Grosz, B. J. & Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics* 12(3): 175-204.

Maier, E. & Hovy, E. H. (1991) *Organizing Discourse Structure Relations Using Metafunctions.* Paper presented to the European Natural Language Generation Workshop: GMD-IPSI, Darmstadt.

Morris, D. (1967). *The Naked Ape.* Vintage, UK

Rheingold, H. (1993). Multi-user dungeons and alternate identities. *The Virtual Community,* Ch 5. http://www.well.com/user/hlr/vcbook5.html

Rosenburg, M. (1992). *Virtual Reality: Reflections on Life, Dreams, and Technology; An Ethnography of a Computer Society.* ftp://ftp.lambda.moo.mud.org/pub/MOO/papers/ethnography.txt

Tromp, J. Dieberger, A. (1995). MUDs as Text-Based Spatial User Interfaces and Research Tools. *Journal of Intelligent Systems,* 5(2-4), pp. 179 - 202.

Designing a Non-Verbal Language for Expressive Avatars

Ben Salem*
University of Plymouth
Exeter School of Art & Design
Earl Richard Road
Exeter EX2 6AS England
+44 1392 475126
electronote@mail.com

Nic Earle
University of Plymouth
Exeter School of Art & Design
Earl Richard Road
Exeter EX2 6AS England
+44 1392 475126
nic_earle@postmaster.co.uk

ABSTRACT
Collaborative Virtual Environments (CVEs) were designed as an expansion of the text-based chat room, rather than a novel application, exploiting the possibilities of online three dimensional graphical space. This initial design direction is observable at the interface level. We put forward the case that to achieve an efficient CVE system, one will have to design and implement a multi modal User Interface based on expressive Avatars as a representation of the different participants, also as an embodiment of software agents. We emphasise the expressiveness of the avatar as a crucial improvement to the efficiency of their communication capabilities, and we describe a vocabulary of expressions to be implemented. We put forward the case that to be more efficient, particularly during a dialogue, an avatar is required to play a role in the communication using non-verbal channels such as body postures, facial expressions and hand gestures. We also suggest conversation circles to facilitate the gathering of participants in a discussion. These circles will address navigation difficulties in CVEs and encourage social exchanges.

Keywords
Avatars, Facial Expressions, Hand Gestures, CVE, Social Communication

INTRODUCTION
A Collaborative Virtual Environment (CVE) is a multi-user virtual space where users are represented by a 3D image termed 'avatar'. The participants of a CVE should be directly visible to themselves and to other participants [1]. This creates a more immersive experience and has greater potential to provide important information to a user about their avatars surroundings as well as reinforcing the perception of the environment. Several papers have been published on the visual aspects of avatars however few have addressed theirs communicative functions [1][2][3]. The issues we wish to address are not only related to the characteristics of the avatars in terms of shape and visual appearance, but are about their functions and behaviours in relation to their owner.

Within this paper the term 'avatar' is also used to describe a graphical representation of a software agent which some CVEs feature. Actions of this type of avatar are controlled by software. The advantage of giving an agent a embodiment is to firstly humanise the agent and secondly to allow a more natural method of interacting with it.

The use of common conventions, that one is familiar with, facilitates the interaction within the CVE as well as improves the communication between users, and between users and agents. Humanoid avatars for the participants are therefore the most efficient design option, they allow the use of everyday conventions such as gestures, postures and body language as part of the communication channels between users. For the embodiment of agents there is more freedom about the realism, and type of avatars to be used while keeping a practical level of expressiveness and interactivity. The use of abstract or symbolic representations for a participant is less justifiable if one wants to achieve a user-friendly environment delivering a rich experience with a degree of multi-modal communication. Having a human-like appearance makes it easy to gather the aim of a conversation, as well as to recognise the participant to whom one is speaking. There is however a need to draw on these principles and develop a set of design guidelines that are useful to the development of avatars and CVEs.

A further problem in developing a non-verbal language of avatars is that avatars that are in conversation are usually scattered around an environment. Even avatars that are within the same conversation are often out of view of one another, forcing the participants to rely on the text window. This is a major problem as any gestures and expression that are performed are likely to be missed. There is a need for the design of a better conversation space which we suggest in this paper.

