ENHANCING TELEPRESENCE WITH MOBILE VIRTUAL PROXIES

SEAMUS HICKEY

Faculty of Science,
Department of Information
Processing Science and
Infotech Oulu,
University of Oulu

OULU 2005

SEAMUS HICKEY

ENHANCING TELEPRESENCE WITH MOBILE VIRTUAL PROXIES

Academic Dissertation to be presented with the assent of the Faculty of Science, University of Oulu, for public discussion in Auditorium IT 115, Linnanmaa, on May 20th, 2005, at 12 noon.

OULUN YLIOPISTO, OULU 2005

Supervised by
Professor Petri Pulli

Reviewed by
Professor Jari Porras
Professor Jukka Vanhala

ISBN 951-42-7683-3 (nid.)
ISBN 951-42-7684-1 (PDF) http://herkules.oulu.fi/isbn9514276841/
ISSN 0355-3191 http://herkules.oulu.fi/issn03553191/

OULU UNIVERSITY PRESS
OULU 2005

Hickey, Seamus, Enhancing telepresence with mobile virtual proxies
Faculty of Science, Department of Information Processing Science, University of Oulu, P.O.Box 3000, FIN-90014 University of Oulu, Finland, Infotech Oulu, University of Oulu, P.O.Box 4500, FIN-90014 University of Oulu, Finland
2005
Oulu, Finland

Abstract

Traditional telepresence systems are comprised of a person remotely controlling a robot in a hostile environment while receiving visual feedback from a camera mounted on the robot. While useful for a number of applications, this model is not particularly useful for everyday work applications. The size of the robot is intrusive, the robot needs to be designed for specific interactions and only one person can use the robot at any one time.

This work seeks to address this problem by replacing the physical proxy with a virtual proxy. The purpose of this virtual proxy is to enhance the system by providing improved support for multiple users, interaction and navigation in the remote environment. This enhanced, or improved, version of telepresence is termed TeleReality as it combines elements of virtual reality with traditional telepresence technologies. To achieve this goal, the basic building blocks, or constructs, of traditional telepresence systems need to be changed. This thesis identifies and evaluates the base constructs needed to build any TeleReality system. These constructs include the need to support navigation of the remote environment and this is achieved by using a network of cameras and image processing software to calculate the various perspective viewpoints of the users. These constructs govern the means in which this collection of cameras are organised and connected. Each user receives a common set of video images from which they calculate their own perspective viewpoints, and consequently supporting a multi-user system. Interaction within the remote environment is promoted using ad-hoc networks and augmented reality technologies. Constructs also cover security and privacy issues that arise from using multiple cameras by adopting both an organisational and technological viewpoint. The focus is on establishing trust within the system by divesting control to the user. An example of these constructs is given by the implementation of a TeleReality model called a 'Visual Cell' system.

The conclusion of this work identifies the constructs that are needed to support a telepresence system using a virtual proxy where the primary interaction framework is informational exchange, although physical interaction can also be supported depending upon the environmental support. This work also identifies the technical issues that require additional research for the implementation of a TeleReality system, from the need for improved image processing, video codec's, broadcast and ad-hoc security protocols, software architecture, registration and the availability of suitable head mounted displays.

Keywords: augmented reality, collaborative virtual environment, media spaces, telepresence, ubiquitous computing, virtual reality

Preface

I had for many years worked in the fascinating world of mobile phone technology before I had the pleasure of actually owning my own mobile phone, which I acquired through a rather unusual scheme from my then employer, Ericsson. At that time it was not unusual for a large number of my colleagues in the area to totally forswear ever acquiring a mobile phone due to the intrusiveness of the device, but the sheer convenience in coordinating various forays into the city nightlife was the essential tipping factor.

In research, it is quite often that we are called upon to envision trends, technologies and visions whose benefits are not always immediately tangible, or have uses other than what we forecast. This was my position as I left the rather narrow focus of a company project to take up work in the exciting field of mobile augmented reality in 1998 as part of the Academy of Finland funded PAULA project. At that time, my knowledge of the world of virtual reality was in stark contrast to my expertise in mobile communications, (beside a few games of free space), and I was certainly eager for the challenge, as well as the learning curve involved.

My first task was to build an augmented reality meeting environment based on some ideas from my supervisor, Prof. Petri Pulli and the pictures of Metsavainio. I approached this problem by first looking at what constituted a meeting and concerned myself with the formalities involved and how to support this by augmented reality means. Problems occurred immediately, or specifically, in our definition of augmented reality at that time. Augmented reality is a system where information is overlaid on the surrounding environment to the benefit of the user. The construction of an augmented reality system is adapted for this process. A video conferencing system, on the other hand, requires the user to see an entirely different world, and one that the user cannot take part in a normal way. An augmented reality user can physically interact and walk around their environment, but in video conferencing, the user must stop moving around (for safety) and cannot interact without giving specific means (robots) to do so. The general conclusion therefore was that augmented reality describes a condition where a user uses a technology in the local vicinity, whereas to use the same technology over a video conferencing/telepresence system described something else. That ‘something else’ became the focus of my research for the following years and is the subject of this thesis.

But immediately, the challenge was to have some form of augmented reality leaning mobile video conferencing system. The key part here was to instil a sense of presence for the remote user using AR technology. At this time, a parallel project was experimenting with a remote controlled car using a head mounted display connected to an omni-directional camera. I then had the idea of borrowing this system and using it for a meeting experiment. First, the car was removed because we did not want to have control problems with it on the table. Second, a great deal of effort went into determining what position we should place it in. Finally, we conducted our experiment with some interesting results. Mainly, by removing the car, we had inadvertently created a system that was multi-user. The second problem that I noted was that getting the right position for the camera was still a major challenge and finally, that our system really had no means for informational transferral, which we quickly noticed was essential. Without this ability to handle information, our system was an interesting one, but only marginally useful, at least from a practical point of view.

It was at this point that I no longer considered what we were doing as an augmented reality based system, but a telepresence system. I then checked for solutions to the problems encountered. Taking the problems we had with our users' visual viewpoint. I discovered that while some researchers had addressed the problems of creating multiple view points, none covered the degree of freedom I envisioned, i.e., the ability to 'get up' and move around, maybe travel to another room. It was not, I felt, from a lack of technical difficulty, but simply, the effort was considered too much, and while some researchers hinted at doing this, for example, Paul Milgram, that was as far as it progressed. Telepresence systems do deal with this, by the use of robotic devices, but this presented an enormous problem for the context of our experiment, which became the support for collaboration. An analysis on the use of robots followed, and I concluded that it was impractical for the accomplishment of the goals set out in PAULA, as I wished to have a collaborative system and there are good reasons why teleoperated robots are only commercially used for operation in hazardous environments away from people, or at least friendly people. I decided therefore that robots were limiting the thinking and envisioned doing without them and seeing if this was possible. I also determined that informational interaction was essential to the practical acceptability of our system, i.e., offering a good degree of presence or immersion was not sufficient for a system to become acceptable. After these conclusions, in mid 1999 I proposed a solution called visual cells.

Work continued on building and experimenting with this system, although it became clear that a satisfactory implementation for usability testing would have required resources beyond that available from the excellent PAULA project. Work, therefore focused on proof of concept with the goal of achieving cohesiveness among the constructs, the development of a user interface, as well as identifying the most limiting technological factors. By December 2001, the basic work had been completed and only the writing of the thesis was left, as well as acquiring the necessary course credits. A rough draft was prepared in august 2002, but serious work did not recommence until the summer of 2003, which was made more difficult by the fact that I was recovering from a traffic accident, and the first version was submitted for review in September 2003. The production of this thesis work has taken six and a half years, making it the longest commitment of my life so far and its completion has being a most satisfactory conclusion.

Seamus Hickey February 2005

Acknowledgements

I received considerable useful advice and moral support throughout this thesis. Of these, my supervisor Prof. Petri Pulli has had the greatest contribution. Not only has he taken the time to twice arrange for me to come to Finland, he also proposed the basic topic from which this work has evolved. Prof. Petri Pulli, proposed as a research activity, the study the use of Augmented Reality technology in mobile meeting settings. Prof. Pulli also suggested the use of the Multisphere model as an improvement on an existing model of the authors. Added to this is the considerable amount of time spent providing hints for future research directions, critical analysis and the reviewing of my thesis. He has also given me considerable assistance on dealing with a range of problems from administration, teaching and cultural differences.

I would like to thank Prof. Pentti Kerola and Prof. Olli Martikainen who assisted me in the selection and application of the research methods utilised in this work. Additionally, Prof. Martikainen provided an invaluable internal review of my work. I also need to thank Peter Antoniac and Tony Manninen for the realms of useful advice, analysis and general moral support that got me through the difficult periods, and a special note for their participation in regular 'research' sessions. I would also like to acknowledge the contribution of Prof. Kari Kuutti, who gave me initial advice and help with the world of CSCW and help with the mobile VR experiment. I would also like to thank Mr. Tino Pyssysalo for help with the VR experiment, but most importantly for the great talks we had discussing our ideas that resulted in the formulation of my research. I would also like to thank Sami Ahola, for lending me his equipment for the VR experiment and subsequently giving me the basic code for handling the meteor framegrabber. I would also like to acknowledge the work carried out by two summer students working under my direction, Mr. Fabien Arrive and Mr. Koldo Echvarren. Mr. Arrive implemented a broadcast protocol while Mr. Echvarren implemented the objects and cycle/select scheme for the visual cell.

On a personal note, I would like to thank my parents, James and Breeda Hickey who financed my university education. Lastly, and certainly by no means least, I would like to take the time to offer my appreciation for my wife, Taina Hickey and family for their patience and understanding.

Abbreviation

AA	Access/Access
AR	Augmented Reality
AV	Augmented Virtuality
AVO	Audio Video Object
BAN	Body Area Network
CC	Cyberworld/Cyberworld
CE	Communication Element
CP	Cyberworld/Platform
CSCW	Computer Supported Collaborative Work
CVE	Collaborative Virtual Environments
FOV	Field Of View
GPS	Global Positioning System
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
HMD	Head Mounted Displays
HRTF	Head Related Transfer Function
IA	Interconnectivity/Access
II	Interconnectivity/ Interconnectivity
IP	Internet Protocol
ISDN	Integrated Services Digital Network
IST	Information Society Technologies
LCD	Liquid Crystal Display
MPEG	Moving Pictures Expert Group
MR	Mixed Reality
MTV	Multiple Target Video
PDA	Personal Digital Assistant
PI	Platform/Interconnectivity
PP	Platform/ Platform
QCIF	Quarter Common Intermediate Format
RFID	Radio Frequency Identification Tags
RGB	Red, Green, Blue

RV	Reality -Virtuality
UDP	User Datagram Protocol
UMTS	Universal Mobile Telecommunication System
TCP	Transmission Control Protocol
VE	Virtual Environments
VR	Virtual Reality
WEP	Wireless Equivalent Proxy
WLAN	Wireless Local Area Networks
WSI	Wireless Strategic Initiative
WWRF	World Wireless Research Forum

Table of contents

1 Introduction

The history of man is dominated by the need to work, travel and communicate across long distance. Each of these has become fundamental for trade, diplomacy, the exchange of ideas and knowledge. These pursuits have been hindered by the restrictions of our physical being and man has consistently sought ways to overcome our spatial limitations. The development of electrical based communication, starting in the mid 1840s, brought a new medium in which people could communicate instantly with their peers without the need to travel. The continuous development of this new medium brought advances in telephony, film and telemetry. The science fiction author Robert Heinleins was intrigued by these new technologies and the concept of our own presence. Our ‘being in this world’ is an age old concept that has intrigued philosophers for centuries. The constraints in which we humans operate are determined by our physical senses, the way in which we receive physical stimuli and how they affect our behaviour and actions. In 1942, Heinleins set out to write a short novel ‘Waldo’, in which a genius by the name of Waldo F. Jones who lives in a zero gravity home in orbit around the earth, builds hardware devices called ‘waldos’ that respond to his hand movements to control robotic arms on earth, while all the time receiving stereo images of his actions. These were the first concepts of teleoperated robots, and from these writings came an understanding that using such systems requires a degree of presence that was unreachable before. These ideas about combining our sense of presence, robots and long distance communications were to inspire future thinkers into working out the means by which one can achieve presence and perform work over long distance. The development of technologies has allowed us to replace real stimuli with artificially created stimuli so that people can be convinced that they exist in an alternative world other than the one they currently exist in.

Teleoperation has become an important area of study whereby people can operate devices in distant, remote and hostile environments like space, and this feeling of ‘being there’ for the operator is called telepresence. While telepresence systems have become important over the years in many activities such as space exploration, military systems and maintenance, the promise of telepresence has not met the reality. Telepresence has some limitations which must be understood so as to understand why it is confined to hostile environments and not the everyday world in which we live.

1.1 Limitations of Telepresence

A basic fundamental component of Telepresence involves a camera mounted on some type of physical *proxy*. A proxy is '*the authority to represent someone else*' (Pearsall 2001) and in telepresence this usually means having something like a remote controlled robot representing the user in the remote environment. Importantly, this standard configuration of a telepresence application means that only one person i.e., the controller can use the physical proxy and, hence, gets the maximum benefit of the experience. Others may watch, but they have no control. Other persons that wish to explore, or *navigate*, the remote environment would require their own proxy.

A second issue relates to the nature of the physical proxy itself. In communication between people, it is important to maintain a degree of eye contact or gaze awareness. In practice, this means that the camera position should be located at a natural eye level, requiring that the proxy should be quite large. Additionally, to perform tasks within the environment and with others, i.e., *interact*, the proxy must be equipped with specially designed manipulation devices, which also must be carefully designed and controlled to prevent damage to the environment. This combination of size and complexity means that each physical proxy is large and expensive. It also means that it has an obtrusiveness quality when deployed in confined work spaces were people are present, e.g., an office workplace (Paulos & Canny 1997, 1998).

Giving each person a physical proxy in an everyday working environment is not practical, therefore it is not easy to support multiple users. Taking video conferencing as an example, for many people to use this system, the camera perspective must be fixed. Additionally specially designed equipment is needed to carry out basic interactive work. Therefore, to make a telepresence system *multi-user*, one must sacrifice the ability to *navigate* and restrict the ability to *interact*, while to support *navigation* and *interaction*, one cannot ensure *multi-user* support.

1.2 Enhancing Telepresence

The goal of enhancing telepresence is to find a solution for a comprehensive system that supports a collaborative work environment between people who are located both remotely and locally. Such a comprehensive telepresence system should be *multi-user* and support *navigation* and *interaction*. Navigation describes the ability of people to maintain spatial awareness of their location and the ability to move about that world (Stanney *et al.* 1998). Steve Benford and his research group (Benford *et al.* 1996), differentiate between the ability to navigate around a space and to interacting with objects within that space. As an example, Bhatia and Uchiyama (1999) uses the term path planning to describe the process of automatically moving a robot around a remote location. It is important to define what is meant by the term interaction in this research topic. According to the NRC (1997), the meaning of the term 'interactivity' is not fully explored. For Collaborative Virtual Environments (CVE), Singhal and Zyda (1999) describe the need for a shared sense of presence, time, a means to communicate and a way to share. The interaction definition used here follows the natural interaction occurring in real life, which means that most of the issues are considered as content matters, not merely tied to any specific input/output type or interaction technique (Manninen 2001). As a consequence, interaction is mainly concerned with the flow of information, e.g., data, visual cues, between people sharing the environment and with the environment itself. Multi-user in this context means that each person has a presence within the environment that they control and are responsible for, or in other words, that the limits of their behaviour is not controlled by others. The author places no limits on the number of people who can participate other than that which is constrained by the technological limits, rather than the conceptual limits.

The primary obstacle to this solution is the physical proxy itself. The solution proposed in this thesis is to replace the physical proxy with the concept of a *virtual proxy*. The virtual proxy attempts to replace all of the functionality of the physical proxy, but without any of its problems, such as obtrusiveness, cost, and lack of flexibility.

It should be possible to create a virtual proxy using advances in a number of current technologies. To do so would mean creating a support system. The virtual proxy could be constructed by using Virtual Reality (VR) technologies to generate computer-generated avatars, which can be viewed by devices such as wearable Head Mounted Displays (HMD). For the remote user, this avatar acts as the point of their camera perspective, and studying the latest advances in image processing systems is needed to calculate the correct perspective.

Interaction is also difficult. Many different interactive tasks involve transferral of information between people and between people and the environment e.g., by the use of a gesture to turn on a light switch. Many interactions could be transmitted by telecommunication means rather than physical interaction. Additionally, many tasks often occur spontaneously and dynamically in an ever-changing environment. Using the latest techniques from ad-hoc networks could help create a dynamic and manageable environment to support a complex collaborative environment. Using the latest advances in user interface techniques from Augmented Reality (AR) should help simplify the interaction process within the environment.

1.3 Research Question

In order to develop such an enhanced telepresence system, researchers are confronted with the problem of defining what exactly the components that comprise that system are. That is not to say that some parts of the problem have not been dealt with already, but rather, that those researchers only focused on narrow problem domains of related systems. As an example, research has looked at providing a flexibility of view in a remote environment purely as a computer vision problem, but never dealt with interaction in that environment. Research in Augmented Reality has focused on the means of interactivity within the environment, but never applied those findings to telepresence systems. The most likely practical explanation for this is that those researchers did not expand beyond their field, most likely due to the singular disciplinary approach taken. By taking a more multi-disciplinary view, one can draw from the research of these fields, as well as from work done on telepresence, collaborative virtual environments, video conferencing, awareness systems, augmented reality and human-computer interaction, to define the basics of a more comprehensive system.

The key to enhance telepresence is to look at developing the components of a virtual proxy, as opposed to a physical proxy, such that it includes such features as multi-user support, navigation and interaction. Therefore the basic research question is framed as

> *'What form should an enhanced telepresence system relying on virtual proxies take so that its multi-user, interactive and navigable'*

The purpose of this thesis is to address this research question. The key areas of multi-user support, interaction and navigation are specifically addressed. It is anticipated that the solution will involve the use of a large number of static cameras. A consequence of using a system involving cameras is that the issue of security and privacy will arise and consequently, *security and privacy* related to telepresence will also be examined. The means in which the form of this enhanced telepresence system is determined through the selected research method.

1.4 Research Method

The choice of the research method was carefully made, as it determines the approach and success of the work. The research is based on many existing systems, no entirely new scientific theory on presence is likely to be forwarded, and that would rule out scientific methods. The action research approach was not used as it was considered that in order to carry out this form of research, an existing experimental system should be available. This was not possible, as no such TeleReality system existed. Different theory research methods were also considered. It could be argued that the resulting work from the thesis could be considered a theory. However the breadth and multi-disciplinary scope of the work was considered to be too broad a topic, as the theory research method requires a more narrow focused approach. In addition, the basic elements of the system would still have been required in order to test out a single theory. Also, as it was anticipated at the beginning that many of the different concepts were already tested and established, the focus of the work concentrates on combining these into a new single system. After careful consideration of all of the available research methods, the Constructive Research approach, described by March and Smith (1995), was considered to be the most relevant and applicable for this research work.

1.4.1 Constructive Research method

The constructive research approach has four levels, shown in Figure 1, the *construct*, *model*, *method* and *instantiation*. March and Smith (1995) define these as

Constructs:	Also called concepts, these '*form the vocabulary of a domain. They form the specialised language and shard knowledge of a discipline or sub-discipline*'
Model:	The model is a '*set of propositions or statements expressing relationships among constructs*'.
Method:	A method is '*a set of steps (an algorithm or guideline) used to perform a task. They are based on the underlying constructs and methods*'.
Instantiation:	This is the realisation of an artefact in its environment.

The constructive research approach is mainly favoured by those of the Engineering profession who are more used to constructing new objects or approaches. Each level has a Build and an Evaluate part. Essentially, one must build the construct or implementation, and then evaluate that build in order to determine their accuracy.

	Build	Evaluate
Constructs		
Model		
Method		
Instantiation		

Fig. 1. Constructive Research framework

1.4.2 Addressing the research question with the constructive method

In applying the constructive research approach to the research question outlined for this thesis, one must first determine at which level that research question is currently situated. Currently, there are no systems that match the enhanced telepresence system requirements as proposed in this thesis and as a result, the constructs that constitute the system are not known. In order to develop an instantiation, method or model of the problem, one must first know what those basic constructs are. These basic constructs are essential to any engineered system. As an example, let's take a pen. The basic constructs or constituent parts would be ink, a container to hold the ink, a head of some type to govern the transfer of the ink to paper and an enveloping case. There can be different relationships between ink colours, different models, or different methods of how the pen works, yet the pen still has the basic constructs of what classifies it as a pen, which is the starting point for all of the following variations.