THE ROLE OF AVATARS
The role of avatars in CVEs is essentially to represent a user's presence, orientation and location. We propose a set of design guidelines to deliver avatars that have further capabilities, namely to facilitate the communication of ideas and the expression of emotions as well as identifying its owner. We address issues concerning group communication, personal identity and expressions. There is a need for avatars to convey relevant information about their users real identity and his intentions during discourse

and conversation. Our research problem therefore divides into two main areas. The first aims to raise issues that need to be addressed about personal identity of human participants. The second is concerned with aspects of the individual and group communication for both human to human collaboration as well as human to agent collaboration.

In the latest group there has been a steady increase in the use of software agents for the assistance and guidance of new participants in a CVE unfortunately this has been done without a common set of behavioural rules and vocabulary of expressions. Leading, to a difficulty in understanding the agents, or to the false impression that one is facing another user.

Figure1: Anthropomorphic avatars (from Active Worlds)

Some of the avatars used in CVEs are humanoid but inadequately designed, as shown in Figure 1, the level of detail is either too poor to permit rich facial expressions, or the design is such that few expressions and gestures are possible.

Facilitating the User's Discourse and Interaction

Once logged-in, communication between the CVE participants is possible by text chat and a library of body

Figure 2 : High level of detail Avatar (from Poser).

and hand gestures. Text messages appear on the screen either above the avatar of the message author or within a separate dialogue window. Users are able to interact with the agent software through the exchange of text messages.

Avatars can be realistic, abstract or naturalistic in form. We ask what is the best style for CVEs. Realistic avatars strive to provide an accurate representation of the user. For example a participant with a realistic avatar may use their real name as a username and real time video image of their face wrapped onto their avatar face [2] [4].

Naturalistic avatars are usually humanoid in form, of a degraded level of details, but emulate natural paradigms just enough to achieve recognition of familiar features [5]. We advocate the use of this style in CVEs. As a humanoid form it will open up the possibility of using conventional conversational habits, such as looking at the speaker, and using all the non-verbal communication techniques one has learnt to employ during a conversation.

Much research has gone into realistic modelling of virtual humans. Notable success are the JACK system [6] and the Xian Terracotta Soldiers [7], however this is a difficult path paved with problems when attempting to apply results to 3D avatars. This is due to technical restrictions such as bandwidth and computational power.

Despite several years as a research topic, avatars in existing collaborative systems play little role in the communication that takes place. Communication is achieved almost entirely on the exchange of text or in some cases voice messages, while in the real world communication is far richer then this. It involves other non verbal channels of communication (NVC) such as face expression, eyes' gaze, eye opening, duration of glances, facial expressions. As well as body postures and gestures, which makes the difference between formal and casual conversations. Also to be included are hand postures and gestures as part of the discussion, for example, to emphasise words, and quantify them, body contacts, hand shakings and slap on the back.... Finally, Clothes, makeup and other aspects of appearance, formal and casual dresses and, uniforms play a significant role as well [8][2][9]. NVC play an important part in our social interaction. When several channels of communication work together we use the term 'multi-modal'. With a human-like avatar, it is possible to exploit some of the non-verbal communication (NVCs) capabilities, facial expressions, hand gestures and, body postures. At the same time these NVCs will be perceived and easily understood by the other participants in the CVE.

A relevant system known as Comic Chat developed by Microsoft integrated the visual language of comic strips with text chat [16]. The result was a brilliant system where symbols were extracted from text messages to automatically generate body gestures, face expression and background scenery. While some useful lessons can be learned from the project its carton like aesthetic encourages anonymity and playful behaviour. It is therefore more suited to chat and role playing games than CVE applications.

Existing CVEs

In existing systems however there are usually two separate modes of operation. The first is controlling an avatar to get into a spatial position where one is able to enter discussion.

The other is typing text or speaking, which in current systems implies a halt to the control of the avatar. In order to develop a multi-modal communication, a visual language for the avatar needs to be developed that supports the discourse that takes place, with a merging between the two. This is particularly true for CVEs that that have a separate window for the textual dialogue. Such a separation removes the user's attention from the avatars and the environment, in which case any visual clues performed by an avatar would be missed.

A final problem concerning social communication that is addressed within this research is that of activation. Some existing virtual worlds such as the commercial system Active worlds offer a limited number of gestures such as waving and some body expressions such as Laughing. However these gesture need to be activated by selecting buttons from tools bars within the virtual world software. We believe this to be totally inadequate solution. In the real world we use gesture in an effortless manner. Often we perform them subconsciously and even involuntary. This makes explicit activation of them from menus completely inadequate.

AVATARS KEY ROLES

We suggest that avatars are designed to address keys issues in the interaction process between the user and the CVE. We classify the key issues as the identity of the owner, the accessibility of the owner, the status of the owner, and the avatar functions and characteristics.

Identity of the User

Within the limits specified by the owner, it should be possible to gather as much information as one would wish about an avatar's owner such as identity, gender, location, interests.

Accessibility of the Owner

Has the owner got a web cam, an audio link, or is text chat the only possible link?

Status of the Owner

Is the owner of the avatar a guest user, a privileged user, a manager who is monitoring the CVE. Is the user on pause mode, how long has he been silent, on a poor internet connection. Does an avatar represent an agent or a human (is this important?).

Avatars Functions and Characteristics

What are the actions within the CVE that can be performed by the avatar, dialogues, displacements, manipulations, expressions, gestures.

Some multi-user environments such as the commercial Blaxxun world [10] allow participants to use more symbolic avatars. These avatars are abstract and usually encourage the real participant to remain completely anonymous. Participants with a symbolic avatars have the tendency not to use their real names when assigning themselves a username. Symbolic avatars encourage anonymity which tend to eliminates shyness and facilitate purely casual chats. This is not appropriate for most applications of CVEs. Furthermore the development of a gestural language for this type of avatar can not be based on the assumption that the gestures will be understood. Due to the abstract and symbolic nature of the avatar any suggested gesture or expression may be far too remote from the familiar ones to be easily perceived let alone understood.

GESTURES & EXPRESSIONS FOR AVATARS

There are several ways of performing gestures and expressions for avatars, which include the following: Real time manipulation of avatar limbs and face using the participant real body connected to sensors. Real time streamed video of a participant onto their avatar. Calling up predefined animations from a local library of gestures and expressions.

The first two methods are computational expensive and require specialist equipment. They also have a major problem in that the participant body resides in a very different space to their virtual body. As a example if a participant wishes to turn around to see who is standing behind them, they will have to turn away from the screen. One of the drawbacks of video streaming is the impossibility of achieving eye contact. To look someone in the eyes, requires looking at the camera instead of the screen and consequently not paying attention to the environment anymore. These configurations also do not transfer well to desktop systems. We therefore favour the predefined library of gestures and expressions.

Within this approach there are several issues that need to be addressed which are as follows:

1. What is the best form of NVC i.e. gesture or facial and body expression.
2. How do we arrive at a set of useful gestures or expression for both human participants as well as agents?
3. How can we ensure that the gestures and expressions are not missed by others when performed? Or how do we ensure that the communication becomes truly multi-modal.
4. How can they be triggered in a more natural way than selection from a tool bar?
5. How can we ensure that they are synchronised with the correct word or message?

What is the best form of NVC

The first issue to address is the choice between using facial expressions or body and hand gestures. In the real world they play slightly different roles. The face shows emotion and punctuates words and sounds while the body shows mood and status. In the virtual world however these two uses do not easily translate. While users engage in one to one conversation with another person, their head and shoulders may fill the screen. This results in facial

expression being noticeable while body and hand gesture are out of view. This is reversed when two people are communicating with a relatively large distance between them or when participants are conversing within a group of people who's avatars are gathered together. In this situation facial expression can be too subtle and are easily missed. Many styles of stage performance in history such as Commedia Dell Arte [11] give the actors masks to hide expression and encourage the use of expressive behaviour via the body. This principle is useful in this type of interaction.

We therefore propose that there should be two sets of animations developed for each meaning that is to be visually conveyed. One set is implemented through gesture, the other is through facial expression. These sets load up when required, based on proximity. When two avatars are within a certain range, the facial expression set loads. When they move out of range the gestural set loads.

This raises the question what will others in the environment see when they look over at a group from a distance? This should depend on their actual distance. It should be gestures unless they are close enough to view the face.

Body Postures

Body postures can be used to express the general state of the avatar. For example during a phone call the participant may want his/her avatar to express a temporary idle state.

Facial Expressions

Facial expressions are an efficient carrier of emotions. They are generally used to punctuate discourse at an informal personal level. When used in a context of a CVE application, facial expressions are less of an important

Figure 3 : Blinking face (from Poser)

feature for a participant avatar, however they keep all their usefulness for agent avatars independently of the application. A powerful channel of NVC is gaze. This is different to facial expression although it involves facial features. We have found from experience that a crude simulation of gaze can be achieve through the heads direction of face.