1.4.3 Research process

The research method followed an iterative process based on an *analysis–build-evaluate* approach shown in figure 2. The driving force behind the analysis phase was based on the use of scenarios. These scenarios were built by a through understanding of the research field that was acquired through literature surveys and the specific goals set out within the Paula project to develop new mobile telecommunication services based on mobile Augmented Reality technologies. The build part of the cycle implemented these scenarios by a process of identifying the constructs from the scenarios and related literature and building an experimental system. The outcome of the build stage was evaluated on a qualitative basis. The goal was to understand if the approach proposed by the scenarios and the identified constructs were valid in terms of their logical cohesiveness, practicality, strengths and weaknesses. There were two phases, or iterations, of this process with the outcome of the evaluation used as the input to the second iteration.

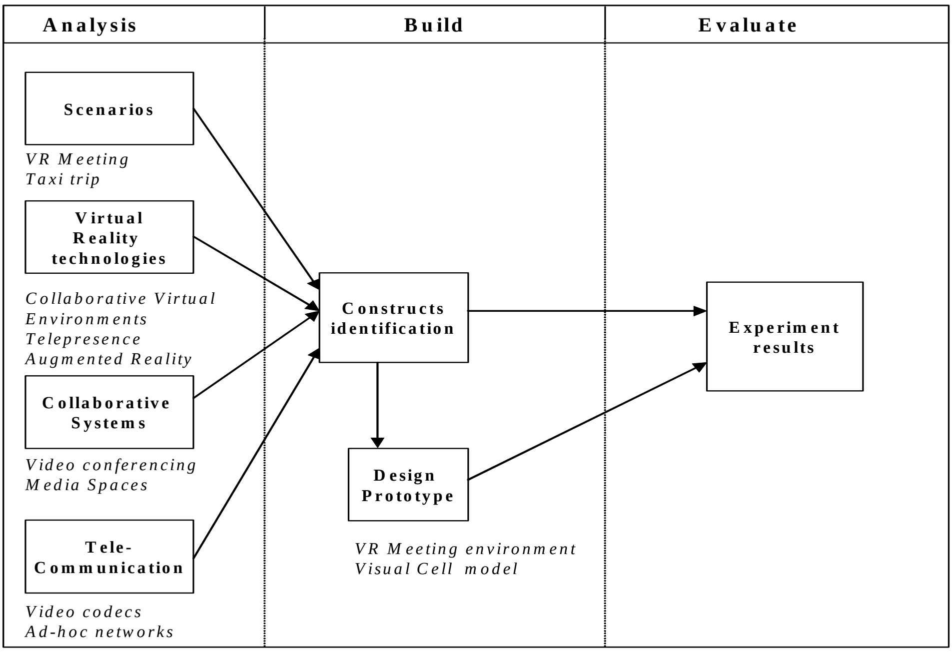


Fig. 2. The Analysis-build-evaluate process. Analysis consists of scenarios and a literature survey. The outcome of the analysis phase was used to identify and build constructs which resulted in a design prototype. These design models were developed and evaluated on a qualitative basis, with particular attention on the validity of the constructs.

1.4.3.1 Phase I

The focus of phase one was to study how one could use mobile augmented reality systems to support meetings, whereby the mobile participant is remote. This resulted in the basic scenario for a simple mobile meeting (Hickey 1999). In this first iteration of the process, the goal was to support a feeling of presence by the mobile user in the meeting environment so as to improve their performance.

Analysis: To support this work and help build an implementation of this model, a literature survey on Augmented Reality, video conferencing and media spaces were carried out. The literature survey on video conferencing also included studying the behaviour of people in meetings as well as the behaviour inherent when people communicate in meetings through technologies. The analysis showed that the technological needs were different. When using augmented reality in a local space and when operating over a telecommunication medium from a different space. The analysis phase also showed the importance of the users perceived position in the remote environment, in particular gaze awareness.

Build: The build section was carried out by identifying the main constructs for a mobile meeting environment. This was greatly facilitated by the use of an exiting system whereby an omni-directional camera was attached to a remote control car to provide feedback to an augmented reality user controlling the car. Taking the camera from the car and keeping it in a fixed position in the meeting room provided a more immersive visual feedback to a remote augmented reality user. The use of this system provided advantages that were not originally identified by the analysis phase, namely the support for multiple users. Had the car been used as part of the experiment, it could not be said with certainty that the importance of a multi-user system would have been concluded.

Evaluation: The mobile meeting experiment was then evaluated on a qualitative basis. There were considerable technical difficulties in the performance of the system but the experiment did support some basic ideas about the system. More importantly however, was that the experiment underlined a number of flaws in the original thinking of the analysis phase. In particular it became clear that an improved mechanism for supporting interaction was essential to the usability of the system. Also, while the omni-directional camera supported some flexibility of view, it was not comprehensive enough to match the requirements proposed in the analysis.

1.4.3.2 Phase II

The results of the evaluation from phase I were used as a basis for phase II. Phase I evaluation identified the key characteristics for an enhanced telepresence system as that of multi-user support, interaction and navigation. Phase II was therefore focused on developing constructs and a system to support these characteristics. It was not clear at the beginning of phase II if those constructs would differ significantly from existing systems, hence the need for this second iteration.

Analysis: The analysis phase used the results obtained from phase I. This phase looked at a number of potential scenarios. The primary scenario of interest was a taxi trip scenario developed as part of the CyPhone project (Kuutti *et al.* 1999). The basic scenario offered some promise, and some modifications of this taxi trip scenario was made to explain the workings of an enhanced telepresence system. The analysis phase also looked at strategies for constructing world environments based on video data. The combination of these scenarios and research activities were fed into the build stage.

Build: The build phase took the scenario and technological solutions proposed from the analysis phase to identify a set of constructs necessary for an enhanced telepresence system. These constructs were used as the basis for a visual cell model, one capable of meeting the goals set out in the analysis stage. A description of the analysis and build phase that led to the visual cell model (Hickey *et al.* 2000).

Evaluations: To evaluate the visual cell model, and hence the constructs, an implementation of the visual cell model was made. This implementation was evaluated to ensure that the constructs proposed fitted together logically, were located in the correct position and to see if there were any elements missing from the system. As a result of this phase, the existing constructs were deemed satisfactory. However, it was also determined that issues of privacy and security should be incorporated as integral parts of the system. This was the final stage of the research process.

1.4.4 Construct charts

March and Smith (1995) do not propose a visual method for presenting the research constructs. This thesis does visualise them by building a ‘construct chart’. This is supplemented by a table which shows how the different constructs would relate to different technologies. This chart *should not be confused with a model* in the constructive research approach. The construct chart shows the general relationship between each construct. The model in the constructive research frame can have multiple elements as part of one research construct. The model in the constructive research field defines highly detailed and specific technological relationships between every part of the model. So as an example in a virtual reality environment, a construct may be sensor display, which means that to use a virtual reality system one needs to provide a means for viewing sensory data, a fundamental construct for all VR systems. One model based on this construct would define a system using only a head mounted display and explicitly define the relationship with other elements of the system. Another model of the sensor display construct would include a HMD, force feedback and sound.

1.5 Introducing a name - TeleReality

The enhanced telepresence system proposed in this thesis is described as TeleReality, which is a combination of the words **tele**presence and virtual **reality.** There are two advantages for coining the term TeleReality. The first advantage is that it negates the need to continually refer to the solution proposed here as ‘an enhanced telepresence system as set out by the research question’. The second advantage is that it is useful to differentiate the system proposed here from other technological implementations of teleoperation, CVE’s and augmented reality.

1.6 Contributions and Results

The results of this thesis show that it is possible to have a telepresence system that does not depend upon physical proxies, but rather virtual proxies. The results also show that it is possible that such a system can support multiple remote participants. The thesis also shows that the remote environment can be navigated by the user, the extent of which is determined by the sophistication of a supporting camera systems used. The thesis also shows that a highly interactive environment can be supported between the user and the remote environment and with other people located in the remote environment.

The constructs for such an environment are identified and revolve around the need for a multi-user system that can be navigated, and that supports interaction between remote TeleReality users and local users. This system is based on the use of a fixed communication network connecting the TeleReality user to the remote environment, and from there, ad-hoc network elements form the basis of connections with the remote users in the environment.

To create an environment that can be navigated, a ‘*world*’ is created by establishing a system of cameras from which image processing software can calculate multiple perspective images. This world can be 2D or 3D, dynamic, i.e., updated in real-time, or static, i.e., not real-time. The world is further sub-divided into smaller groups of cameras covering sub-spaces, called *partitions*. As a consequence, the system must support navigation, i.e., the ability for users to *handover* from one partition to the next. The important factors in the calculation of this are the user’s *position*, *perspective* and *movement*.

The advantage of this camera system is that each user can receive the same video images, but then utilises their own image processing software to calculate their own perspective viewpoint. The video images may be sent in a number of different ways, e.g., broadcast video. This ability to calculate unique perspectives based on common data enables multiple users to utilise the system simultaneously, thereby supporting a multi-user solution.

Interaction was identified as an important component for providing presence and functionality in the environment. The system is divided on the basis of *actors*, who initiate actions in the environment, *artefacts* which provide services to the environment and *transactions* which provide the communication between them. A multi-sphere model (IST 2000) was applied to define the organisation of these entities into a geographically based model of personal user spheres, local area based spheres and global spheres. This organisation enables this model to be overlapped to some degree with the partition system used in creating the visual world.

When using a camera based system like that proposed here, issues of privacy and security arise. The solution is partly organisational, partly technological. A ‘trust contract’ is established between the people in the system and those operating the system. This guarantees various levels of awareness, control, access rights, and the integrity of the system. Ad-hoc networks are considered to be unreliable at the present and hence I propose that each user forms a one on one connection to the fixed part of the network, ensuring the possibility for third party certification.

An example of a TeleReality model called the ‘*Visual Cell Model*’ that is based on these concepts was proposed. A partial implementation of the visual cell model was constructed within the project PAULA (2003). In implementing this model, it is shown that many of the core constructs are essential for its operation. It was also possible to show that the deficiencies within the implemented system were the results of the constructs that were not implemented.

1.7 Importance of the Research

The importance of this research is that it addresses a long term problem with telepresence systems that heavily restrict their acceptance as a widely available commercial technology. Telepresence applications are normally used for specialist work environments, usually hazardous to humans. The most successful telepresence technology is videoconferencing, but this also has many restrictions, especially in its support for presence and as an interaction medium. The comprehensive solution presented here shows how a more immersive, interactive and multi-user system can be implemented and is the first attempt to group together everything that is necessary for such an environment. It also identifies the negative consequences of the system to privacy/security and proposes solutions to these from fields that would not normally be considered by telepresence researchers. Since the start of this research, computing power, network bandwidth solutions, HMD design and ad-hoc network solutions have steadfastly advanced. At this rate, the system I propose may be available within the next few decades. Therefore the early identification of the constructs will speed the development of these solutions.

Outside of the core issue dealt with in this thesis, elements of this work can be applied to other areas. The first area that could benefit is videoconferencing. Applying the interaction framework to existing videoconferencing systems should reduce the need for location dependent specialised interactive hardware devices. Consider that most video conferencing systems work around the exchange of information, and that this information is often taken from electronic format, transferred to a physical medium e.g., printouts and it is the manipulation of these physical items that constitute the majority of the specialised interaction devices used in video conferencing. This research shows ways in which information can be exchanged directly by staying in the virtual informational space. The advantage of this is that fully interactive videoconferencing meetings can be held anywhere there is a camera, hence no longer requiring specially designed videoconferencing locations.

The repercussions of this improved mobility of videoconferencing might not have had such a great impact with current or recent technologies. However, future mobile phones are being equipped with still/video cameras. This means that simple (and primitive) videoconferencing systems can be set up anywhere and anytime, as all that is needed is a video camera and communication link, both provided by a mobile phone. This work shows how interaction could work in this scenario. It also shows how cost benefits may be achieved with one phone acting as a proxy for the meeting and others attaching to it using ad-hoc networks. This reduces costs by having only one long haul connection, instead of a long haul connection per meeting participant.

Another area of importance that this research will impact is in remote maintenance activities. Augmented Reality is been used in some cases to help maintenance workers. This research shows ways in which a remote person can help guide an operator in a maintenance activity. In this case, a maintenance worker can call upon senior personnel and experts for assistance. The “team” can then exchange information, e.g., the assistant could put notes on the maintenance workers visual field. This just in time maintenance expertise could work in areas such as car and household maintenance, medical checkups (human and veterinary) and house appliance maintenance. In short, this can be applied to any situation where one person needs help with a task from a person who is remotely situated. The benefits come from the reduced needs of a professional to spend time travelling.

1.8 Organisation of work

The thesis is organised as follows. Chapter 2 describes the theoretical background of presence, telepresence, virtual reality and ubiquitous computing. These contain important concepts which are evident throughout the thesis. Chapter 3 gives a more technological view of the important research fields that affect this work. In particular, the constructs for existing telepresence systems are described. The other technologies deal primarily with the fields of virtual environments, augmented reality and media spaces, and ad-hoc networks. Chapter 4 describes the shortcomings and limitations in current telepresence systems. These limitations describe the research problem in more detail and the approach taken to meet these problems. A set of constructs for the system to achieve the goals of multi-user support, interaction and navigation are proposed and described. A comparison of the basic constructs between TeleReality and the other areas in the field is also given. Chapter 5 describes the components that the constructs need for interaction. Chapter 6 describes the constructs needed to provide detailed visual information in a telepresence system and support navigation. Chapter 7 describes the main constructs to ensure security and privacy within the system. Chapter 8 describes a simple model called visual cell systems which partially implements a TeleReality system. Chapter 9 contains an evaluation of TeleReality by focusing on a sample meeting and evaluation of the visual cell model. Chapter 10 contains an evaluation of the constructs, the restrictions on the system, a theoretical impact summary and a discussion on the technologies that are important but neither mature nor available within this work. Chapter 11 concludes the thesis with a summary of the rationale for the thesis, the main findings and the future research work and applications that result from this thesis.

2 Presence, Virtual Reality, Telepresence and Ubiquitous computing

This chapter summarises the different theories which are often found when discussing telepresence. The most important theoretical base is the concept of *presence*, which is a goal that those working in the field try to achieve. This is followed by a discussion of the different theories on telepresence that permeate throughout the research field. No one theory is dominant and hence one must look at them each in their own context. Lastly, the Ubiquitous computing and the Wireless Strategic Initiative (WSI) reference model are presented as these play an important role in enhancing the way that telepresence systems work.

2.1 Presence

A common measure of a successful VR system is the degree in which the system convinces the user that what they are experiencing is real. This is usually referred to as the level of *presence* the subject has while taking part in the system. For Virtual Environments (VE), this means that the user should feel such a degree of immersion that they believe they exist in a real world. For Collaborative Virtual Environments (CVE) the users have the same level of immersion as with virtual environments, but that they are also able to enhance the experience by interacting with other users present in the environment. For Augmented Reality this would mean that the user is able to accept the synthetic object overlay as been part of the world they inhabit. In telepresence, the user must feel that they are active and exist in the remote world and are aware of their surroundings. As presence is one of the means for measuring system efficiency, a common understanding of what is used to measure the system is needed. Unfortunately, no consistent means is available for this measurement process. Not only do the ideas of presence vary depending upon type of VR systems, but also the fundamental idea of presence is an open question. The study of the nature of existence is called ontology, a subfield of metaphysical enquiry. Zahorik and Jenison (1998) identify two approaches, one based on the rationalistic view based on the philosophical theories of Descartes, Spinoza and Leibniz, and the other based on the viewpoint of Heidegger and Gibson.

2.1.1 Rational View

The rationalistic view is a theory of knowledge acquisition and holds that knowledge is obtained on the basis of reason, or rationality. This is classically opposed to the idea that knowledge is acquired through experience, which is the empiricist stance. Winograd and Flores (1986) outline the assumptions for the rationalistic view.

- We are inhabitants of a 'real world' made up of objects bearing properties. Our actions take place in that world
- There are 'objective' facts about that world that do not depend on the interpretation (or even presence) of any person.
- Perception is a process by which facts about the world are(sometimes inaccurately) registered in our thoughts and feelings

In this view, the central argument of the rationalistic view is put forward. The process first starts by identifying the objects in the real world and creating a mental picture of that world. This mental picture is what is termed as reality and our presence is the manner in which we feel part of that world. There is a separation between this mental picture and the actual physical interaction with the real world. Hence, the internal mental picture is a rational logical view of the world that is perceived, as separate to the world that is physically real.

This definition is an attractive means to evaluate VR systems for researchers who wish to develop immersive environments, but find them notoriously difficult to measure. The central problem is that to measure it, one must have knowledge of the internal mental image of the subject in order to measure the system, which is an extremely difficult process of information extraction.

In VR systems, two types of measurement are proposed, subjective measures and objective measures. Both subjective and objective measurements are needed to evaluate systems (Held & Durlach 1992, Sheridan 1992). The subjective measure is acquired by obtaining the user's sense of immersion or presence in the environment and the objective measure is acquired by observing the users' performance in the system. That such a dual measurement system exists, implies that there is a possibility of conflicting results, e.g., as with systems with low presence scores, but high usability measures. Zahorik and Jenison (1998) say that researchers are continually forced to relate subjective feelings of presence to objective facts of presence, both through theory and measurement.

2.1.2 Heidegger/Gibson view

Zahorik and Jenison (1998) combine the views of Martin Heidegger, a mid-twentieth century German philosopher, and J.J Gibson, a psychologist, to provide a different view of presence in VR. Martin Heidegger was interested in the question of what it means *to be*, while Gibson is interested in perceptual theory.

Heidegger believed that it was not possible to adopt a detached and analytical viewpoint for examination of what it meant to be, since such an examination takes place in the context of certain physical, social and historical states of affairs. Heidegger advanced two concepts of '*throwness*' and '*readiness-to-hand*'.

Throwness: Heidegger claims that we are thrown into situations that we must continually act and interpret. By throwness, the subject is placed in an everyday situation which calls upon him to act. In this state of action, one is not able to stand back and analyse the situation and actions as one is too occupied with the act, in what is called '*being in the world*'.

Readiness-to-hand: While throwness describes our situation within an environment, he contends that interaction with objects follows a similar path. For example, if one were to hammer a nail, then the subject usually concerns oneself with the work at hand, i.e., hammering the nail, rather than the tools that do the work (hammer). The tools themselves become transparent to the user. The action of hammering precludes the user from having a stable representation of the equipment that he uses. The hammer is then 'ready-to-hand'. Under certain conditions, the user does form a representation of the tools, but in this case the tools are changed to 'present-at-hand', a process called breakdown, e.g., when the user drops the hammer while using it, the action is broken and the user becomes fully aware of the hammer again. Heidegger offers some ideas by linking presence to task completion, in other words our presence in the world depends upon action/interaction with the system, which is easier to measure than subjective impressions.

Gibson argues that perception is a direct process of picking up information from an information rich environment. Hence, the environment is both the object of perception and the source of perceptual information, i.e., the perceiving organisms are intimately related, and namely that perception for the organism is the pickup of information and supports action. This action-supportive information is termed *affordance*. Actions of the organism (e.g. hand) have consequences for the environment and the nature of the environment has consequences for the organism. Gibson's central idea is that there is an ecological relationship between the perceiver (the subject) and the perceived (environment).

Both Heidegger and Gibson, then, have a similar view that existence and perception are formed on the basis of interaction with the environment, or, that presence is tantamount to successfully supported action in the environment, that environment may be real, virtual as well as local or remote.

Perhaps the difference between the rationalistic approach and the Heidegger/Gibson viewpoint is that whereas the rationalistic approach concentrates on subjective/objective measurements, the Heidegger/Gibson rationale would only focus on the degree to which the subject is working in the environment. In this sense, if the subject is fully involved and present in the environment, they have little time to develop subjective views.

The essence of the debate between the rationalistic and Heidegger/Gibson views is how to measure the effectiveness of a VR system. A desktop VE with no sensoring or VR equipment might engage the user so effectively that it would score very high in a Heidegger/Gibson form of measurement, but much lower in a rationalistic setting. Taking the Gibson discussion on the relationship between the organism and its relationship with the environment, he maintains that the organism evolves to suit the perspective around us. This may well be true, but discrepancies in the visual field can cause motion sickness and other side effects. It may be true that the organism may evolve to fit the environment, but the less optimistic scenario is extinction, or our refusal to use an organism that causes users to be violently ill. The action of immersion is to effectively reduce conflicts between what is perceived in the real world and the virtual.

Sheridan (1999) argues that, in reality, there is little practical difference between the different ontology's, and that, ultimately, the design and implementation is an engineering task. He suggests that, for this reason, an engineering approach should be taken. Instead of mental and physical separation, he uses control theory to expound the sensor and motor aspects.