The body direction of face can be used to show group membership, and the head can be used to simulate gaze within the group.

Hand Gestures

Hand gestures are the most comprehensive NVC skill. We can classify gestures as being either :

Symbolic:

These are context independent expressions such as signs (e.g. sign language), the message is contained fully into the gesture performed.

Metaphorical:

Used in parallel to a spoken or typed message for quantification, expression of virtual and interface objects (e.g. a handle and geographical navigation), the message, in this case, is gathered from the gesture itself and the context within which it is performed. An example is using the hand to specify the intensity of the light, by moving it up and down.

Pointing:

To select objects or to reduce the scope of a spoken statement (e.g. this, that), this is a special gesture, since it can have so many meanings, most of them context dependants. Pointing could be performed to indicate a direction of displacement, an object or an option selection.

GENERATING THE EXPRESSIONS, POSTURES AND GESTURES

The generation of the facial expressions, body postures and hand gestures, is driven by key parameters, the current context within the CVE, the interaction task being performed, the current discourse of the participant if applicable.

We are primarily interested in two forms of NVC. These are hand gestures and facial expression. This is because they possess the widest range of capabilities for communication. We emphasise hand gestures since our principal way of interaction with the world where we live is through our hands; we perform almost all of everyday tasks with them, with a great dexterity and naturalness [12].

Two sorts of expressions are required: to illustrate the avatar status and to punctuate an avatar discourse. The first group is essentially made up of body postures (i.e. idle, thinking), while the second group is made of facial expressions and hand gestures. They are generally performed in parallel to a spoken or text-based dialogue.

Within textual message users often use text symbols to convey emotion and describe a bodily action. This is an attempt to replace the body in interaction. It is therefore logical that these textual symbols are given a visual clue as to their meaning. These symbols include acronyms, emotion icons and performative words.

- *Acronyms* – are abbreviation that make up a common expression e.g. LOL = laughing out loudly
- *Emotion Icons* – keyboard symbols used to create faces e.g. :-) = smile

- *Performative words* – these are words embedded within text message to describe actions. These words are usually identified by being contained within two symbols such as the asterisk e.g. *wave*

Text messages contain as well elements from traditional grammar that are used to punctuate messages. For example the question mark indicates a request for feedback and the exclamation mark emphasises a message, capitals letters are used for shouting. All these need to have a visual interpretation. There are also a number of words that are so commonly accompanied with a gesture in real world conversation that we believe they should be included within the set such as the words 'yes' and 'no'.

Examples from this section are shown in table 1:

Text	**Meaning**	**Gesture**
Yes	Agree / Yes	Nod head
No	Disagree / No	Shake head
?	Questioning	Head back, one eyebrow raised, hand out stretched
!	Emphasise message	Head back, eyebrows raised, torso upright
:-) LOL	Happy / Smiling Laugh Out Loudly	Smile Laughing
:-(	Sad / Upset	Head and shoulders dropped.
AFK	Away From Keyboard	Busy
OMG	Oh My God	Surprise
IMHO	In My Humble Opinion	Neutral pose
:-* *Kiss*	Kiss	Blow a kiss
wave CUL8R	Hello/goodbye See You Later	wave

Table 1: Examples of keywords with their visual clue

TRIGGERING GESTURES & EXPRESSIONS

We propose that a more natural method to activate gestures is through the text messages that are written and sent. A system needs to be developed where the text messages are scanned for keywords. When detected these keywords trigger an appropriate gesture. A similar approach could be adopted for spoken discourse. A speech recognition module, would extract key words from the discourse. This kind of system has already been implemented in the 3D Online Traveller world where verbal messages are scanned for phonemes that activate lip synchronisation [13].

Acronyms used within CVE discourse can be directly mapped onto their corresponding gesture. This is also true of punctuation marks used in traditional grammar as well as emotion icons.

Another technique that can be used is when a user sends a message intended for a particular group member by including their nickname within a message. The software should detect this and turn the author's avatar head towards the intended recipient's avatar.

Figure 4: Waiving and Thinking/busy poses (from Poser)

A user should also be able to inform the system, whether a trigger word or symbol is displayed within the message on the other group members screens. This is so that a gesture can be performed without a text message. For example a user may wish to indicate a desire to speak without explicitly interrupting the person who has the floor. This can be achieved by typing a symbol such as a forward slash (/) as a message. The gesture will be triggered but there will be no text sent.

Another useful feature is to allow the user to customise what gesture is assigned to what keyword. Changes can be saved as a separate set which can be loaded up at any time. This is useful for developing different pallets or gestures that are called up for different contexts i.e. social or formal. Some gestures such as signalling an intent to leave a group conversation need to have there own key assigned. For example the function keys can be used to activate appropriate gestures.

A MULTI-MODAL CONVERSATION SPACE

Investigations into the structure of virtual world social encounters reveals that the process of interaction typically breaks down into three sequential stages [9][14]. In the start part people seek an individual or a group for meaningful conversation. In the middle users interact and in the end they negotiate a way breaking out of the interaction.

In current CVEs avatars are scattered all around the environment. This distribution of the avatars is a direct consequence of difficulty in navigating in the CVEs. Another reason for the lack of clustering, one would expect when a conversation occurs, is due to the fact that although each participant's text message appear above their avatar in a bubble text, the text is often difficult to make out. This is even more the case in a crowded environment where text messages tend to overlap each other. Participants have the tendency of relying on the text window, and the experience shift from participating into a CVE to taking part in a text chat session. This itself has further reduced the need for participants to move their avatar into a gathering when engaged in a group conversation.

In order to achieve multi-modal conversation, not only should other channels of communication be developed such

as gestures, but users avatars must be in a space where they can see the other group members. We believe that the need for two windows i.e. text and visual, can be eliminated. Text message can appear in the same space as visual ones. This is a necessary step toward achieving a multi-modal interaction.

We have developed a concept we call the 'conversation circle'. This is a configuration for achieving multi-modal communication with chat systems. Users have been reported to arrange their avatar into a circle formation for discussion in existing virtual worlds [15]. Our proposal builds on this formation and presents a more sophisticated application. Figure 5 shows a screen shot of six avatars within a conversation circle. A circle begins when one user makes an offer of conversation by clicking on the recipient avatar. This is detected by the computer and the users avatar is animated over to the recipient on all user monitors and an appropriate gesture is triggered. If the recipient accepts the offer of conversation by clicking on the users avatar, both avatars are placed and locked along the circumference of an invisible ring. A user must signal that the conversation is over before they can leave the ring and move on. This can be achieved by detecting that a mutual obligation has being meet to leave the conversation for example, the wave gesture or a textual symbol embedded with the final text message sent to the group such as a forward slash (/).

Figure 5: Screen Shot of a conversation Circle

When a group is formed an invisible light fades in from the circle centre and just above head height. This light illuminates the avatars faces. It also creates shadows around the parameter which helps to visually define the groups space. Users external to the group can not move in this area. When last two members of a group end their interaction the light source fades out.

Newcomers can join the group by clicking on it with the mouse. The circle expands and the newcomer is placed along its circumference. Avatars are evenly distributed along circumference of the circle so in terms of spatial importance all users are equal. When a user departs, the circle contracts to bring together the remaining avatars.

The users viewpoint over the group needs to be third person view so that all participants are visible. This includes a users own avatar which is important as s/he must be aware of what gestures their own avatar is performing.

Text messages appear in speech bubbles that emanate from the avatars. They remain on screen for a limited duration determined by the length of the message in characters, or until a new message is submitted by that user. Text message within other groups are not visible while in a conversation circle as other discussions are not considered to be relevant. This also ensures that the messages are not overlapped. The text bubbles themselves are opaque so as not to completely obscure the environment.

This concept is still in its early stages. Although multiple social circles can exists within a single environment, each circle currently only accommodates for up to six group members. This particular formation has been designed with purely chat applications in mind however we believe that it can be modified to support other uses of CVEs. A final point to note is that few CVEs currently support the use of shadows. It is however possible to simulate them with the use of algorithms. This technique is commonly used within computer games such as Tomb Raider which is available for the Sony Playstation and the PC.

The Convention of the Conversation Circles

Stage 1: Initiation or joining a conversation

Table 2 shows some gesture that can help visually support or this part of the interaction.

Meaning to Convey	**Visual Gesture**
Available for conversation	Remain Neutral
Unavailable for conversation	Ghosted avatar (semi-transparent effect)
Ask for conversation (close/far)	Move into a space, meet gaze, perform questioning gesture
Accept an offer or conversation (close or far)	Smile/move towards the conversation space

Table 2: Table of Gestures to support the initiating of a social interaction.