2.2 Telepresence Theories

Telepresence has its roots in the coupling of robotics and telematics. The earliest known work on telepresence comes from the writing of the science fiction author Robert Heinleins short novel 'Waldo' (Kac 1993). In this tale, a genius by the name of Waldo F. Jones suffers from a disabling disease and lives in a zero gravity home in orbit around the earth. He builds hardware devices called 'waldos' that respond to his hand movements to control robotic arms on earth, while all the time receiving stereo images of his actions.

It was Minsky (1980) who is credited with first coining the term 'Telepresence' in his article for OMNI, a magazine that specialised both in science fiction and science fact stories. Minsky states that the word was actually suggested by his friend Pat Gunkel. The promise of telepresence as seen by Minsky was that he proposed the development of a whole new economy of mining, nuclear, space and underwater exploration based on this technology. Since then, a number of different definitions for telepresence have emerged.

2.2.1 Sheridan's Telepresence

In 1992, Thomas B. Sheridan helped launch the journal, 'Presence', dedicated to the study of presence for teleoperator systems and virtual environments. In that first issue, he advanced his theories on telepresence. Few people have written as widely as Thomas B. Sheridan on telepresence (Sheridan 1991, 1992a, 1992b, 1999). Sheridan is dedicated to the development of teleoperator environments, i.e., the control of robotic devices at a remote site by an operator, or also called a *teleoperator*. Sheridan's view on telepresence can be summed up as the '*feeling of actual presence at the worksite*'. Sheridan also calls telepresence a '*compelling illusion*'. The goal, as he sees it is '*the ideal of sensing sufficient information and communicating this to the human in a sufficiently natural way that she feels herself to be physically present at the remote site*'.

The initial definition of telepresence from Sheridan's point of view is that it is a subjective experience where the teleoperator loses their awareness of their local environment, becoming convinced that they exist in the remote environment. The goal of this achievement is aimed, from Sheridan's point of view, at increasing the effectiveness of the teleoperator in performing their task in the remote environment. By argument, the more control and presence the teleoperator has of the remote environment, the more efficient will be their working task. Towards this task, Sheridan defines three constituent parts to necessitate a telepresence system,

- Fidelity and richness of information presented to the operator, i.e., the transmission of important and relevant sensory information to the operator.
- Dexterity of the control of sensors in the remote environment, e.g., the ability of the observer to modify their viewpoint, or to reposition his head to alter binaural hearing.
- Ability to make an impact on the remote environment e.g., the extent of motor control in which objects in the environment can be changed relative to each other.

The strength of a telepresence system is determined by how effectively a system meets each of these three criteria (Sheridan 1991). These determinants to presence are shown in figure 3. Sheridan noted that providing sensory information was the dominant factor in information flow, i.e., accounted for the bulk of the transmitted information over a telecommunication link. Sheridan does not give accurate measurement values for this axis, more a relative guideline which says that as you increase along each axis, one should near the point of perfect presence.

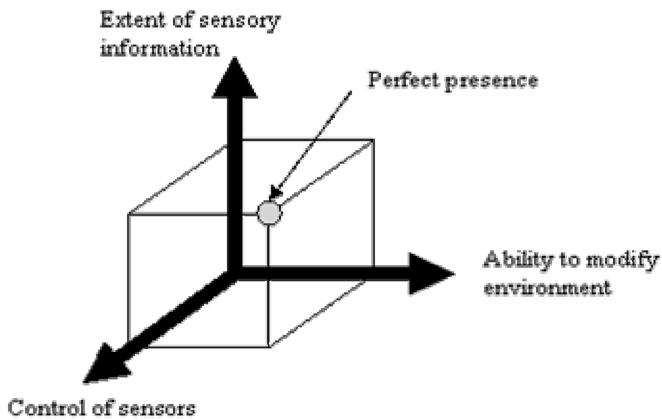


Fig. 3. Sheridan's (1991) determinants on presence for a given task in a teleoperator environment.

Sheridan determines that the effect on presence can also be heavily influenced by the normal filtering used in a teleoperator system, i.e., the effect of interfacing with the control panel, and the effect of delays brought about by the telecommunication media, and the degree of automaticity of the robot.

Sheridan suggests that in measuring this model, there are two effective ways, one based on subjective feelings of presence, and the second, an objective feeling of presence. However, he does not present a means by which these themes can be measured and how they affect the axis in his model.

2.2.2 Steurs telepresence

Jonathan Steur (1992) provides a different take on the meaning of telepresence which differs in some important features compared to Sheridan, It does not confine the definition of telepresence to teleoperator systems. Steur states that telepresence '*is the experience of presence in an environment by means of a communication medium*' (Steur 1992). This definition determines that if the user feels a degree of presence in an environment through a telecommunication medium, then they are telepresent there. The consequences of this are that the topic of telepresence is not restricted to teleoperator environments, but it also includes computer generated virtual environments within the definition of telepresence. To unify both teleoperator and virtual reality systems under one theory, Steur identifies two primary technological characteristics, *vividness* and *interactivity* which are shown in figure 4. These refer to the power of technology to represent a telepresence environment, and not the internal mental representation of the individual.

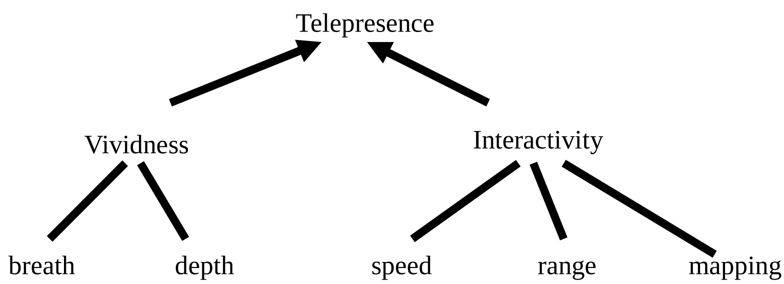


Fig. 4. Steurs (1992) variables influencing telepresence.

Vividness refers to the capability of a technology to produce, and present a sensorial rich mediated environment. Two important variables for vividness are *breath* and *depth*. Breath refers to the number of sensory dimensions simultaneously presented, e.g., touch (haptic), sound, visual imaging and taste/smell. Depth describes the quality of the sensory information received. An image with a higher resolution, and hence greater depth, is considered to be better than a low resolution image. Likewise, such is the difference between high quality 3D sound systems and normal 2D sound systems.

Interactivity is defined as 'the extent to which users can participate in modifying the form and content of a mediated environment in real time'. Steur identifies three factors important to interactivity; speed, range and mapping. Speed describes the real time rate at which information is mediated within the environment. It is based on real-time interaction where interactions by the user in the environment are updated quickly and efficiently to maintain the actual representation of the system with the user's mental representation. Range of interactivity is determined by the number of attributes of the mediated environment that can be manipulated and by the amount of variation possible. In other words, its the degree of change that can occur in the mediated environment. Mapping refers to the way in which human actions, such as moving, turning ones head, affect the mediated environment. These actions are captured by control devices, e.g., joysticks, which are then mapped or connected to actions in the mediated environment.

Steur rates different media systems by constructing a two axis chart for interactivity and vividness shown in figure 5. At the low end of this chart are things like books, at the high end, concepts such as star trek 'holodecks' and MIT's Kidsroom (Bobick *et al.* 1999). In the middle are media such as CD's and VCR.

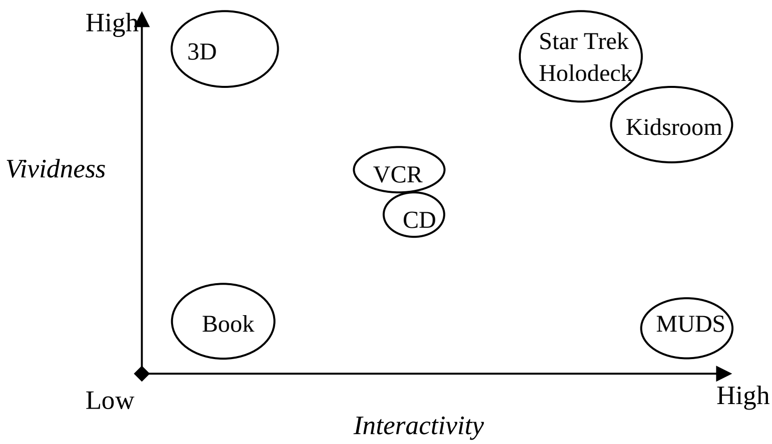


Fig. 5. Various media technologies classified by Vividness and Interactivity

2.2.3 Zeltzers AIP

David Zeltzer proposed a taxonomy for graphical simulation systems based on three components, *autonomy*, *interaction* and *presence* (Zeltzer 1992). While Zeltzer focused more on the effect of presence in virtual environments, the taxonomy is general enough so as to also apply to telepresence systems (Steur 1992). For Zeltzer, autonomy is a '*qualitative measure of the ability of a computational model to act and react to simulated events and stimuli*'. Autonomy describes the model of the space in virtual environments.

Interaction means the degree of access to model parameters at runtime with immediate response. This can vary from a state were there is no interactive response to a state of rich, immediately responsive interactions. The difficulty in interactive design, according to Zeltzer, is to organise the interaction environment so that the user is not overloaded with potential interactions that would limit their ability to perform tasks.

Presence is a measure of the number and fidelity of available sensory input and output channels. In VE's, this presents a problem as there are limits to the degree of photorealism and processing power to build a system that is entirely convincing. Therefore, Zeltzer distinguishes between the degree to which models can be accessed at runtime (interaction), and the means for accessing those parameters. Hence, the focus on achieving presence depends upon sensory information received by the user, such as visual, auditory, haptic and control, through devices. Presence provides a '*measure of the degree to which input and output channels of the machine and human participant(s) are matched*'.

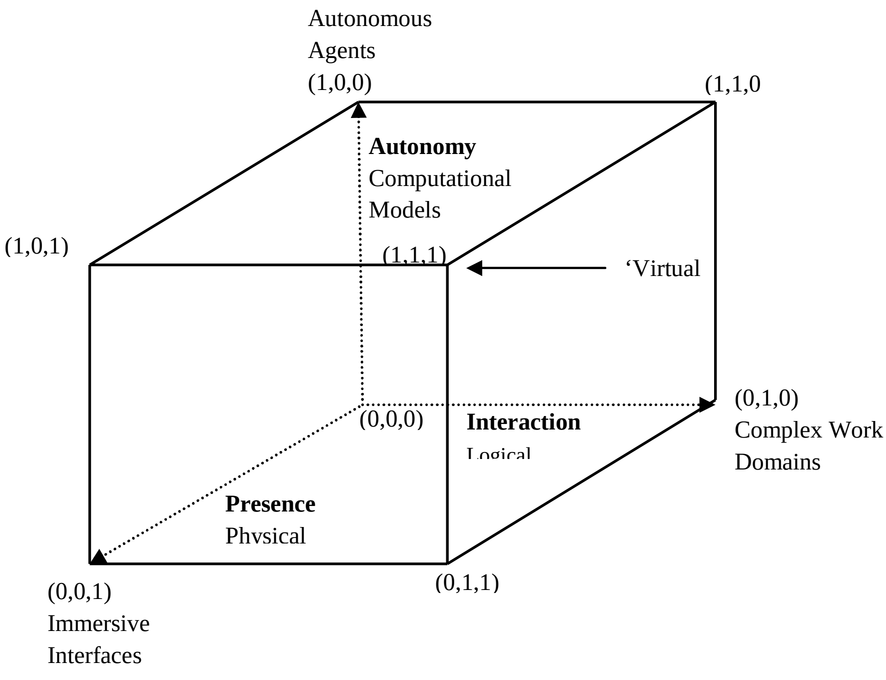


Fig. 6. Zeltzer's (1992) AIP - cube

Zeltzer organised these three components into axis ranging from values 0 to 1, to form the AIP cube shown in figure 6. The value (0,0,0) represents the point where the user has no interactive ability, presence, and there are no autonomous agents. The point (1,1,1) represents the goal, full virtual reality with the simulated model reacting continuously to stimuli, responding to the actions of human participants and offering an environment that is indistinguishable from the real world. Different graphical simulation systems exist somewhere between these two points. Zeltzer concludes that the problem for interaction is to understand how to abstract complex control spaces. For presence, it is to improve our understanding of human perception and design devices accordingly, and for autonomy, work is continuing on developing stronger models e.g., motion models.

2.2.4 Robinetts Theory

Warren Robinett set out to classify systems that incorporate a head mounted display, such as tele-operation and virtual reality (Robinett 1992a, 1992b). Robinett proposes that the main reason for needing this classification is that we have both direct action with the real world and also a technologically-mediated experience. Robinett sets out to classify eight features based on this technologically-mediated experience. These are *causality*, *model source*, *time*, *space*, *superposition*, *display type*, *sensor type*, *action measurement* and *actuator* type.

- Causality makes a distinction about what type of synthetic experience is received, whether the actions affect a virtual world, a real world (teleoperation) or are recorded (film). Model source describes the mathematical model stored in a database that acts as the basis for building a virtual model. That model source can be scanned in with real world data or crafted piece by piece by a designer.
- Space and time refers to the scanning and display of the environment and the time dimension in which it occurs. In teleoperation, the scan space is displaced from the display space, and, in some cases, the display space differs in scale, e.g., such as in microsurgery. Additionally, the time dimension covers the period in which the scanned space is shown in the display space. It can be instantaneous, delayed or, in some cases, scaled differently to capture important information events, e.g., slow motion replay of a goal during a soccer game.
- Superposition allows a merging of a virtual world with a real world. This form of superposition can be applied by superimposing objects on the subject’s view using HMD’s with half silvered mirrors. It also applies to other areas such as superimposing sound and tactile responses.
- Display and sensor type cover devices that support every phenomenon that human beings sensory organs can detect. Each display type has a list of corresponding sensors that provide information for it. Display devices include visual displays, tactile devices, auditory, smell and taste.
- Actions and actuators describe how actions of the users are enforced in the remote environment by various actuator devices controlled by the user. Actuators in Robinetts model are confined only to tele-operation classed systems, while action events are also applied to VE systems.

2.2.5 Schloerbs Telepresence Model

Schloerb defined telepresence as occurring when '*the person perceives that he or she is physically present in a remote environment*' (Schloerb 1995, 64). He attempted to distinguish between objective and subjective telepresence. The ability to modify the remote environment determined whether objective telepresence occurs at all and the degree of telepresence is decided by the ability to perform a remote task. Schloerb defines subjective telepresence as the subjects impression as to whether they feel that they exist physically in the remote environment. This can be accomplished by the subject identifying so strongly with the remote task that he believes himself to be actually operating within the remote environment.

Schloerb therefore says that one is either telepresent or not. He does, however, separate strongly between objective presence (teleoperator performance) and telepresence (subjective telepresence). For him, the most important thing is the ability to do work in the remote environment, i.e., objective presence, with some exceptions such as entertainment.

2.2.6 Milgram's continuum

Paul Milgram expressed an interest of somehow characterising the differences that exist within the definitions of telepresence and virtual reality. Unlike traditional teleoperator designs, Milgram mixed computer graphical environments with standard teleoperator systems. This lead Milgram to develop a taxonomy to explain how this mixing of reality could be formulated when using a video based head mounted display.

Milgram and Colquhoun (1999) described the problem in terms of the Reality-Virtuality continuum shown in figure 7. To the extreme left of this continuum is the real world, i.e., an environment that is in no way modified by any virtual synthetic object. Examples of the real world are the present local real world around us and telepresence systems that provide a live video feed from another real world location. To the extreme right is the Virtual World, a completely computer generated synthetic world. In between these two extremes we have some mix of both the real world and the virtual world. Technologies that are located on the right side of the continuum are basically computer synthetic worlds where real world images and objects are included. As an example, a 3D world with real-time voice or dynamic video feeds. The world is basically synthetic, but is augmented with real world components. This is referred to as Augmented Virtuality (AV). To the left of this continuum, we have what is essentially the real world which is augmented with synthetic objects. In this scenario, the world about us is real and synthetic images are overlaid onto the real world. This area is referred to as Augmented Reality.

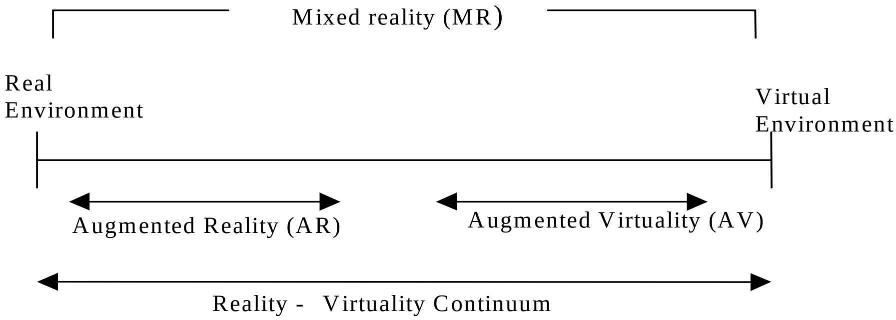


Fig. 7. The Reality –Virtuality Continuum

2.3 Ubiquitous Computing

Ubiquitous computing describes the situation where computing is everywhere and operates largely in the background of everyday activity, and, to all intent purposes, becomes invisible to the user (Weiser 1991). It takes many small computers located throughout the environment, mostly invisible to users, performing small tasks, but connected together by wireless links to enhance the overall computing power available to the user (Weiser 1993). Achieving this goal meant studying various aspects of the production of these small devices, such as smart appliances (Hansmann *et al.* 2003), and connecting them together using radio links, one such example being ad-hoc networks.

2.3.1 WSI Reference Model

The World Wireless Research Forum (WWRF) has worked to develop a reference model whereby ubiquitous computing platforms can be integrated into 4th generation mobile networks. The result of this has been to propose a reference model for mobile systems based on I-centric design (Popsecu-Zeletin 2001). I-centric design aims to develop services and applications that are human driven as opposed to technology driven. The technology adapts to the users needs, which requires the technology to support the user and modify its behaviour based upon the users context. It also adapts the communication system to meet the demands of each individual (Arbanowski *et al.* 2001).

The Wireless Strategic Initiative Reference model covers all aspects of the wireless world from business models and user issues down to radio interfaces. It describes the grand building blocks of the wireless world and how they interact at reference points. The combined definition of business models and reference points enable the early definition of roles and business relationships as well as assumptions on business topology and market value chains and value networks. (Arbanowski *et al.* 2002).

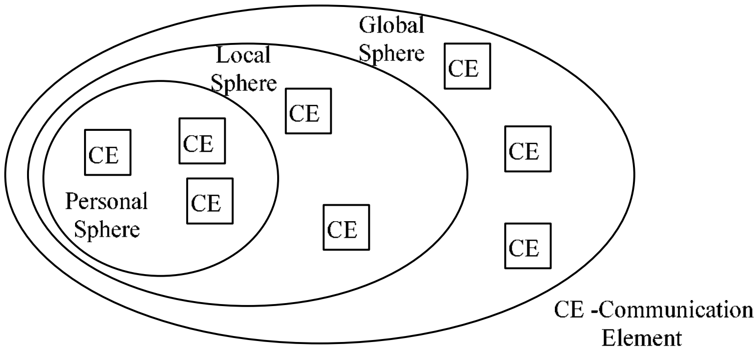


Fig. 8. Structure of the WSI Reference Model

The WSI reference model describes the Wireless World as a set of concentric spheres inhabited by the networked communication elements (CEs) shown in figure 8. These concentric spheres are a simplified version of the multisphere model set out in the WSI Book of visions 2000 (IST 2000). The main achievement is the definition of the 'Communication element', which acts as a communication entity in the different spheres. The different spheres reflect the vicinity of building blocks in respect to the user communicating in the wireless world. The wireless world is made of a large number of these CE elements, similar to Weiser's view. The Local sphere can be seen as a Body Area Network (BAN), the local sphere serves for local networking infrastructure and the global sphere is responsible for global connectivity. The communication elements are divided into four layers: the Cyberworld, Open Service Platform, Interconnectivity and Access, as shown in figure 9.