Stage 2: Engaged within Conversation

Once engaged in conversation there are several visual operators that can help facilitate fluent conversation. examples of these are shown in table 3

Operator	Visual Gesture
Greeting	Wave a hand
Give floor	Arms forward offering, then pulled back /offer token
Show who has floor	Hold token
Show desire to speak	Raise arm above head

Table 3: Gestures to ensure the smooth running of conversation at the middle part of an interaction.

Stage 3: Leaving Conversation

Users usually signal an intent to end an interaction some time before it actually ends. The interaction will then end only when both participants or in the case of a group, all participants, have come to a mutual agreement.

Signal	Gesture
Signal intent to leave	Body quarter turn, gaze still connected.
Say goodbye	wave

Table 4: Necessary visual clues for ending conversation

Some gestures need to be adapted and there is a need for some additional gestures for when a user joins and then participates within a group interaction. Examples are shown is table 5.

Joining	
Ask for conversation	Move onto circle parameter
In Conversation	
Show newcomer so that others can involve him	Avatar is highlighted
Show that a message is addressed to the whole group	Sweeping gaze and arm with hand out stretched.
Show a message addresses the an individual group member	Head turn towards
Leaving	
Signal to group mutual agreement has been recognised	Sweeping wave.

Table 5: Additional Gestures Needed When Conversing Within A Group.

Agent Avatars in Conversation

The expressions and gestures that need to be performed by an agent are slightly different for the middle stage of the interaction. Within this stage gesture performed by agents should serve to inform the human participant of its status. Table two provides some examples of states that the agent needs to convey to achieve a more fluent conversation using the visual channel.

Status	Visual Clue
Available	Normal
Unavailable	Ghosted
Type of agent i.e. help or policing agent	Dress code
Processing a query	Thinking gesture
Cannot interpret request	Confused expression
Require more information	Questioning expression

Table 6: Examples of necessary gesture for agent avatar in conversation

For the leaving stages a simple 'goodbye' should suffice. As conversation with agents is likely to be less involved and more functional, following the conventions of social interaction in agent conversation will be less of a requirement. All that is required to break off an interaction with an agent is a single signal to show the agent that they are no longer needed for the time being.

The Participants Field Of View

When designing the conversation space one of the key issues to be addressed is the point of view from which the participants will see the circle as well as the other avatars. Three options are available a first person view, a third person view and a bird's flight view. The first person point of view is the one delivering a realistic visual observation of the environment. Such a point of view doesn't deliver a comprehensive perception of the environment, this is due to the lack of any feedback from the avatar actions (such as balance, touch and force perceptions). Without visual feedback, one cannot be fully aware of what the avatar is currently doing with a first person view [3]. A third person view is basically located behind the avatar slightly above the head, this lets the user not only see what the avatar is looking at but at the same time be aware of the avatar actions, postures and immediate surrounding. Because this view is fixed relatively to the avatar no disorientation can occur when there is a mismatch between the avatar gaze and the user's point of view. Unfortunately this is the case with the bird's flight point of view. The user has to be aware of this fact and not attempt displacements when such mismatch occurs. With the third person view it is possible to have all the participants in the circle of conversation within the field of view as one would expect it to be the case. With a bird's eyes point of view, comes the possibility of viewing more than just the local conversation circle. As a thumb of rule, first person view is to be used only for formal or urgent one to one dialogues, third person view in most cases, and bird's eye view when inquiring about the environment.

GROUP STATUS

We argue that there is a group behaviour that needs to be visually represented. The social circle concept allows for a visual grammar of a social group to be developed. For example it may be necessary to provide a visual clue that indicates whether or not the group is full. If a group is full then the light source can turn red. Other users can then understand at a glance the state of the group.

Social circles are owned by the participants contained within them. This means that only group members can change the status of the group which could be :

- Private : Invited participants only.
- Semi-private : Only acquainted participants can join.
- Public : open to all.

From another perspective conversation spaces are characterised by the type of conversation they are holding. Casual conversations are the least regulated, and no control or restriction is put upon participants regarding interruptions and entering and leaving the group. Formal conversations have more restrictive rules. Participants exchange a token which act as a moderator of the conversation.

Group Condition	Visual Symbol
Group is close to newcomers	Emanate a red glow
Group is open to newcomers	Normal
Group is looking for newcomers	Open spaces around the parameter
The group context is social or formal	Dress code of the avatars

Table 7: Some Visual Clues to Group Status.