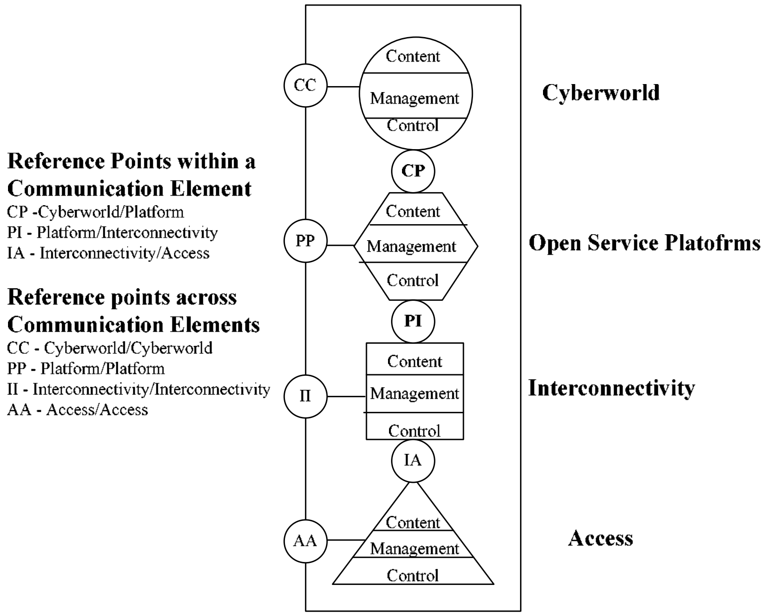


Fig. 9. Building blocks of Wireless World Communication elements

- The Cyberworld hosts all application specific functionality. It relies on a generic service infrastructure provided by the 'Open Service Platform' and exploits it to implement application service elements. It has the means to generically describe and explain their characteristics and demands to ensure that the underlying infrastructure is being used efficiently and to the user's satisfaction.
- The Open Service Platform is responsible for providing a flexible and generic service infrastructure to the cyberworld, in order to facilitate the creation of new services according to user's and telecom operator needs. The restrictions imposed on the creator of services have to be reduced to a minimum by providing reusable generic service elements.
- Interconnectivity can also be referred to as the networking part of the Wireless World reference model. The functions located there take care of logically linking communication elements from different spheres together and maintain and manage these links even when they are subject to a change of network topologies or access networks.
- Access implements all aspects of the physical connection(s) between different CEs. These may also be radio or other types of connections. Due to the hierarchical structure of the reference model, a connection in higher spheres could use multiple connections in underlying spheres, relying on services provided by the interconnectivity block.

2.4 Summary

This chapter described the background theory that is prevalent in the area of telepresence and ubiquitous computing and forms the basis for the central philosophy of presence in this work. It describes the philosophical basis for subjective telepresence, as advanced by Descartes, and the more objective telepresence school of thought, as supported by Heidegger and Gibson. Numerous theories of telepresence, their mechanics and classifications are described by Sheridan (1991), Steur (1992), Zeltzer(1992), Robinett(1992), Milgram and Colquhoun (1999) and Schloerb(1995). These have as a common framework, the importance of sensory information, the importance of being able to interact and work in remote environments and the need for feeling subjective presence to support this work. Finally, the WSI reference model was described as the basis for the implementation of the ubiquitous computing vision.

3 Virtual Reality Technologies

In this chapter, the term Virtual Reality is examined, and it is explained how it has become a misused term to describe the use of certain technologies such as Head Mounted Displays, position sensors, audio, computer generated synthetic objects and haptic devices. These technologies are used to realise a degree of presence as described in the previous chapter. This chapter describes a set of VR research fields that use common technologies, which are shown in figure 10. The chapter describes how different research fields have evolved under the common name of VR and how these research fields address common problems. Different research fields within Virtual Reality, such as Augmented Reality (AR), Collaborative and standalone Virtual Environments, Media Space, Telepresence and Teleoperation systems will be described as these provide solutions and constructs that are needed to address the research question. In particular, the constructs of existing telepresence systems are described. An overview on the operation of ubiquitous computing and associated security issues is also provided, as this is an important concept in supporting interactions for the virtual proxy.

3.1 Virtual Reality

In 1968, Dr. Ivan Sutherland developed the first computer graphic based Head Mounted Display (HMD) (Sutherland 1968). This device was comprised of two cathode ray tubes, mounted on a subject's head with the rest of the subject's vision of the world sealed off. A head position sensing system was also incorporated into the system. Both systems were connected to a computer where a computer generated 3D graphical world resided. By taking the head position sensor data, a view of this virtual world was rendered and displayed on the cathode ray tubes in the HMD. The effect was to create a virtual experience for the subject by placing them in an entirely virtual world. Other researchers strove to create systems to monitor and impersonate the other senses. 3D audio systems were used to imitate sound (Begault 1994), haptic devices, such as dataglove were used to simulate different sensations of touch (Zimmerman *et al.* 1987), and treadmills were used to simulate motion through large areas (Darken *et al.* 1997).

By 1989 a number of research teams were working on these ideas independently and in parallel with each other. These projects had different names such as Virtual Worlds, Virtual Environments, or virtual workstations. At this time, Jaron Lanier, coined the term *Virtual Reality* (Kalawsky 1993) to describe this collective research. The term then became a popular word within the media and among researchers to describe the virtual experience possible from these technologies.

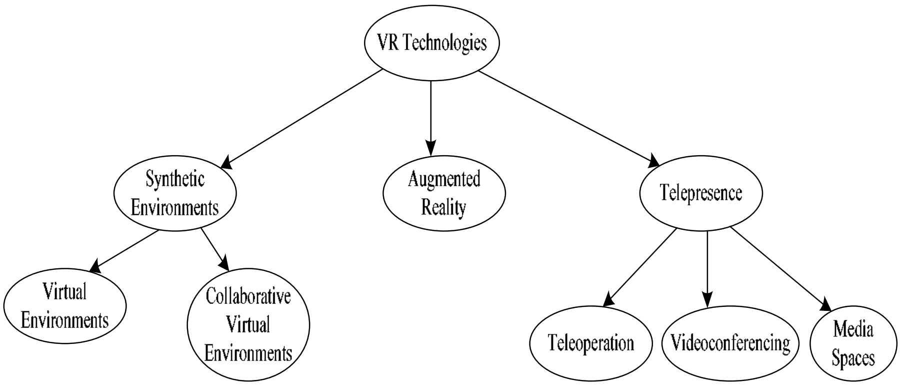


Fig. 10. Related research fields covered by the use of VR technologies such as Head Mounted Displays (HMDS), haptic devices, position sensors and audio devices

3.1.1 Defining Virtual Reality

Having now given a name to the research area, which has managed to survive for some time and gained acceptance, there came the resulting necessity to define exactly what the term means. The result was not consistent. Most popular definitions of virtual reality are based on the particular technological systems employed. Coates (1992) describe VR as electronic simulations of environments experienced via head mounted eye goggles and wired clothing enabling the end user to interact in realistic three dimensional situations.

Greenbaum (1992) considered it to be an alternative world filled with computer generated images that respond to human movements. These simulated environments are usually visited with the aid of an expensive data suit which features stereoscopic video goggles and fibre-optic data gloves. Krueger (1991) said that the term typically refers to three-dimensional realities implemented with stereo viewing goggles and reality gloves.

While each description varies, the one thing they have in common is that they all define VR based on the type of equipment used. This usually involved a computer simulated environment and various goggles and gloves systems to access those environments (Steur 1992).

Defining VR by the equipment it uses means that VR can cover a very wide range of technologies and these are summarised in figure 10. Synthetic environments described systems with the creation of entirely virtual worlds and the use of VR technologies to interface with them. There are two fields here, standalone and collaborative virtual environments (CVE), the difference between them resulting from the need of collaborative systems to support multiple users working together, while stand alone systems act as a single user experience. To complicate matters further, researchers started to focus on augmenting the real world with computer generated objects, a system that is generally referred to as Augmented Reality (AR) (Azuma 1997). AR uses many of the same pieces of hardware and software as systems with completely computer-generated worlds, so it could also fall within the category of VR. Additionally, teleoperator systems can also use some of the same hardware equipment to view remote real world locations, and this also had some effect on other telepresence technologies such as videoconferencing and media spaces. Researchers then started to focus on the effect of the system, the sense of being present or '*being there*' in the environment, whether that sense of presence was based on a totally computer generated world or based on the real world. The VR equipment became a means to an end, to provide a high level of presence or experience within the system, not just in completely synthetic environments, but also real environments. A consequence of the common use of technologies is that solutions for one field can be found in other fields, e.g., solutions to CVE can also be applied to collaborative augmented reality systems (Vlahakis *et al.* 2002). As a result, a good understanding of these fields should lay the basis upon which constructs can be developed to address the research question.

3.2 Synthetic Environments

Synthetic Environments are computer-generated worlds that attempt to replace everyday reality with one that is entirely virtual (Singhal & Zyda 1999). The user is then immersed in the environment using a HMD, tactile gloves and surround sound. There are two different types of synthetic environments of interest, stand alone Virtual Environments (VE) and Collaborative Virtual Environments (CVE). The difference is that a standalone VE is designed for the benefit of a single person and offers a platform for testing various immersion concepts and equipment (Slater & Wilbur 1997), as well as having a strong commercial relevance in the computer gaming industry. A CVE has the same structure as standalone VE, but includes support for multiple actors within the environment, a distributed communication network and support for work collaboration (Greenhalgh & Benford 1999).

3.2.1 Stand-alone Virtual Environments

A Virtual Environment is a computer simulated world consisting of mathematical and software representations of real (or imagined) agents, objects and processes, and a human-computer interface for displaying and interacting with these models (Zeltzer

1992). The user is then placed in the environment in a number of ways. The standard method is to place a HMD on the users head and attach it to a tracking device. For additional immersion, tactile feedback using gloves, motion feedback and auditory feedback can also be used. Figure 11 shows the different components of a VE.

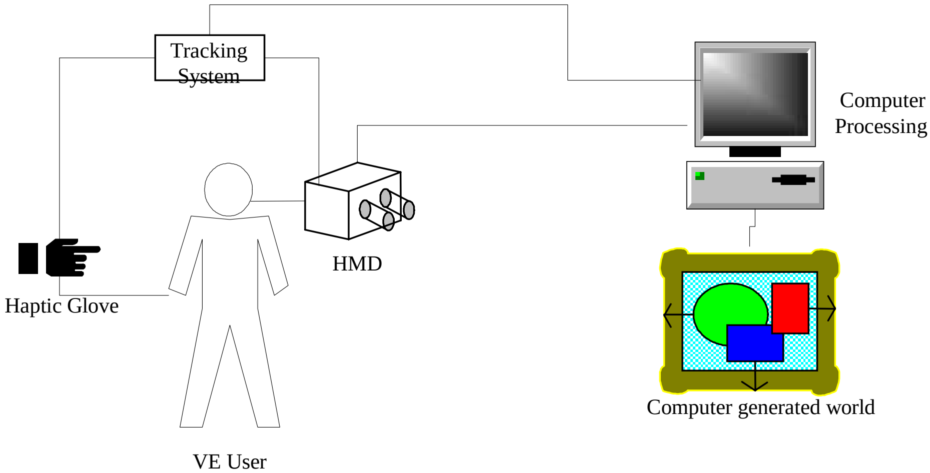


Fig. 11. Different VE components, the HMD, tactile devices, tracking sensor's and graphical world

Graphical world: The basic world in the VE is a graphical representation of a world that is built using 3D graphics software and hardware. The sophistication of these 3D worlds is heavily dependent upon the processing power of the software and hardware. The most popular means to create 3D-world content is to define a small polygon with points in the Cartesian plane. These small polygons are then combined to form more sophisticated objects. Visual colouring or images are then mapped onto the surface of each polygon in a process called *rendering*. To effect animation, transforms are applied to the polygons, altering their position relative to each other to give the impression of motion. The complexity of the model, as in the number of polygons, and the complexity of the rendering images determine how much processing power is needed for the creation of a smooth world. Special 3D graphical video cards are available which are dedicated to the processing of large volumes of polygons and rendering. In general, the more visual complexity offered in a VE, the greater the lag will be induced (Reddy 1997) as the system struggles to keep up with the users' change of position. This lag, which is also made worse by delays in position system and errors, can cause effects of motion sickness, or cybersickness (Regan 1995). Other contributing factors to cybersickness include vection and HMD Field of View (FOV) (Stanney *et al.* 1998).

Head Mounted Display: The HMD provides the visual sensory information of the environment to the subject, and hence is one of the most important means in which immersion in the VE is achieved. For VE's, the HMD is usually a closed system that excludes any vision of the outside world, other than the image shown on the virtual display. The HMD usually consists of an image source, mirror, an eyepiece, and for see through systems, a combiner (figure 12). The image source is usually a display technology such as a cathode ray tube or a Liquid Crystal Display (LCD). The mirror in the system acts to bend the light source, to direct the light source to the subject's eyes so as to have a more compact piece of equipment. The eyepiece acts as a magnifier so that the image appears larger. The Combiner is a half silvered mirror which reflects the image source to the user as well as leaving a portion of the outside world scene into the eye. Some of the most important restraints on the HMD are the display resolutions, Field of View (FOV), binocular overlap, and various technical distortions introduced in the system. (Kalawsky 1993)

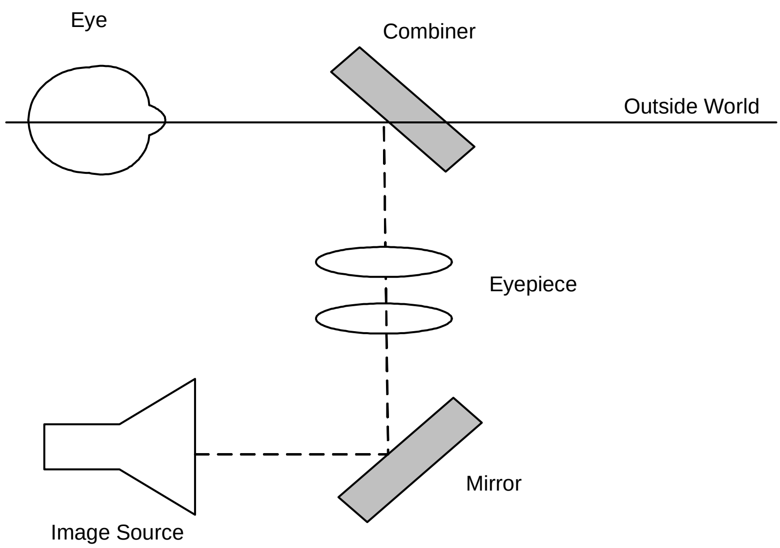


Fig. 12. Components of a HMD. The half-mirror can be replaced with a full mirror for non-see through HMD

Position Sensors: Acquiring the position and orientation of a person's body elements is important, e.g., the position of the head is needed to calculate the correct perspective image in the 3D environment. Other position information of interest may be the hand position, for interaction, and walking position, for navigation. There are four basic approaches to sensing position, electro-mechanical, electromagnetic, acoustic and optical technologies. One of the most common of these is the Polhemus tracker, which uses an A.C. electromagnetic system. The Polhemus tracker has low frequency magnetic fields using a three-axis dipole source with three magnetic sensors and works by tracking the changes in this magnetic field. For this to work, the system must first be calibrated to determine the starting reference position. Such systems have very low range and need constant calibration. (Kalawsky 1993)

Haptic device: Haptic devices allow the human operator to select and operate computer systems with a virtual hand controller that provides natural feedback via tactile stimulation and force feedback. The user's hand position and orientation is determined and is input into the virtual world generator so that it can display an image of the hand in the correct position on the Head Mounted Display (HMD). This allows the user to grasp and manipulate items in the 3D environment. (Kalawsky 1993)

3.2.2 Collaborative Virtual Environments

Collaborative Virtual Environments (CVE) systems combine multiple distributed VE's together to form one world where multiple people can work together. An initial early driving force was the need for collaborative military simulators, that had to combine and synchronise multiple geographically distributed VE's together to form a single environment (Stytz 1996). The CVE contains all of the characteristics of a VE, but also includes support for distributed and collaborative actions. The key concept for CVE's is that the occupants are represented to one another in graphical form and can interact with each other and with various representations of data and computer programs (Benford *et al.* 1996). A number of systems exist which support this approach. Some of the most widely available CVE's are DIVE (Carlsson & Hagsand 1993), NPSNET (Macedonia *et al.* 1995), MASSIVE (Greenhalgh & Benford 1999), (Greenhalgh & Benford 1995), BrickNet (Singh *et al.* 1995) and Mixed Reality (MR) Toolkit (Wang *et al.* 1995). Three key issues are the object distribution model, network communication and network latency.

3.2.2.1 Object Distribution Model

In the distributed VE, each element in the network has its own geometrical objects which form a subset of the total VE. The objects stored at each node of the CVE are said to reside on the local node. Each node must inform all of the other nodes within the network of their local objects from which each node can then generate the total environment. There are 3 methods in which this is performed (Broll 1995).

Active replication: On the creation of an object at a node in the network, a copy of that object is made and distributed amongst all of the other network nodes. Once distributed, changes to the object must be distributed amongst all of the sites and the system must guarantee that simultaneous changes are resolved to provide one definite state within the system

Replication on demand: With replication on demand, objects and changes to those objects are only distributed when a process expresses some interest in that object. When the object changes, only those processes that had expressed an interest in the object are informed. There are two approaches to managing the copies, either the objects do it themselves, or a special manager for all the objects of a node is established. The manager is then responsible for all object copies and updating.

Migration: Migrations do not distribute copies, the objects themselves are distributed, but not duplicated. The distribution is done according to the load at each node or their frequency of access. Occasionally, some duplication is done in order to reduce bottlenecks for commonly accessed objects, in which case, either of the two previous methods may be used.

3.2.2.2 Network communication

In principle, any communication medium may be used to connect distributed VE, but in practice, all of the systems developed use the Internet Protocol (IP) for network layer distribution, with either Transmission Control Protocol (TCP) or User Datagram protocol (UDP) used for the transport layering. TCP/IP suits the bursty nature of the CVE data transfers. (Singhal & Zyda 1999). The communication strategy can vary depending upon the system and context. Possible approaches are

Peer-to-peer: Communication is established on a peer to peer basis. A change in one object in one node is communicated to each other relevant node on an individual basis. An example is a 5-node system, with each one sharing a common object, so that changes will have 4 individual messages sent to each of the other four nodes.

Multicast: Multicast subdivides the system into a number of subsets, which listen to a channel for any messages. With a multicast system, one message is sent when an object changes state, the other nodes listen in to this message, determine if it is meant for them and act accordingly.

Broadcast: When an object changes state, a single message is sent to all nodes irrespective of whether the nodes require that information.

These three techniques both have their advantages and disadvantages. Peer-to-peer communication works satisfactorily for low-end systems that do not have a large set of objects. For higher volume systems, the messages sent increase drastically, requiring high bandwidth to negate network latency. Broadcast is bandwidth cheap, but sends messages to every node whether it wants it or not. Multicast is the most efficient for systems with many users e.g., 100 plus. As this number of people cannot exist within the same presence or awareness of the objects, there is no need to communicate with all of them, but rather smaller groups. These smaller groups then belong to the multicast group, only receiving information relevant to them.

3.2.2.3 Network Latency

As with all communication systems, information transfer is not instantaneous. There is normally a small delay between the transmission and reception of packets throughout the system. This has a varied effect upon the CVE. When an object in a local node changes, the copy of that object in other nodes will also need to change. As information transfer is not instantaneous, there is a small lag between these two actions. Depending upon the frequency of state refreshing, the lag can become quite significant to the participants in the CVE. For proper operation in a CVE, the participants must have trust that the position they see an object in, is the actual position of the object and not the delayed position. Even a minor delay can have a significant effect on interaction, especially in games such as Doom (Singhal & Zyda 1999). The nature of this lag is then a combination of a number of factors: the object refresh rate, the bandwidth, rendering, communication protocol used and frame rate. Network latency should be kept as low as possible, which is one reason why network VE tends to have lower photorealistic models than standalone games (Singhal & Zyda 1999).

One means of reducing the effect of network latency is to apply *dead reckoning*. Dead reckoning works by estimating the future position of an object based on the object's current position, speed and orientation. This result is checked against the actual position of the object, which is broadcast at intervals. Dead reckoning works by predicting the future position of an object based on the current knowledge. A number of dead reckoning algorithms exist with different benefits and results (Capin *et al.* 1999, Aggarwal *et al.* 2004).

3.2.3 Interaction in CVE

The purpose of CVE's is to create an environment that has some useful applications. The main use of such environments is to facilitate communication, co-operation and collaboration between different participants of the system. This differs from the interaction in VE's, where interaction is a singular experience between the user and environment on a local machine. The purpose for interaction in CVE's may be for training (Stansfield *et al.* 1995), gaming or work collaboration. CVE's are heavily dependent upon a spatial form of awareness. What this means is that the relative position of the actors within the environment determines their ability to communicate. This is done to mimic real world interaction as much as possible. The principles for this spatial interaction were established by Greenhalgh and Benford (1995) and concentrates on the idea of the aura, focus and nimbus, as shown in figure 13.

Aura: An aura may be defined in a VE as a 'region in which a person's presence may be perceived. Thus to be perceived, a persons aura must extend to the perceiver' (Fahlen et al. 1993). The aura is used as a means of scaling as it limits the number of objects where interaction can be considered. When 2 auras intersect, the users virtual avatars are fully aware of each other and may perform more extensive interactions.

Focus: Focus describes the general awareness a virtual object has of other objects within the environment (Greenhalgh & Benford 1995). As an example, if one were in the front row of a lecture theatre looking at the lecturer, then the lecturer is within your focus as you are aware of him, whereas people sitting behind you are outside your level of awareness, hence are not in your focus.

Nimbus: Nimbus is a subspace in which an object makes some aspect of itself available to others (Greenhalgh & Benford 1995). This could be its presence, identity, and activity. The nimbus allows an object to try to influence other objects.

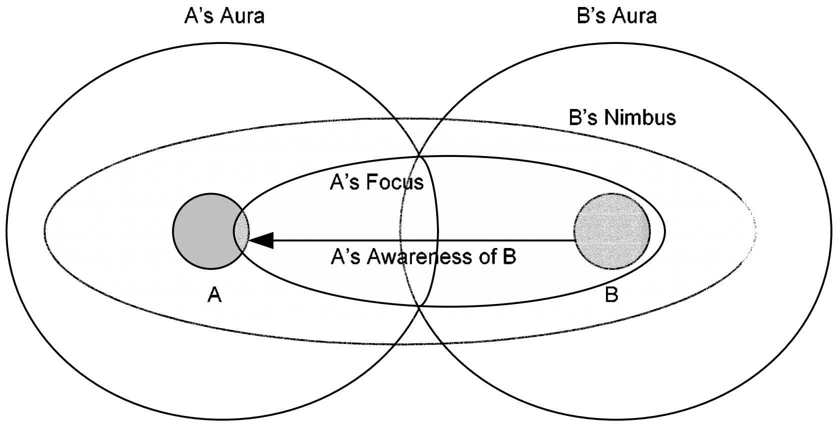


Fig. 13. Interaction begins when A and B's auras collide. A has a high awareness of B as it is in is in its focus and within B's nimbus.

These three levels of awareness act as a pretext to interaction and have some powerful applications. The boundaries can be modified to fit the context or potential applications. For example, if a person was to enter your aura, he could automatically send you a message to introduce himself. The standard means in which the aura is implemented is to use standard collision detection software. When a collision is determined between 2 auras, a peer-to-peer communication is established and the two objects negotiate the level of mutual awareness using values for both their nimbus and focus.

3.2.4 Usability issues in CVE's

The concept of immersion in CVE's is somewhat different from that of VE's. Not only is there a need to feel some presence in an environment, but the primary purpose is to support collaboration between people and hence awareness. From this point of view, the relationship between the participants and their avatars is the primary interest. Bowers and his associates (Bowers *et al.* 1996) noticed that users had difficulty getting a common point of view in a strange environment and this reflects the effectiveness of a CVE (Hindmarsh *et al.* 1998). Fitzpatrick and his colleagues (Fitzpatrick *et al.* 1996), say that it is more useful to talk about the ideas of locales within the CVE as a means of organising work tasks. This concept is expanded upon to the idea of the aura (Fahlen *et al.* 1993). Greenhalgh and Benford (1995) makes the aura the chief means by which interaction between avatars is performed. Interaction only occurs when the aura of two avatars overlaps each other. This approach is an attempt to mimic what happens in the real world, where people normally only interact when they are in such a close physical proximity with each other that they can see and talk to each other. Therefore the aura represents a similar concept of the space around the avatar in which interaction between people can occur.

As with most technologies, the success of CVE does revolve around the perceived need for people to use it and the effectiveness with which they can participate in the system. For CVE’s, the ability of people to interact and communicate with other users is the primary purpose of the system, and hence, that is the chief means for determining the usefulness of the system.

With a CVE, the intent is often to mimic real world situations in a virtual environment, replacing the real actor with a virtual representation. Hindmarsh and his colleagues (Hindmarsh *et al.* 1998) say that the limited field of view of the HMD can cause a sense of fragmentation within the environment. Such field of views can vary from 40 –70 degrees of peripheral vision whereas normal humans have a peripheral view of 150 degrees. That is, the amount of information that can be viewed in a CVE is less than what could be viewed in real life. The loss of some of this information in a CVE makes interacting in CVE’s more difficult. Participants have to use their voice to specify actions so that others can understand their intentions. Greenhalgh and Benford (1995) suggest using different camera angles to compensate for this, using behind the shoulder camera angles, birds eye, trading off immersion for improved co-operation.

Another problem relates to understanding the other participants’ orientation and perspective. If, for example, another virtual user says’ look at this object, and gestures towards it, the other participants must first search for the relevant avatar, then see which way he is pointing to and then look for the object, often needing additional aural descriptions to discern the object.

This is further complicated by network latency. Hindmarsh and his colleagues (Hindmarsh *et al.* 1998), describe a situation where one user is pointing to a fireplace in a CVE, thinking that another user is looking at it. However the other user was in the process of turning and was actually facing a different direction, the updated information had not yet reached and updated the first user’s visual display. Hence, neither user had an accurate understanding of the other’s perspective, which caused some confusion.

CVE's also seemed to suffer when communication was taking place. The lack of any visual feedback while people were talking together caused communication breakdown, as participants did not know if the others were really listening. This is particularly an avatar problem, as it does not reflect the normal visual cues that a normal person would give in a real life communication. A possible solution is to texture map real-time video onto virtual embodiments as in the ‘talking heads’ (Olives *et al.* 1999).

The other problem with CVEs is that the degree to which the controller of the virtual avatar is actually present in the environment is never sure. If the avatar is not active, it is possible that they have engaged in external activities, just leaving their avatar hang in the virtual world. This can be difficult for other users, as they try and determine if one unresponsive actor is actually present, and not gone for tea, bathroom or holding parallel discussions.

3.3 Augmented Reality

In contrast to VE's, where users are immersed in a synthetic environment, Augmented Reality allows the user to see the real world, but with synthetic objects superimposed on the real world (Azuma 1997). Azuma (1997) states that an AR system is any system that combines

- The real with the virtual.
- Is interactive in real time.
- Is registered in 3 dimensions.

Two different research teams lead by Kim and Davey (Kim *et al.* 1997, Davey *et al.* 1994), say that AR is any case where an otherwise real environment is augmented by means of virtual (computer graphic) objects. Milgram and Colquhoun (1999) suggest that the practical question is whether reality is being augmented, or whether virtuality is being augmented, and promoted a RV continuum to differentiate between the augmenting of a real world and that of augmenting a virtual world with real images.

Augmented Reality applications can be grouped under two broad categories, environment based and collaborative based systems. These environment based systems have undergone the longest research, as early systems focused heavily on the individual's interaction with the environment (Poupyrev *et al.* 2002), whereas when the technology developed further, more and more collaborative interaction between people started to become important. Environment systems focused more on the relationship between the user and the environment, including ARQuake systems that allowed people to play quake in outdoor settings (Piekasski & Thomas 2002, Thomas *et al.* 2002). Other activities focused on maintenance, such as that developed by Bertelsen and Nielsen (2000) for assisting workers in wastewater management, and science exhibition displays in museums (Woods *et al.* 2004). These systems also started to allow people to add virtual objects, such as voice notes or photographs to the physical environment (Rekimoto *et al.* 1998).

Collaborative applications included a variety of techniques, such as tangible interfaces, information cooperation and synchronous editing of data. Tangible interfaces mapped virtual objects to real objects, and people could cooperate by moving these objects around (Slay *et al.* 2002). Examples include TILES (Poupyrev *et al.* 2001, Tan *et al.* 2001), which used paper markers to represent data and actions. Gausemeier lead research on using tangible interfaces to help in the design of flexible manufacturing floor plans (Gausemeier *et al.* 2002). Another example used an AR videoconferencing application by superimposing a 3D video avatar on a fiducial marker, although the avatar is only a few inches high (Prince *et al.* 2002). Cooperative AR systems started to look at information exchange, using various notepads to enter and control data between users and the environment. Examples include the Archeoguide experiments (Vlahakis *et al.* 2002), chess game (Reitmayr & Schmalstieg 2001), and MAJIC mobile AR collaborative system (Nigay *et al.* 2002). Collaborative applications supporting synchronous updating of objects by AR users in the same physical location (Renevier & Nigay 2001, Schmalstieg & Hesina 2002).

3.3.1 AR architecture

The implementation of an augmented reality system uses the same equipment as used by VE, but with some differences. These main differences are

Scene generation: Rendering is not a major problem with AR in comparison to VE’s. VE’s need to render complicated images, in order to replace the real world with a virtual world. Therefore AR systems require less graphical processing power and complexity compared to VE’s.

Display device: The display device used in AR has less stringent graphical processing requirements than VE’s. The resolution of an optical see through HMD might be lower than that which could be tolerated by a VE HMD, because the optical see through HMD does not reduce the resolution of the real environment. Optical see through HMD can suffice with a small field of view, as the user can still see the real world with their peripheral view.

Video see through HMD works by combining a closed view HMD with one or two head mounted cameras. The video provides the users a view of the real world. Video from these images are combined with the graphic images scene generator, blending real and virtual (Klinker *et al.* 1997).

Optical blending is simpler and cheaper than video blending, as only one stream of information is used to display the graphics required, and the real world is seen through the combiners. The added advantage is that the resolution is not set by the resolution of the display device, but rather by that of the real world, as well as providing normal peripheral views. Neither does the system have to concern itself with defining the correct eye offset for the generation of stereoscopic 3D images of the world.

Video also allows the option to better match the delay between the real and virtual views. With an optical system, the user can see the real world scene instantaneously, whereas there is a delay in updating the virtual image. This delay can cause problems to the user. Video systems can delay the real image to match the delay of the virtual rendering. Video systems also provide additional registration options using visual recognition of visual markers to control registration. (Rolland *et al.* 1994)

Tracking and sensing: While the two previous cases had lower requirements for AR than VE, this is not the case with tracking and sensing. The requirements for AR are much stricter than VE's, as objects must be matched with the real world. In VE’s, this can be controlled more easily as the world is virtual, and the scene generator has total control over the scene generation and object placement.

3.3.2 Registration

The principle issue with AR, and one that prevents their commercialisation, is the problem generally referred to as *registration*. Most AR applications require the accurate superimposition of a virtual object over a real object, otherwise the illusion that the two worlds coexist would be compromised. An example of where this is necessary is needle biopsy (State *et al.* 1996a). Also, registration disparities caused significant difficulties in picking up virtual objects, especially when they seemed within reach (Thomas *et al.* 2002). To successfully match the real and virtual world, the scene generator needs the ability to interpret the exact position of the HMD and it's orientation in the world. For this, accurate tracking and sensing is necessary.

There are two types of registration errors that exist in AR systems, static errors and dynamic errors. The four main sources of static error are

- *Optical distortion*: optical distortion exists in most cameras and lenses. Distortion is a function of the radial distance away from the optical axis. Near the centre of the view, images are undistorted, but the further one goes the worse is the phenomenon. The use of optics to magnify and focus graphical images can cause a distorted graphical image that is mapped onto the real environment, causing static registration errors. Static registration errors can be removed with some careful planning of additional optics (Edwards *et al.* 1993) or image processing to pre-distort the image (Watson & Hodges 1995).

- *Errors in tracking system*: these are errors generated by the tracking systems. They are difficult to deal with, as they require another 3D ruler, which is more accurate than the tracker that is being used.

- *Mechanical misalignment*: This is where errors in the physical realisation of the system are not as specified. This can happen by some parts of the mechanical system changing slightly, due to wear or physical accidents, or the HMD attachment becoming loose.

- *Incorrect viewing parameters*: The viewing parameters define how to convert the reported head or camera locations into viewing matrices used by the scene generator to draw the graphic images.

Dynamic errors occur because of system delays or lags. The end to end delay is defined as the time difference between the moment that the tracking system measures the position and orientation of the viewpoint to the moment when the generated images corresponding to that orientation appear in the displays. Typical reasons for this delay are delays in the tracking subsystem, communication delays, and scene generation rendering scan out delay from the frame buffers to the display. End to end delays of 100 ms are fairly common. These delays only cause errors when motion occurs. This is because as the user moves his head, what they see is delayed by a time equal to the system delay and is incorrect for the time that they see the resultant image. System delay is the largest single source of registration error in existing AR systems, outweighing all others combined (Holloway 1995). Techniques to reduce these errors are to reduce lag, reduce the apparent lag (Burbidge & Murray 1989), match temporal streams for video based systems or predict the future viewpoint and object locations (Emura & Tachi 1994).

Bajura and Neumann (1995), point out that registration based solely on the information from the tracking system is like building an 'open-loop' controller. The system has no feedback on how closely the real and virtual actually match. For this reason, the use of visual based tracking systems are highly popular. Visual based tracking systems use image processing to aid registration. A video image of the scene is generated and the image processing software attempts to identify fixed features in the environment.

This process is usually assisted by placing special markers in the environment (Mellor 1995). These markers are detected and their locations are used to make corrections to enforce proper registration. With this scheme, the result can be as accurate as 1 pixel. Vision based systems can be computationally problematic, so a combination of vision based tracking with magnetic tracking systems can also be used (State *et al.* 1996b). This system worked in real time with errors typically less than 1 pixel, but encountered problems due to the difference in lag between the visual and the video tracking system

3.3.3 Position, mobility

In VE, the participants can fly through the environment. In the real world, their movement is usually confined to a specific place and movement in the VE's is enacted using controls such as a mouse or keyboards. For AR users, movement is physical, one cannot fly through the environment, but must walk or run. The AR user must then carry the tracking and sensing equipment with them. Unfortunately, current tracking and sensing systems do not work very well when the user is highly mobile. Polhemus trackers have a range of a few meters, which is not suitable for people walking long distances. Iltanen research group used an optical beacon based system, but the range was restricted to just six meters (Iltanen *et al.* 1998).

This mobile nature proves highly problematic for the tracking system used in AR. Most tracking systems require markers placed in the environment, and these are read at various points to calculate the position. Starner and his colleagues used a system of visual code tags combined with a GPS tracker to help with positioning and registration for outdoors, and an infrared tracking system for indoors (Starner *et al.* 1997). Pyssysalo (1998) used a combination of GPS with gyroscopes and acceleration sensors, and Radio Frequency Identification cards (RFID tags) for indoor environments.

3.4 Telepresence Systems and their constructs

Chapter 2 describes the theories that make the foundation for telepresence systems. This section looks at the practical systems used in telepresence and summarises their main constructs. The principal goals for teleoperation systems were to allow humans to operate in hazardous and hostile environments (Minsky 1980) such as space (Weisbin & Lavery 1994) or nuclear facilities (Greaves 1994). There are two existing approaches to teleoperator systems, one is the traditional teleoperator system, relying on visual feedback, and the second is VR based teleoperator systems. Using Robinett (1992a) classifications, variations on the classification of these two types of systems are shown in table 1. The remaining part of this section describes the structure for these two approaches and identifies the constructs in each system.

Table 1. Robinetts classification of as applied to different telepresence

	VR based excavator	Tele-operator
Causality	Transmit	Transmit
Model Source	Reconstruct	Scan
Time	1-to-1	1-to-1
Space	Remote	Remote
Superposition	Isolated	Isolated
Display	Screen/graphics	HMD, force feedback
Sensor	Cameras and radar	Camera
Action measure	Joystick	Force feedback arm
Actuator	Excavator	Robot arm

3.4.1 Traditional Teleoperation

The typical structure for a teleoperation system as described by Sheridan (1991), Held and Durlach (1992) and Minsky (1980) is shown in figure 14. The physical proxy can contain a manipulator or/and a motor system to enable movement. The physical proxy can also include sensors, including a video camera, which help in the control of the system itself, and which provide feedback to the operator. A communication medium is used to transmit information to and from the physical proxy. The operator possesses devices that enable him to view the sensory information they receive over the medium, as well as an action measure that enables them to control the remote actuators. The system is a single user only system, i.e., only one person may use it any time.

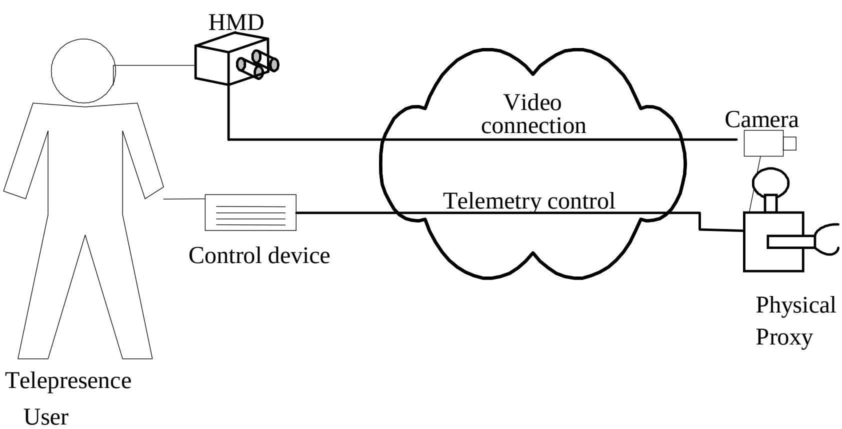


Fig. 14. The typical system configuration for a telepresence system, involving a remote controlled robot with an attached video camera

The basic constructs of the system are shown in figure 15, and the technology needed for each construct is shown in table 2. They can be divided into three sections, the operator side, the proxy side and the communication system. The system goal is for the proxy to provide sensory information to the operator and for the operator to provide commands based on that information to the proxy in order to complete a task.

The *sensor display* construct is the means by which sensory information recorded in the remote environment is presented to the teleoperator. This information can be presented in a variety of ways for different experiences, e.g., through the use of HMDs, force feedback, displayed on TV screens and 2D or 3D audio.

A *Control device* construct is that part of the system where the teleoperator inputs commands that the remote proxy must obey in order to carry out tasks. These control devices can come in many forms, depending upon the complexity of the remote proxy. A remote control car would be easier to control than a remote control aircraft. Typical control devices include joysticks, datagloves, 2D and 3D mouse and keyboards.

The *operator communication element* is the front end that is responsible for transmitting the data bits for the sensor presentation and control devices to the physical proxy. This communication element is always owned by the operator. The *communication medium* is responsible for the transmission of data over long distances, be it terrestrial networks or long range space radio signals. The *proxy CE* is a communication element, usually found on the physical proxy. For mobile physical proxies, this CE is usually connected to the operator by a radio link.

The *user's sensors* construct are those elements that take information about the environment that are necessary for the operator to gauge a feeling of presence there. These sensors include the scan types, such as video which the operator can use to look at what is going on. These sensors not only present the operator with the facts that he needs before he can begin work in the environment, but also provides valuable feedback on his work progress.

Actuators are the devices in the proxy that are directly controlled by the operator. It is the manipulation of these that enable the operator to work in the environment. Typical functions include moving the proxy, operating mechanical arms and modifying the camera position. Included in the actuators is any intelligent control process needed to interpret and respond to the control sensors.

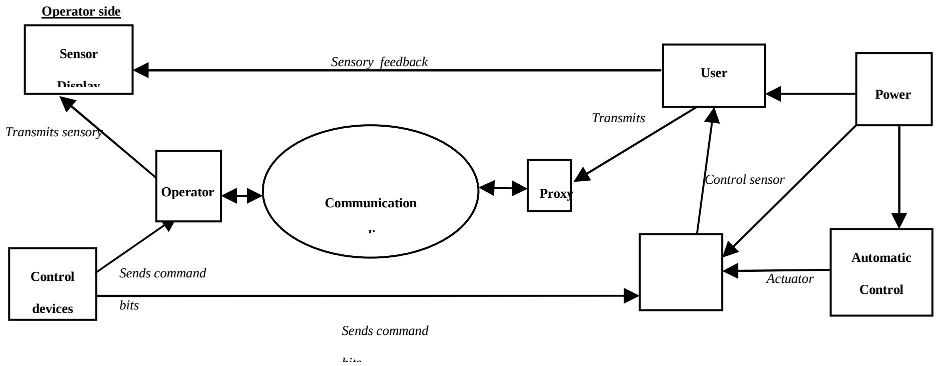


Fig. 15. A construct chart indicating the main elements of the traditional teleoperator system

Automatic control sensors are sensors that the proxy needs to control itself without the direct intervention of the remote operator. Examples of these sensors are those used to provide position feedback on mechanical arms, safety features to prevent collisions and damage to the proxy, and also to operate the proxy when it is not being controlled by the operator.

Power provides the proxy with the energy to perform its tasks. For many systems this is a simple battery, however, that battery must often be recharged, which is especially complicated for space missions.

3.4.2 Virtual Reality teleoperation systems

An advancement on the traditional teleoperation system uses virtual environments to assist the teleoperator in their task. With a VR based teleoperation system, the primary means for viewing the remote world is a VR based computer model that maps itself directly to the remote environment. Information about the remote environment is acquired by sensors and these mathematical values are fed to a computer which generates an accurate computer model. The proxy is also computer modelled and position sensors keep a track of the proxies' current position and orientation. The goal here is two fold, first, to improve the operators handling of the proxy, and second, support teleoperation in environments where video is insufficient. Two examples of this system are described here.