CONCLUSION

We propose that to arrive at a useful set of gestures, one should draw on a number of sources. Firstly gestures that support the social conventions of a CVE, secondly from bodily actions that are described by acronyms, emotion icons and keywords that are found within the text messages.

We have Set guidelines for avatar design that take into account the identity of the owner and some of his or her relevant characteristics, and we have proposed a library of predefined gestures and expressions. Such a library should facilitate communication and expression in a CVE.

We have proposed conversation circles that address the following issues :

1. Navigating into a conversation space, allowing a clustering of conversation groups.
2. Visual definition of the conversation circles within the CVE.
3. Allows all participant to be visible at once and a multi-modal communication is achieved.
4. Delivers conversation groups which can be expanded and reduced according to the number of participants.
5. Ensures that text messages do not overlap

Further developments should result in the spatial arrangement of the conversation circles in social groups. Enhanced avatars so that members of a group are recognisable by some visual characteristics. The possibility for participants to initiate a conversation simply by clicking on an individual to initiate a group, or by clicking on an existing group which is open to newcomers.

This paper has proposed a set of design guidelines towards a truly multimodal CVE which would facilitate the communication between users and provide better ways for the participants to express themselves.

ACKNOWLEDGMENTS

This work was undergone at the now disbanded NVRCAD centre. The NVRCAD and part of this work have been funded by a JISC-JTAP grant. Poser is a software for animating 3D human figures.

* The first author is now at the Electronic Research Group, Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, England.

REFERENCES

1. Benford S., Bowers J., Fahlen L.E., Greenhalgh C., Snowdown D., Embodiments, Avatars, Clones and Agents for Multi-User, Multisensory Virtual Worlds, *Multimedia Systems*, *5*, 2, 1997, 93-104.
2. Capin, T.K., Pandzic I.S., Thalmann N., Thalmann D., Realistic Avatars and Autonomous Virtual Humans in VLNET Networked Virtual Environments, *Proceedings of From Desktop to Webtop: Virtual Environments on the Internet, WWW and networks,* (Bradford, UK, April 1997)
3. Guye-Vuillieme A., Capin T., Pandzic I., Thalmann N., Thalmann D., Nonverbal Communication Interface for Collaborative Virtual Environments, *Proceedings of CVE 98,* (Manchester, UK, June 1998), 105-112.
4. Pandzic I., Kalra P., Thalmann N., Thalmann D., Real-Time Facial Interaction, *Displays 15,*3 , 1994, 157-163.
5. Damer B., Putting a Human Face on Cyberspace: Designing Avatars and the Virtual Worlds they Live in, *Proceedings of SIGGRAPH97* (Los Angeles, CA, USA, August 1997), ACM Press, 462-464.
6. Balder N., Philips C., Webber B., Simulating Humans, Oxford University Press, Oxford UK, 1993.
7. Thalmann N., Pandizic I., Moussaly J., The Making of the Xian Terra-Cotta Solders, *Digital-Creativity*, *8*, 2; 1997; 59-66.
8. Argyl M., Bodily Communication, Second Edition, Methen & Co., Ltd., London, UK, 1975.

9. Vilhjalmsson H., Autonomous Communicative Behaviors in Avatars, available online at the following location: http://hannes.www.media.mit.edu/people/hannes/msthesis/

10. Blaxxun Virtual Worlds is a commercial product available online at the following location: http://www.blaxxun.com

11. Fisher J., The Theatre of Yesterday and Tomorrow: Commedia Dell'Arte on the Modern Stage, The Edwin Mellen Press, Lewiston.

12. Sturman D.J., Zeltzer D., A Survey of Glove-Based Input, *IEEE Computer Graphics and Applications,* Special issue on Virtual Reality, *14*, 1, 1994, 30-39.

13. Online Traveller is available at the following location: http://traveler.onlive.com/community/trav2dest.html

14. Earle N., Beardon C., The Role of Obligation Within Virtual Encounter, *Proceedings of CVE,* (Manchester, UK, June 1998) 57-66.

15. Damer, B. (1996) Inhabited Virtual Worlds, *ACM Interactions*, Sept-Oct, Vol III - 5, 27-34.

16. Kurlander, D. Skelly, T. Salesin, D Comic Chat, *Computer Graphics Proceedings, Annual Conference Series, SIGGRAPH,* (New Orleans, Louisiana, August 1996), 225-236.

THESIS CONTAINS

VIDEO CD DVD TAPE CASSETTE