Table 2. Construct-technology chart

CONSTRUCT	Technology	CONSTRUCT	Technology
Sensor Display	HMD, Screen, force feedback	User Sensors	Video, tactile, Sound
Control devices	Joystick, dataglove, mouse	Actuators	Manipulator, Wheels, Tracks, Camera position
Operator CE	Fixed e.g., ISDN, TCP/IP	Actuator Control sensors	Feedback , Collision, Detection
Communication medium	Fixed and Wireless -video streaming -audio streaming -telemetry	Power	Battery, solar, mains, radioactive,
Proxy CE	Wireless, sometimes fixed		

In Paul Milgram's research group (Milgram & Ballantyne 1997, Ballantyne *et al.* 1997), a 3D model of an excavator is used to represent a real remote controlled excavator. The joint angles are sent to the local user, updating the 3D-model position to give an accurate reflection of the current state of the remote excavator. A 3D-laser range finder is used to record actual terrain elevations (Gagnon *et al.* 1997) and this is mapped onto a deformable 3D map of the terrain. A video image of the area is inserted or 'billboarded' into the scene to ensure that the 3D model of the environment matches the reality of the environment. The primary issues with this system is that

- The user lacks wider situational awareness of the environment.
- The camera position for the 3D perspective image and the video image perspective do not match, causing a degree of contortion.
- Sending correct control signals to the excavator as a result of the previous image distortions.

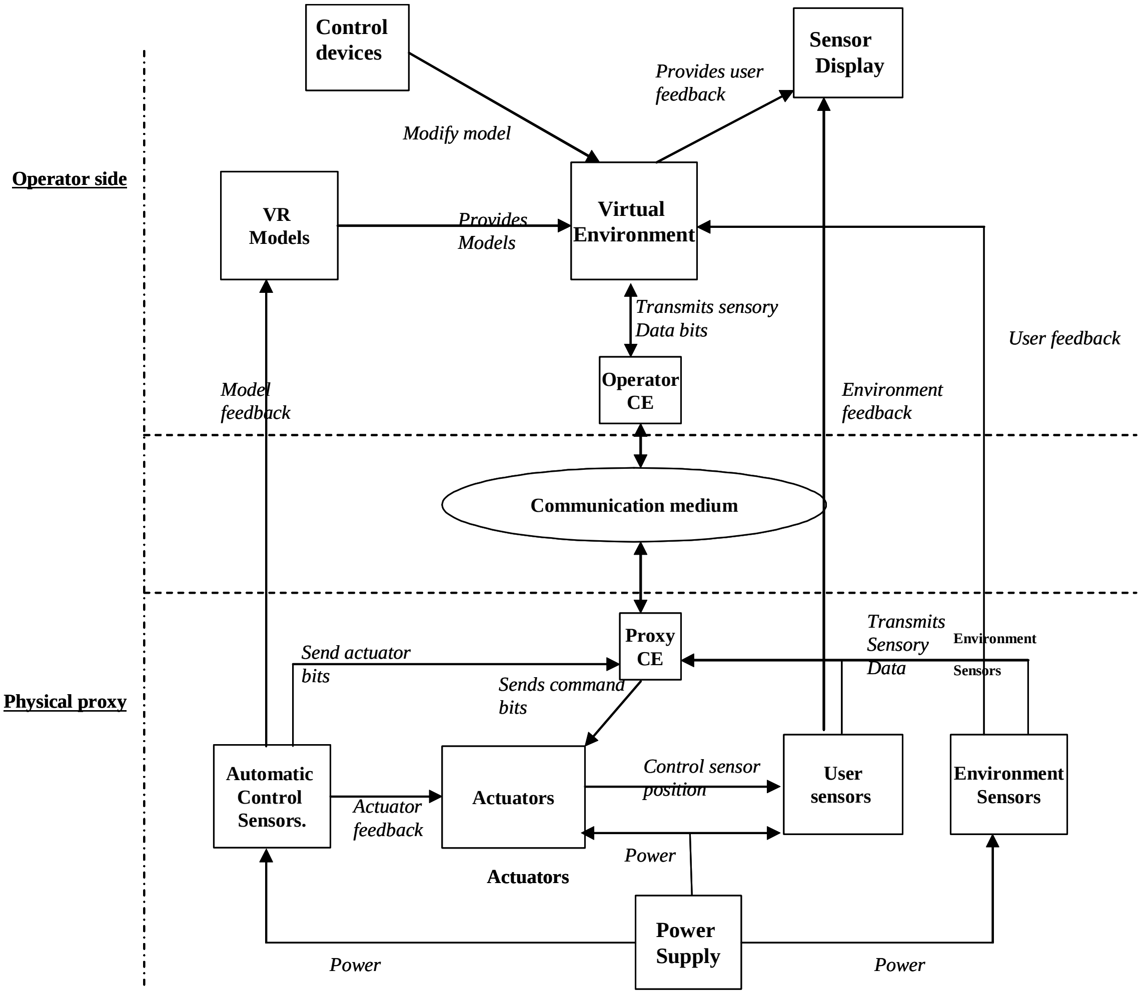


Fig. 16. A construct chart for VR based telepresence systems. Note the additional constructs to provide information to the VR model base and the two sources of sensory feedback to the operator, one directly from the remote environment and the other from the VR modelled environment.

The second type of system advanced by Lin and Kuo (1999) is for controlling a real unmanned underwater vehicle. The problem in underwater environments is that sediments in the water inhibit visibility, making it difficult for the operator to navigate the submersible. They created a 3D synthetic model of an offshore platform and the unmanned vehicle. They then used sonar to base the position of the vehicle with the known position of the platform. When the system detects that the vehicle is close enough to the platform, then the operator may switch to a video feed.

The general issue that can be concurred from these systems is that establishing the correct working perspective of the remote environment with the teleoperator is the greatest challenge for work that is beyond simple navigation.

Some of the constructs for VR and telepresence systems are the same as traditional telepresence systems, and a summary of these constructs is shown in figure 16 and table 3. While many of the constructs are the same, there are some additional constructs and the relationships between them also changes.

Table 3. Construct –technology chart for a VR based Telepresence system.

CONSTRUCT	Technology	CONSTRUCT	Technology
Sensor Display	HMD, Screen, force feedback	Proxy CE	Wireless, sometimes fixed
Control devices	Joystick, dataglove, mouse	User Sensors	Video, tactile, Sound
VR Models	Synthetic objects e.g., in Java3D, Opengl	Environment Sensors	Radar, sonar
Virtual Environment	3D computer generated graphical world	Actuators	Manipulator, Wheels, Tracks, Camera position
Operator CE	Fixed e.g., ISDN, TCP/IP	Actuator Control sensors	Position feedback , Collision Detection
Communication medium	Fixed and Wireless -video streaming -audio streaming -telemetry	Power	Battery, solar, mains, radioactive,

Sensor display receives information for the operator from two sources, the primary source being the computer generated 3D models, with the video images acting as a back up. The control device construct now works by directly modifying the computer model of the physical proxy and so does not logically speak directly to the physical proxy. Rather the commands are interpreted by the model which then determines the command bits sent to the physical proxy.

The VR model construct acts to store the different independent models used in the virtual environment. Prime among these are the models of the physical proxy. The information received from the actuator’s sensors are used to update this model’s orientation and ensure that it coincides with the picture seen by the operator. Other models can include environmental structures such as oil platforms.

The virtual environment operates the rendering of the VR models and also the rendering of the remote environment based on environmental sensor information. Information received from the environmental sensors is used to construct an accurate representation of the remote environment for viewing by the operator.

The automatic control sensors differ only in that not only are the sensors being used to help the physical proxy perform its task, but the sensor values are also used to update the VR model with the correct position and actions of the actuators. This means that there is a need for additional information flow across the communication medium compared to traditional teleoperation systems.

An additional sensor construct, environment sensors, is needed to capture information about the environment. As in the case of Lin and Kuo (1999), where sonar is used to position and orientate the submersible, this information is not used directly by the user, but by the virtual environment which presents it in a visual framework that the user can understand.

These types of systems require heavier sensor work than traditional teleoperation systems, and consequently, because they still use video, additional communication bandwidth.

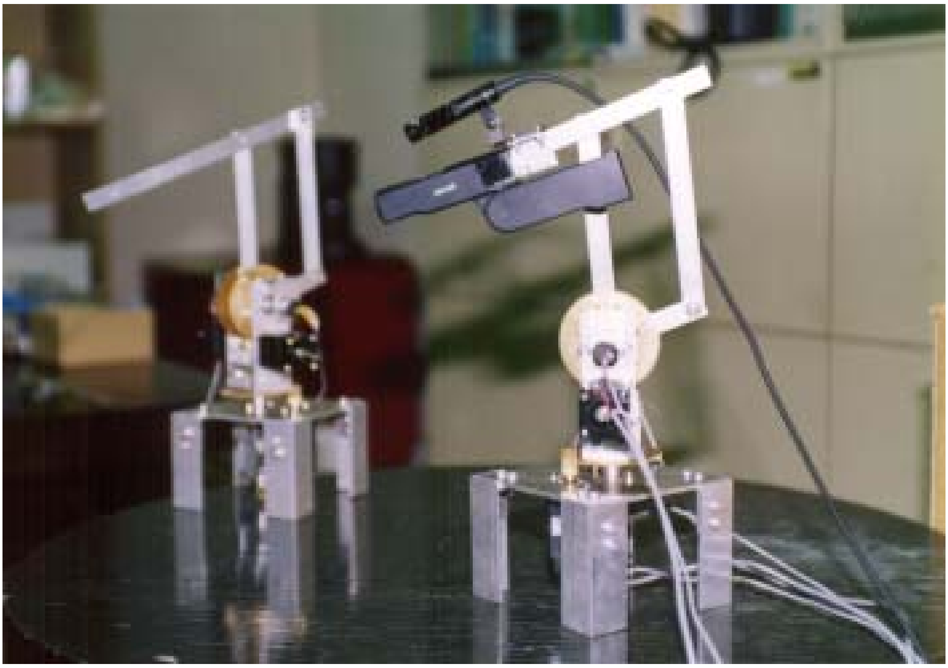


Fig. 17. The GestureCam Master-Slave system. The slave (right) has a laser pointer. The master (left) controls the movement of the slave.

3.4.3 Collaborative telepresence systems

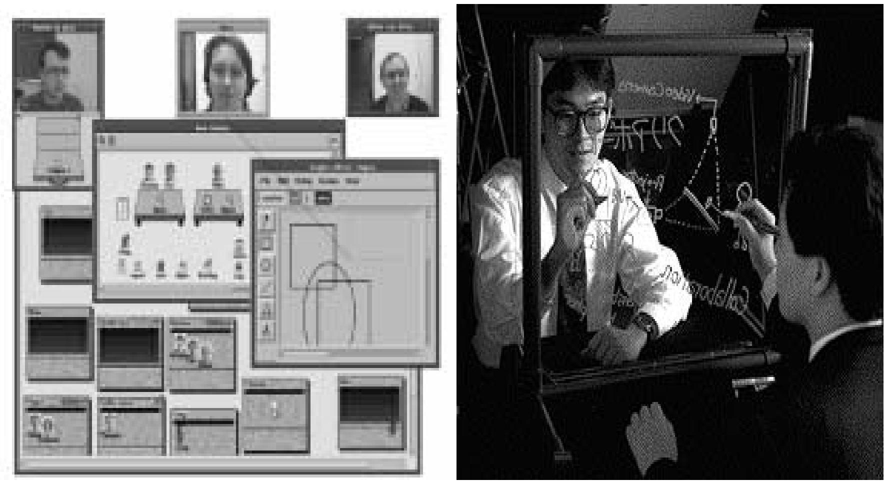
Initially, teleoperation systems focused on the exploration of hostile environments, but later on, these were also used in supporting collaboration tasks and tele-embodiment behaviour. Kuzuoka (1992) and Kuzuoka and his co-authors (1994) examined the use of teleoperation to support spatial workplace collaboration, i.e., defining an area where a remotely connected person can perform tasks in an environment, whereby collaborative tasks can be accomplished together. Normal teleoperation focuses on the robot being in an environment where people do not exist and where the transactions are between the robot, environment and the user. Kuzuoka's system, GestureCam, shown in figure 17, had a remote environment, a robot with a laser pointer and local and remote users also taking part. The issues then changed to developing variations of focal points, sharing focal points, confidence in the communication system and the ability to confirm viewing intentions and movements. These systems had a remote person help and instruct a user of a machine in the operation of that machine. It was very important that both users had knowledge of their *field of views* (FOV) in order to work smoothly.

Nardi's research team describes a surgical operation that had a video camera co-mounted on the optics of a microscope, which is then shown on a video screen for the rest of the operation team, watching students and experts (Nardi *et al.* 1993). The purpose of the system is to assist co-ordination of activities by the surgical team during the operation. What they found was that rather than having people working together, the video allowed them to act independently to accomplish the task.

3.4.4 Video conferencing systems

Video conferencing systems establish an audio-video link between two rooms or desktop systems (Edigo 1988). This is restrictive also on data exchange, which must be managed in parallel or with specialist hardware and software in the system. The cameras are placed in fixed positions for the duration of the meeting, but can have features such as zoom, tilt and pan movements. In some systems, the remote user can switch between different cameras to have different perspectives. In desktop systems, a number of collaborative workspaces can be used for data exchange and collaborative work.

Some of the problems associated with video conferencing systems are the high logistic and support costs on the users of the system, as well as the necessary technical staff responsible for the maintenance and operation of the system (Bergmann & Mudge 1994). Most video conferencing set-ups are not face-to face systems. Okada's research team say that video conferencing systems find it difficult to support multiple eye contacts between participants and the images are small, so important facial expressions and gestures are lost (Okada *et al.* 1994). They proposed the MAJIC video conferencing system that erected life size video images of the participants on a curved half-transparent mirror. The nature of the MAJIC system is that it requires a specialist room with a limited number (2-3) of participants. Other variations of this team is the Teleport system (Gibbs *et al.* 1999) that uses a special full wall display surface that merges real time video with synthetic objects on the wall. Desktop video conferencing systems rely on straight audio/video links between the users. To support the meetings and to provide some basic gaze awareness, systems such as Diva (Sohlenkamp & Chwelos 1994), shown in figure 18(a) and Forum (Isaacs *et al.* 1994, 1995) were designed.



(a) (b)

Fig. 18. (a) The Diva meeting room, (b) Clearboard (picture courtesy of NTT Human Interface Laboratories).

Video conferencing systems also need to support collaborative work between its users, which is difficult by just using video cameras. Specialist work environments have been created such as Clearboard (Ishii & Kobayashi 1992) and Liveboard (Elrod *et al.* 1992). Clearboard, shown in figure 18(b), uses a specially constructed whiteboard, using half-mirrors with a projection screen underneath. A camera situated above the whiteboard captures the user's drawings and the half mirror also reflects the user's facial details to the remote user. The video image is then sent across the network and projected onto the projection screen. The advantage of this system is that it maintains a video link and supports gaze awareness. The downside to the system is the loss of clarity from the poor quality of the video images, and the double hand problem, where the hand position is obtained from the camera and also by the reflection, and the inability of erasing their partners drawings. Liveboard is a similar system but does not provide personal video information, and is mainly used within a video conferencing system to support collaboration.

3.5 Media Spaces

Media Spaces could be regarded as a version of a desktop Telepresence system. Media space systems use audio/video links to try and create a shared work environment for collaboration, using both formal and informal communication. William Gaver's research team, define a media space as a *'computer-controlled network of audio-video equipment used to support collaboration'*(Gaver *et al.* 1992, 27).

For others, the advantages of such a system is that it can also support a continual and casual awareness of others in the system, as well as supporting informal communications

(Fish *et al.* 1992). Support for informal communication is one of the main motivations for media spaces. By this, it is meant that such communications are impromptu, short, quick and easy to initiate, informal and lightweight (Tang *et al.* 1994). Informal communication is a key element in supporting work in a collaborative work environment (Kraut *et al.* 1993). So, as well as forming a medium in which one on one audio-video connections can be made for communication, most anticipate media spaces as forming a community in which participants are aware of the others present in the system. An example is the Portholes system (Dourish & Bly 1992) which presented each user with a video image of 15 different distributed locations. The cruiser system (Fish *et al.* 1992) used an 'autocruise' feature whereby audiovisual connections were made to each of its 30 nodes for a brief period of time. This tries to model the appearance of walking down a hallway while glancing into other people's rooms, which was also the approach used in Montage (Tang *et al.* 1994).

The camera system used in these systems were normally attached to a desktop computer, but EuroPARC (Gaver *et al.* 1993) concluded that for collaborative tasks, one camera systems were not satisfactory as they do not add significantly to the process of collaboration. They built a Multiple Target Video (MTV) system that used a 4-camera system to provide a greater view of the remote environment. While this helped with task related work, the added complexity exacerbated problems in determining a joint frame of reference. Yamaashi and Buxton (1996) used a 2-camera system to increase remote awareness, one camera focusing on the user and a second wide Field Of View (FOV) camera to provide more information on the remote environment. To overcome the limitations of using so many cameras, EuroPARC proposed that a user's head movements should be used to control a remote controlled camera to obtain a better perspective of the other's site (Gaver *et al.* 1995). This also had the benefits of informing the user of the remote user's actual FOV, letting the user interact with the camera as if it were the person. The head tracking system was highly susceptible to ambient light changes, a similar problem as experienced with image based registration in AR systems.

Alternatives to audio/video media spaces are audio only systems. Such systems concentrate more on voice communication and ambient noise, such as typing background noise and music. In the Thunderwire system, high quality duplex audio systems with minimal lag were essential (Hindus *et al.* 1996). Nevertheless, despite the relevant success of the system, Sellen (1995) noted that only 10 % of subjects preferred audio only systems to an audio/video system. Some of the problems associated with audio only spaces occur when people have left the environment without the knowledge of others. The condition then arises where people start to initiate communication with the missing person, although they are no longer present. The Thunderwire system was also a public audio channel, which meant that there was a lack of private confidential channels, an issue that needs to be rectified.

Of great concern in media spaces is the need for privacy and confidentiality. The presence of cameras in each room, providing a continuous video feed of each individual's current status brought out anxieties in some users and opened the means for intrusive communications and potential abuse. A number of systems, such as RAVE, CRUISER and PORTHOLES, tried different techniques to manage this fear. These techniques involve social protocols in reciprocity, i.e., people are aware who is looking at them at all times. This still does not enable the users to preserve privacy or object to intrusive behaviours. For a subject to have a level of control over the system would as Gaver state, *'one way to assure that media spaces will not add new threats to privacy would be simply to remove all audio and video', 'but this would clearly do away with any and all services these technologies offer'*(Gaver *et al.* 1992,33).

The paradox with media spaces is that to incorporate privacy, one must also disable the very purpose of the system, which is general awareness. That is not to say that attempts have not been made to counter the problems, Hudson and Smith (1996) proposed using a 'shadow-view' technique to filter over the video image by shading out the subjects image, until an explicit communication effort is made. Nynex PORTHOLES (Lee *et al.* 1997) use a Blur filter to hide important details. Zhao and Stasko (1998) experimented with pixelization filters and edge detection filters. Obata and Sasaki (1998) tried to establish a rule of "interactional distance" by using virtual avatars as a precursor to communication.

It must be said that there is little evidence that media spaces have become a successful technology. There are probably a number of reasons for this, one is the stress that the system causes on the network resources, although many systems used low-resolution video with extremely low frame updates. The low video quality also probably counted against the system. The second problem is that the degree of privacy invasion was significant, and the means to counter this invasion rendered the system useless. A third fact, as later discovered (Bellotti & Bly 1996), is that many of the cases that result in informal communication and awareness occur by the user's mobility, e.g., chance meetings that occur when walking around. Media spaces, required the subjects to stay near their desktop at all times, trying to mimic the behaviour of walking around, whereas people still had to move outside the office to go for coffee, printers stations, bathroom. Whereas, as research efforts at the time of media spaces focused on supporting group work through desktop systems (Satzinger & Olfman 1992), by the mid 90's mobile technologies, like PDA's (Myers *et al.* 1998) and mobile phone systems became more popular and research effort was then directed at improving the support for these mobile people within the environment (Bergqvist *et al.* 1999, Luff & Heath 1998), rather than concentrating on desktop systems.

3.6 Ad-hoc Networks

Ad-Hoc networks operate by allowing independent devices or appliances to connect together spontaneously (Banatre *et al.* 2000) to create a network. The individual elements of the networks use a pre-established radio frequency and protocols to carry out the network communication. Alternatively, infrared is sometimes used. While many Collaborative AR systems use WLAN technology, e.g., Archeoguide (Vlahakis *et al.* 2002), ad-hoc networks are also used, e.g., in Quazoom, a collaborative AR card game (Reijers *et al.* 2002), and hence ad-hoc networks are of interest in this research topic.

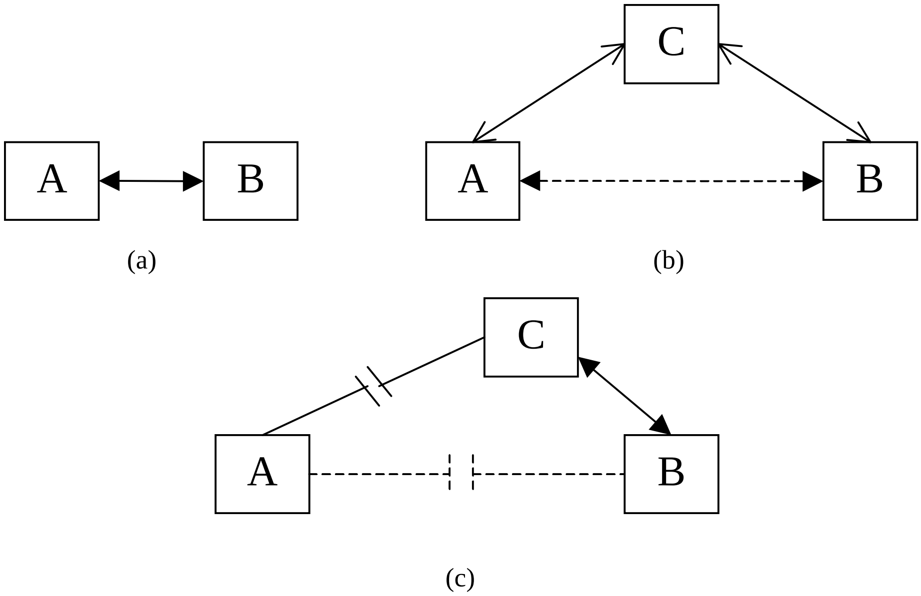


Fig. 19. (a) two devices from a network connection (b) C device join's the network, A and B lose contact, but are still connected through C (c) A loses contact with device C and hence B, C and B maintain connection

3.6.1 Ad-hoc network topology

Unlike current networks, the network plan is not usually known and is not necessary to be known. Sometimes, the composition of the networks is often known, but the means by which they are connected together is usually hidden from the user and indeed, the means by which the independent elements are connected together can change over time

Each element of an ad-hoc network is completely standalone and independent. In Figure 19(a), two ad-hoc elements come within range of each other, they negotiate a common communication channel and, while they are within range, they can communicate as a temporary network. If a third element comes within range of both or just one of the other elements then a 3-way network is established. If the third element is outside the range of one of the other elements, the intermediate element acts as a bridge between them as shown in figure 19(b). This bridge action allows networks of devices with short radio ranges to extend their range and network with a larger number of devices.

If an ad-hoc element breaks contact with another element in the network, either due to a conscious effort of its owner, or by moving out of radio range, that part of the network is now lost. Any device that was using it as a bridge to another device is also disconnected from that part of the network. This is shown in figure 19 (c), where device C has lost a connection with device A. Device B was relying on device C to act as a bridge to device A, but has consequently lost contact with device A. Device B might try and establish contact with device A, if it were in range, and ,if successful, the device C would also have contact with device A again. The loss of the intermediate node (like C) would cause the entire network to collapse instantly. Hence it is important that the protocols governing the communication links and the applications that are running over them are capable of dealing with such a sudden alteration in the network complexion.

3.6.2 Ad-hoc network procedures

The basic procedures used to establish an ad-hoc network is fundamentally the same, regardless of which actual system is used. The process usually involves

– Network discovery

– Channel negotiation

– Service discovery

Network Discovery describes the process by which one network element detects the presence of another network element. This detection process usually operates by one or both elements broadcasting their presence and the other elements listen and respond. In cases where a network already exists, the presence of the newest member is propagated throughout the network.

Channel negotiation is the process whereby the network elements communicate with each other and agree upon a common communication structure. This process also determines any additional responsibilities that may be necessary, such as the network hierarchy that is used.

Service discovery covers the process whereby each element tries to establish what services are on offer from the other network elements. Each element makes a query, or broadcasts the list of services and applications that they have. When each element knows what is available, they can then establish a connection at the application layer and perform some communication action.

Once the network is established, it is up to the participants to try and manage the network and to preside over the addition and disconnection of radio elements. Managing a continuously changing network is quite difficult. Some ad-hoc networks can establish a set of priorities, making one element control the network or having a more distributed network strategy. Once a network element ceases to be part of the network, the remaining elements must clean up the current configuration and any application must be either suspended or deleted.

3.6.3 Security for ad-hoc networks

Fixed and mobile systems have well established means for providing security and authentication within their systems. In exchanging information over a distributed network requires, the order of importance is, the need for authentication, confidentiality, integrity and availability. Authentication means that the participants should some how prove that their identities are what they claim. This normally requires 3rd parties for Public Key cryptography and Certification bodies. Confidentiality means that outsiders cannot intercept the communication. Integrity means that messages should not be altered during transmissions. Availability means that the system is available and is not encountering any denial of service attacks.

The security issues for ad-hoc networks are more complicated in a variety of ways (Stajano 2002). First, the entire physical network is usually unknown from the start and often temporary, mobile and dynamic. They are sometimes called infrastructureless networks or spontaneous networks (Banatre *et al.* 2000). Stajano and Anderson (1999) note that the concerns for distributed networks also apply to ad-hoc networks, but that the order of importance is reversed. Therefore, the order of importance is availability, integrity, confidentiality and authentication.

Ad-hoc networks pose a number of serious security issues. Apart from the standard security issues in relation to distributed networks, ad-hoc networks give rise to a number of security issues that are particular to its network and directly related to the lack of a fixed network topology. As well as the need to insure that 3rd party eavesdroppers cannot listen in on the radio waves, the ability to authenticate other members on the network is rendered more difficult by the lack of a trusted 3rd party. This is due to the difficulty in accessing a 3rd party as a result of the lack of a fixed network topology.

In most networks, there is a central point or person who exercises the security control and policies over the system. This centralised control is absent in ad-hoc networks, and the control of each radio element is the responsibility of the person or entity which owns it. This migration of control and responsibility places the emphasis on the owner of the device to assume responsibility, a responsibility that many consumers are perhaps ill equipped to make. To further complicate matters, a user may have far more than one radio element, especially in the case of Body Area Networks (Van Dam *et al.* 2001), where each element has to be individually managed by the user, complicating their control and manageability over the system.

Ad-hoc networks also face significant difficulty caused by the ease of constructing lateral attacks on the system. The most common form of attack could well be a denial of service based attack that specifically drains the power source of the appliance. As these appliances will have probably a limited power supply, such an attack can cause the appliance to use an excessive amount of power, hence draining the battery as quickly as possible. The attack could be both malicious and unintentional, e.g., advertisement spamming. A summary of the principal problems relating to ad-hoc networks are,

- Baseband attacks
- Application Attack
- Lateral attack

3.6.3.1 Baseband attack

Baseband attacks are attacks over the air interface where a third party attempts to listen into the communication between entities by listening to the baseband communication over the radio waves. This is a standard well-known method for the intercepting of information. Cellular technologies like Global System for Mobile Communication (GSM) use encryption over the air interface to counter 3rd party listeners. Wireless Local Area Networks (WLAN), which some see as a potential ad-hoc system, can use encryption over the air interface, e.g., 802.11 Wireless Equivalent Proxy (WEP), but this is not wholly integrated as part of the standard meaning that it is an option rather than a feature of the system. A WEP enhanced WLAN can talk to a non-WEP WLAN, where the WEP is turned off. Additionally, current WEP protocols have a number of weaknesses which leaves them vulnerable to sustained attacks (Fluhrer *et al.* 2001, Stubblefield *et al.* 2001).

The effectiveness of the 3rd party listening can be limited by the need to get a listening source close to the communication link. This is particularly more difficult for ad-hoc networks that may have a range as short as 2 meters to 300 meters. This means that the interceptor must be able to get physically close to the target, and depending upon the system, just be a few meters away. This still does not rule out the use of bugs close to a network element, however that would require a more concerted effort by the attacker.

Some believe that it is all right to have an open air interface, so long as there is encryption between the applications. The belief here is that the data over the air interface is still encrypted and hence safe. However the strategy then is to determine if the applications are responsible for the encryption or another application layer protocol.

However, those using ad-hoc networks should insist that the other elements have encryption on the baseband, as it will provide them with the assurances on the integrity of the system.

3.6.3.2 Application Attack

While baseband encryption prevents a 3rd party anonymously listening in on a communication, a third party can circumnavigate this by posing as a third party network element and becoming a part of the network. Once accepted in the network, they are then free to eavesdrop and attack other members, possibly by exploiting the use of service discovery protocols to try and gain access to information that they would not normally be allowed to have or infect the network with virus's and Trojan horse's. Bluetooth is especially vulnerable to 3rd parties impersonating network elements. Good authentication, firewalls and additional application layer encryption would then be required to ensure that information between trusted clients is ensured. The authentication strategy is difficult, as there is no 3rd party site that can be used, because, as mentioned, each element is a standalone utility. Other means of controlling authentication are then needed. Stajano and Anderson (1999) propose using a 'Resurrecting Duckling' model for authentication, i.e., the first time two devices communicate, their owners establish a secret unique key between them, which is then maintained for their lifetime.

3.6.3.3 Lateral attack

Lateral attacks are attacks against the system, be it the device or battery. The most common means of attacking the system is to cause a sustained battery attack. This causes the system to use up its battery by deliberately requesting information from an appliance repeatedly. Most of the network elements will be mobile and will probably only process a limited amount of battery capacity. A malicious attacker could set up a system that made repeated calls to the appliance, causing the appliance to continuously use up the battery for pointless radio transmission, which would effectively and quickly drain the battery.

Some of these attacks may not even be malicious but rather a form of spamming. To take one example, a shop has an ad-hoc element outside its store that broadcasts an advertisement to all the people passing by. This forces all of the surrounding ad-hoc elements to use their battery in dealing with the advertising, whether the user is interested in the advertisement or not. Now a row of shops, each with one of these can significantly impact on the power consumption of the appliance.

3.7 Summary

This chapter described the list of research fields that can be loosely connected by the use of the same Virtual Reality technologies such as HMD's, position sensors, audio devices, haptic devices, computer generated worlds and objects. The fields covered include Telepresence, stand alone and Collaborative Virtual Environments, Augmented Reality, Media Spaces and Ubiquitous Computing. The important technology of each field was described. Virtual environments rely heavily on the computer generation of virtual worlds and display devices to view that world. Collaborative Virtual Environments differed from the need to support communication and keep users updated with accurate real time information. Augmented Reality had less emphasis on the generation of synthetic objects, but a much greater emphasis on object registration, accurately mapping the synthetic object to the correct location and positioning of those objects. Telepresence systems were described in some detail, with the most important aspects relating to the control of physical proxies by obtaining and displaying accurate sensor information. Two types of telepresence systems were examined, traditional tele-operator systems and VR based systems. The constructs for each were identified. Video-conferencing and Media Spaces were also examined to discover the problems with using telepresence technologies in a collaborative work setting. Ad-hoc network systems and the WSI reference model were also examined as it is believed that they will be the source for ambient communication in the future and are expected to play an important role in TeleReality. Some of the more important aspects of each technology that will be relevant to TeleReality are highlighted, such as the issues of presence, equipment used, collaborative work, interaction, privacy, shared visual perspective and ad-hoc communication networks.

4 TeleReality: Enhanced Telepresence System

This chapter describes the limitations of current telepresence systems and proposes an enhanced telepresence system called TeleReality to address these limitations. Key limiting factors are multi-user support, interaction, navigation obtrusiveness and security. Many of these limitations are a direct consequence of using physical proxies. This chapter proposes using a virtual proxy with a supporting camera and communication network. This requires modifying the current constructs for telepresence systems and these new constructs are presented.

4.1 The constraints of telepresence

Telepresence and teleoperation have been around for a considerable length of time now. However, the evolution on the central constructs of the technology has not changed for half a century. The normal operation of a telepresence system consists of receiving sensory information and modifying the remote environment (Sheridan 1991). Such systems usually attach a camera to a physical controllable device (such as a robot), which is connected via a telecommunication link to an operator who views the resultant image on a Video Display Unit or HMD, as previously. The operator can usually send telemetry data to control the robot by using various control devices such as a joystick or manual control board and the constructs for this type of system are shown in figures 15 and 16. There are a number of significant constraints that are evident from such a system.

4.1.1 Multi User Access and Navigation

Unlike CVE's, where multiple people can exist in the same place at the same time, Telepresence systems are far more limited. The main problem is similar to that experienced by many collaborative work projects, limited resources required by multiple operators. Many of the resources can be used by only one person at a time and a system of turn taking must take place for the collaborative task to take place (Maxfield *et al.* 1995, 1998). Consequently, each user has to take turns to get access to the system during the experience. This means that the experience is not universal, but rather that those operating the system have priority, while others must act as observers. This is similar to driving a car, one person can drive and the other passengers are just along for the ride. The same phenomenon is apparent with telepresence systems. Each telepresence system is normally represented by a physical proxy, e.g., robot, which is then used to navigate in the remote location. Being able to navigate in the remote environment is one of the chief reasons why many people implement telepresence/teleoperation systems (Kaplan *et al.* 1997), but only one person may control the robot at any one time.

To have multiple people sharing the experience, each person must be assigned a proxy to call their own. The consequence of this action is that, in the remote environment containing the proxies, things can get quite crowded, depending upon the type of the proxy. In Paulos and Canny (1997), the proxies were Blimps, helium filled balloons weighing less than 500g, ranging from 180x90 to 120x60 cm, each occupying the space normally taken by a man. Each person was given a Blimp. Later, they used man high robots driven about on cars to compensate for the difficulty in controlling blimps, as even a slight wind would make them impossible to control (Goldberg 2000). Tachi's research group used a man sized robot for interacting with people, using retro-reflective projection technology to show the remote users' video image on the robot's face (Tachi *et al.* 2004). The size and obtrusiveness of these proxies is one reason why they are not actively used in environments occupied by people. Normally, then, the favoured use of such proxies is in well-defined spaces off limits to humans, e.g., space exploration, in fact, this was the primary reason why telepresence systems were of interest (Sheridan 1992a).

There is one means by which telepresence can be used by multiple people without the need for turn taking, and that is in video conferencing systems using fixed camera positions, where no one has the ability to navigate the remote environment and perform only limited collaborative tasks using specially designed equipment e.g., Clearboard (Ishii & Kobayashi 1992).

4.1.2 Interaction

Interaction is a critical feature for a telepresence system, i.e., the ability to actually work in the remote environment (Sheridan 1992a, Steur 1992, Zeltzer 1992). For a user to interact with the remote environment in a telepresence system, it is necessary for them to have some remote controlled object perform the physical act for them. This means that the user has to have some control mechanism, such as a joystick, with which they can directly control the equipment in the remote environment. Hence, every case where telepresence is used to perform some form of interaction between the proxy and the environment requires a custom made design for each possible interaction scenario. This is one reason why telepresence systems are dedicated to highly specialised tasks. The design of robots for Telepresence is a considerable task in itself and the design is not always easily transferable to other environments. Therefore, interacting and performing tasks within that environment requires a careful and accurate design endeavour. As telepresence systems are largely developed for working in hazardous and inhospitable environments, the relationship of interaction is between the operator and the environment. Human to human interaction is not a strong feature of telepresence systems. In cases where telepresence systems were used to communicate with people, it has been more on a strong ‘watch and follow’ basis, as was the case with GestureCam (Kuzuoka *et al.* 1994). There, the operator gave a presentation, and the students just watched.

4.1.3 Awareness

The feeling of awareness also helps to increase a feeling of presence in the environment. People will require the system to operate at the eye levels they are accustomed to. From a telepresence systems point of view, this means the robot must be able to adjust their height to match the position of the human. So, if a person is sitting down, the robot must be at the sitting level position, and when the user stands up, they must alter their height to maintain eye contact level. This requirement means that not only would you need many robots for a large meeting with a number of remote and local participants, but that these too would also need to be large (Paulos & Canny 1998, Tachi *et al.* 2004).

4.2 The Virtual Proxy

The most restrictive element for improving a telepresence system and making it widely accessible in a human dominated space, is the restrictions placed upon such systems by the physical proxy itself. It can be observed that the more constrained a telepresence system is, the less presence is achieved, and the more presence that a system accomplishes, the more complicated and restrictive it then becomes. Hence, currently, there is a relationship between the immersion offered by a system and the practicality of deploying that system. The most restrictive form of telepresence is to attach it to a wall, as in video conferencing, but at a cost of achieving any feeling of immersion. On the other hand, one could have a highly developed robot that mimics a user perfectly, but is too expensive and obtrusive. The resulting conclusion is that the type of telepresence system used and its effect on immersion is determined and constrained by the physical proxy.

As the physical proxy plays such an important part in the design implementation and immersion in a telepresence system, as well as providing most of the major constraints to such a system, one must wonder if it is possible to replace the physical proxy with something else. If one were to do so, what would it look like and how would it provide the same degree of functionality?

The physical proxy is exactly that, physical. It is a physical device that is needed to interact with a physical world. In CVE's, however, one does not have a physical proxy but a virtual proxy, or an avatar representing the user (Kim & Kuc 1998). So to deal with the virtual world one would need a virtual avatar and to deal with the physical world one would need a physical proxy. While this is the case now, it is not necessarily so in the future. It should be possible to replace the physical proxy with a virtual proxy in the real world. Hence, the telepresence user could be able to control and use a virtual proxy and have that virtual proxy represent their activities in the real world location.

The virtual proxy offers a number of advantages over a physical proxy. First, each user of the system has their own independent virtual avatar to control, which is independent of the others. A virtual proxy could also have more freedom of movement compared to a robot, as they are not limited by the restrictions of normal physics, e.g., an avatar can easily 'fly' around the room. The virtual avatar should have a richer set of possible interaction events, without the need of new physical designs, just software updates. A virtual proxy should not be physically intrusive and this opens the possibility of its use in crowded people friendly environments as well as support for more cooperative and collaborative work between people.

To support the virtual proxy, one needs to put in place a system that supports the use of a virtual proxy and allows it to perform its tasks. What is required then is to somehow take the physical world and interpret it in such a manner so that a user can have a virtual proxy. The goal is that virtual proxies would be able to explore and interact with that remote physical world and the people in it, in a similar manner as one could explore and interact with a completely synthetic virtual environment. Such a task is non-trivial in execution. However, by utilising current and evolving technology to capture the real world parameters and presenting these parameters with the use of video and augmented reality technology, it is worth studying if one can realise a virtual proxy.

4.3 Characteristics of TeleReality

The key characteristics that the constructs of a TeleReality system need are described here and originate from the desire to support presence. The derivation of these characteristics is shown in figure 20. The important characteristics are navigation, interaction, multi-user support, communication, presence, privacy and security.

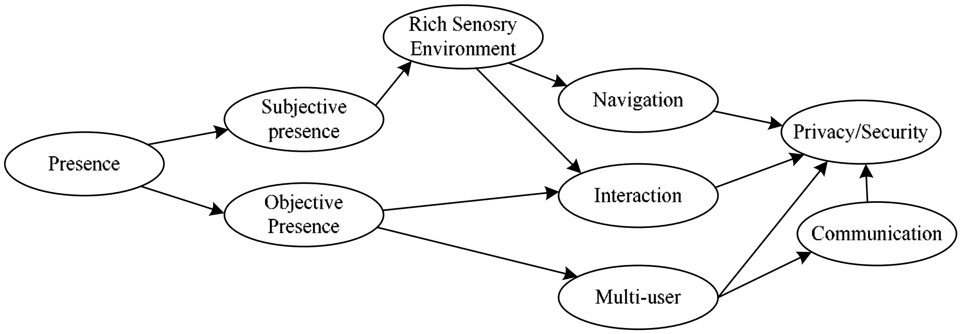


Fig. 20. Derivation of characteristics from the need for presence in a TeleReality system.

Presence: Presence, as described in chapter 2, determines how capable the resulting system is in supporting feelings of immersion and the ability to work in the environment. Presence can be divided into two parts, the subjective feeling of presence and objective presence, and the ability to perform and complete a task. Subjective presence is improved by having a rich sensory environment for the person to operate in. To experience this, the user must be able to move, or navigate the environment as well as interact with it. Performing work (objective presence) in the environment depends upon the interaction paradigms available and their usability. To enhance work between people, collaboration must be supported and hence the system must be able to support multiple users who also have a presence that is interacting and navigating the environment. To support multiple users, efficient communication means are needed, and in some cases, especially in telepresence systems as shown in the previous section on media spaces, privacy and security most be provided. Presence covers a number of topics that are important, such as the effectiveness of equipment used and gaze awareness. What follows is a further explanation of these characteristics that come as the result of supporting presence.

Navigation: The ability for the remote user to explore the remote environment is a central aspect of telepresence systems and also for CVE's. Sheridan (1991)describes this as the 'control of sensors'. Naturally, it should also be a central component for a TeleReality system also. The difficulty is in determining how one can navigate the remote real environment without the use of a physical proxy. The requirement on the system is to provide a system where the user could assume any perspective viewpoint in the remote environment.

Interaction: Interaction is essential for doing work in the environment. This is sometimes grouped with navigation, (Steur 1992), but in this thesis, it is closer to Sheridan's 'ability to modify environment'. Interaction will be performed by using a system based on an AR interface to interact with objects within the environment. A virtual object, including other people and avatars represents each element within the environment that can be interacted with. To interact with these objects, one must interact with the virtual representations of the objects. This does not mean that physical interaction is impossible, so long as the virtual object been manipulated is linked to equipment capable of performing that task. Indeed, if the virtual object represents a robotic device, one reverts to the traditional teleoperation model with all of its problems in collaborative spaces.

Multi-user support: A key feature of TeleReality that is not supported by other telepresence systems, but essential for cooperation and collaboration between people is a multi-user support. TeleReality needs to be able to support multiple users, both remote and local without the need for extra hardware or resources. This characteristic is not essentially a standalone part of the system, but is strongly linked to the characteristics of navigation, interaction and communication

Communication: The issue of communication dictates the ability to transfer the relevant information to the remote user. Unlike traditional telepresence systems, communication plays a significantly larger role in a TeleReality system. Because there are no physical devices to perform interaction, there is a greater reliance on ubiquitous computing communications based on radio signals.

The principal can be explained as follows. If one wishes to turn off the light in a room, one move's near the light and touches the switch. In practice, the brain sends signals to your hands and arms that cause the necessary physical contact. When a robot is used to turn off a light switch, the commands from the user is first sent to the robots central control centre and then to the actuators in the robot arm that performs the relevant action. For this system, the signal is sent directly to the light switch, whether by means electronic or by a servo switch, the light is turned off. What essentially happens is that our desire is expressed as a signal, which is mediated to the target by a radio link, rather than through an intermediate device. Therefore each device in the system must have some mechanism for performing an action, in order for the remote user to interact with them. Communication in our system relies heavily on signals between objects and devices to support interaction behaviour.

Privacy/security: Due to the means by which TeleReality is implemented, substantial security and privacy issues are raised. As with any multi-user system, the possibility of a security attack and violation of privacy becomes an issue. An added complexity is the reliance on multiple video cameras and the possibility of a ‘big brother’ effect (Orwell 1949). Constructs for privacy and security must be incorporated into the system from the beginning to ensure that such a system will not alienate potential users and that they can use it with confidence

4.4 TeleReality Constructs

The TeleReality constructs modifies the traditional telepresence construct chart in a number of significant ways, and this is primarily due to the characteristics of interaction and multi-user support, which are needed to support cooperative and collaborative work. While the goal of receiving sensory information and modifying the environment are still to be found, the means of achieving that are quite different. The biggest change is the replacement of the physical proxy with a virtual proxy. This single change modifies how that sensory information is received by the subject and how the remote environment is modified. Therefore, constructs such as automatic control sensors, power and actuators are not fundamental constructs. These must be replaced with some new constructs. It is also useful to keep in mind how interaction was performed in a telepresence system, it was a rather unidirectional process with the operator effecting the environment. The environment affecting the robot is undesirable, as it complicates the operators control and to the extent that the operator loses control. For TeleReality, the constructs must be able to interact in a two way relationship both with the environment and with other people as shown in figure 21. This is an important characteristic of the remote environment, it is largely autonomous as described by the AIP cube. Traditional telepresence systems would ideally rank zero on the autonomy axis, as for them, the less autonomy the environment has the better.

The means by which sensory information was obtained in traditional telepresence systems was also highly dependent upon the physical proxy, the sensors were always attached to the proxy, e.g., the camera was mounted, and sonar was part of the system. Without the physical proxy, these sensors must be located elsewhere in the environment, the preferred option been to have them in fixed locations.

The requirement for multi-user support also affects the constructs. This was a simple problem for traditional telepresence constructs, there was no multi-user support. The experience was entirely one to one, so all of the relationships between the teleoperator and physical proxy in figures 15 and 16 are one to one. This cannot be the case for TeleReality to meet the interactional framework shown in figure 21, a multi-valued relationship between the core parts is needed.

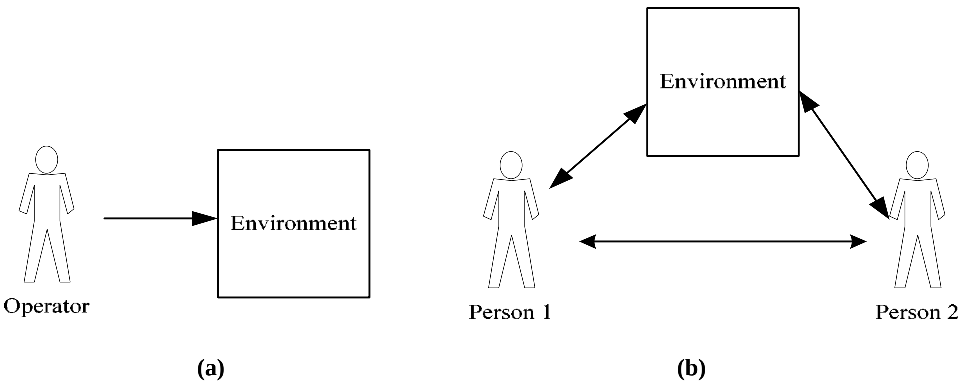


Fig. 21. a) The traditional interaction model for telepresence systems has the user interacting unidirectionally with the environment. b) TeleReality requires a more autonomous environment and also supports collaborative work with other users, either local or remote.

An additional characteristic that is needed to support interaction is to how to visualise the means by which interaction occurs. In traditional telepresence systems, one could rely on a physical tangible object to do the interacting with an object. One could see it and with force feedback get some sensation from it. This reference point will be missing from the use of a virtual proxy. To visualise interaction is an important point for a number of reasons. First, it helps the subject know what they can and cannot do in the environment. Second, it determines the effectiveness of their ability to work in the environment, and third, it offers a degree of presence in the environment. The visualisation of this is carried out by representing interactive objects as computer generated synthetic objects through an augmented reality user interface. This interface should as far as possible be similar to the interface an augmented reality user has in the environment. In fact, the only difference that the remote subject should have from the augmented reality user present in the environment is the fact that they are not physically present.

The derivation of the constructs for a TeleReality is arrived at by the need to support the key characteristics of navigation, interaction and a multi-user system, and is shown in figure 22. The multi-user characteristic is dealt with by the use of actors, of which there are three, the TeleReality actor, AR actor and cyber actor. Of these, the TeleReality actor and AR actor require a common user interface, which leads to the development of the constructs to support this user interface, namely, *control devices*, *sensor displays*, *MR environment* and *environment sensors*. The need for interaction requires autonomous agents in the environment, hence the derivation of the cyber actor. As interaction primarily focus’s on informational exchange, an *information storage* construct is added, and this is linked to the cyber actor and to the TeleReality and AR actor through the *MR environment*. Interaction also requires the use of communication medium to transfer the information between all of the various actors. Navigation supports the need for a rich sensory environment that is comprised of two constructs, *environment sensors* to capture the data, and an *environment model* to build a model of the shared world, which is then used by the MR environment to display to the actors.

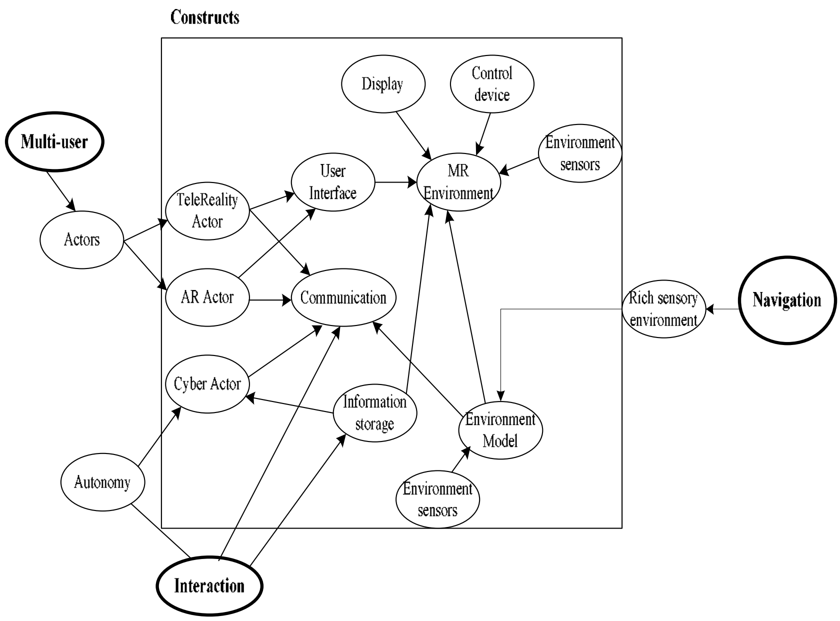


Fig. 22. The derivation of the constructs from the requirements for Navigation, Interaction and multi-user support.

4.4.1 TeleReality construct model

The basic TeleReality construct model is shown in figure 23 and the multi valued roles between the primary parts of the system are shown in figure 24. There are two broad differences between this construct model and those models used for other telepresence systems. First, there is no physical proxy, hence no actuators and sensors required to monitor their position, or power to move the proxy. The second broad difference is that previous telepresence systems were single user systems. The addition of the AR actor, i.e., a person in the remote environment capable of interacting with the telepresence user, and cyber actors, which are autonomous agents in the remote environment, is also a significant change. The AR actor has an identical construct arrangement to the TeleReality user. The cyber actor has an *autonomous agent* construct, which acts as a basic intelligence to replace the human decision process and does not require the Control devices and sensor information constructs.

The construct diagram shown in figure 23 is split between five main parts,

- TeleReality actor, i.e., the telepresence user of the system.
- Communication medium.
- Virtual proxy.
- AR actor, i.e., a person in the remote room.
- Cyber actor, an autonomous agent within the remote environment.

First, the constructs for control devices and sensor display is identical to previous constructs in other systems, although their relationship to other constructs has changed. This change results from the situation that the user receives all of their sensory information through the *Mixed Reality environment* construct. As a comparison, the VR telepresence constructs had two independent sources of sensory feedback, one from the VR model and the second from a video camera. Here, however, the video feedback is essential for the construction of the Mixed Reality (MR) environment, where it is mixed with other information sources and, as a result the MR environment, is the central repository of all of the sensory information needed by the *sensor display* construct.

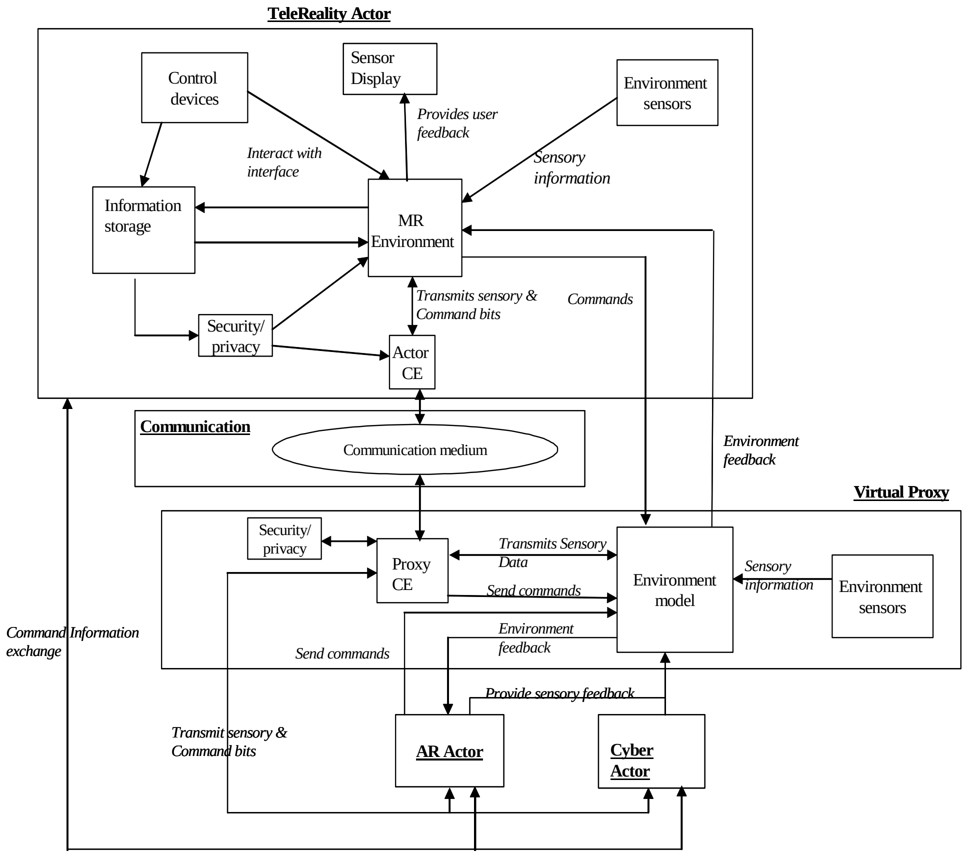


Fig. 23. Construct chart for an enhanced telepresence system with virtual proxies. The multi-valued roles are not shown here, neither are the AR and Cyber actor constructs.

The *MR environment* construct is a significant change from other telepresence systems, but is a common feature of AR systems using video mounted closed HMDs, although, in this case, the video image is from a remote location, not the user's present location. This construct is used to construct information based on received sensory data from the remote environment, as well as synthetic models which represent information both from the user and from the AR and Cyber actors. These synthetic objects are rendered over the video images and presented to the user. Additionally, the MR environment can be interacted with by the user using the *control device* construct so as to modify the user visual perspective in the remote environment, as well as modify the synthetic objects to obtain or transmit information.

The security/privacy construct is far more necessary for a TeleReality system as there is considerably more planned communication between people, and, hence, a need that information is safe from unauthorised personnel. Additionally, the type of camera system used in the virtual proxy gives rise to questions of privacy, which require policies that must be incorporated into the system to maintain trust.

The *Information storage* construct deals with how information is stored, used, when that information is transferred from personal storage areas to the MR environment and how it is modelled there. It also acts as the storage area for information received from the remote environment that is not transitory, e.g., information from the remote environment that the user is interested in keeping. It performs similar functions to the *VR model* construct in figure 16, except that the range of actions performed is greater and it can be independently modified by the control devices, e.g., text editing by keyboard.

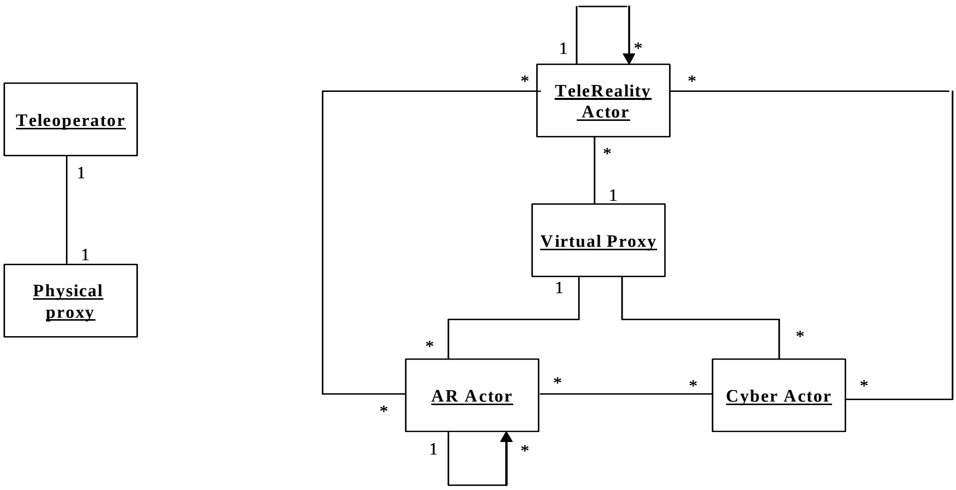
The virtual proxy is an entity that contains some new constructs, but also keeps some old constructs. The environment sensors construct is vital for this system as it obtains all of the scan information about the remote location, such as the video, audio and depth information. This collection of information is then given to the environment model that constructs a world model based on the information received.

The *environment model* is perhaps the location where the most intensive processing is carried out by the virtual proxy. It is responsible for building a real time audio-visual-informational model of the remote world which is presented to the other actors. It is responsible for sending to the telepresence user the correct sensory information that represents their perception of where they are in the remote world. It captures what information the actors wish to make public, it is responsible for ensuring that the synthetic objects representing that data are placed in the correct physical location within the model for others to see. It effectively acts as a broker for the exchange of information based on sensory data. To explain this another way, information the users have is represented as synthetic objects, and the environment model holds the location and properties of these objects.

The proxy communication element is responsible for acting as the CE element within the remote environment that ensures communication between the telepresence users and the actors within the remote environment. This single communication proxy means that the telepresence user does not need a separate parallel long distance communication link with each actor in the remote environment, saving considerably on cost. It also enforces the principle that the telepresence user can only interact with actors that they can get sensory information from, i.e., people they can see and have a sense of co-presence with.

4.4.2 Multi-valued relationship between constructs

An important differentiation between the TeleReality construct model and previous systems is that this system must support collaborative and cooperative work, i.e., the system has more than one independent actor. This adds in a requirement to define the multi-valued relationships between the actors. Multi-valued roles do exist in previous construct models, e.g., a robot with more than one actuator has more than one automatic control sensor. In saying that, the overall relationship is strictly one operator to one physical proxy. In a system where there is more than one robot in the remote environment, they operate as entirely separate and parallel systems with only physical interaction between the robots.



(a) (b)

Fig. 24. a) The normal multi-valued relationship for traditional teleoperation systems. b) The multi-valued relationship between the main actors in a TeleReality system

Here, there is a system connection between each of the main actors, i.e., the virtual proxy, the TeleReality actor, AR actor and cyber actor. The multi-valued relationship is shown in figure 24. There is, in effect, only one virtual proxy, while there can be many TeleReality, cyber and AR actors. Each TeleReality actor can communicate with a number of other TeleReality, cyber and AR actors. This is also the same for AR actors. Cyber actors are normally autonomous agents that exist in the remote environment for the benefit of AR actors and TeleReality actors. Cyber actors are not intended to strike up spontaneous connections to other cyber actors, but that their connections between each cyber actor are known at system start up. This makes it easier to control what the cyber actors are doing and how they expand their resources.

4.4.3 AR and Cyber actor constructs

The constructs that make up the AR actor and Cyber actor are not detailed in figure 23. The AR actor constructs are shown in figure 25. The AR actor constructs are those for a typical AR system, and one can note the similarities with the TeleReality actor. There is one construct added here and that is the privacy/security construct, which is not integral in most AR environments, due to the general non-commercial applications so far. An additional fact to note is that AR systems can have environment sensors (e.g., to locate head orientation) that are independent of the virtual proxies' environmental sensors, although the AR actor should be allowed to use them.

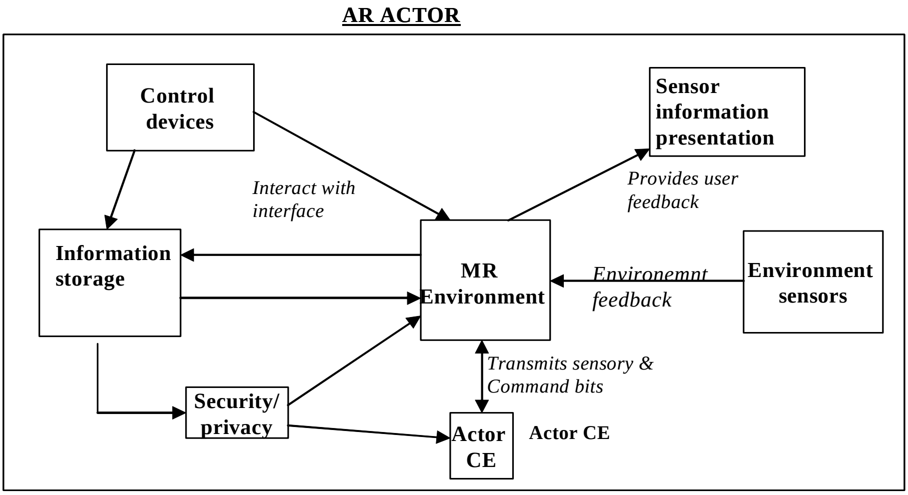


Fig. 25. Constructs for an AR actor. They are almost identical to those for a TeleReality user if video based closed HMD are used.

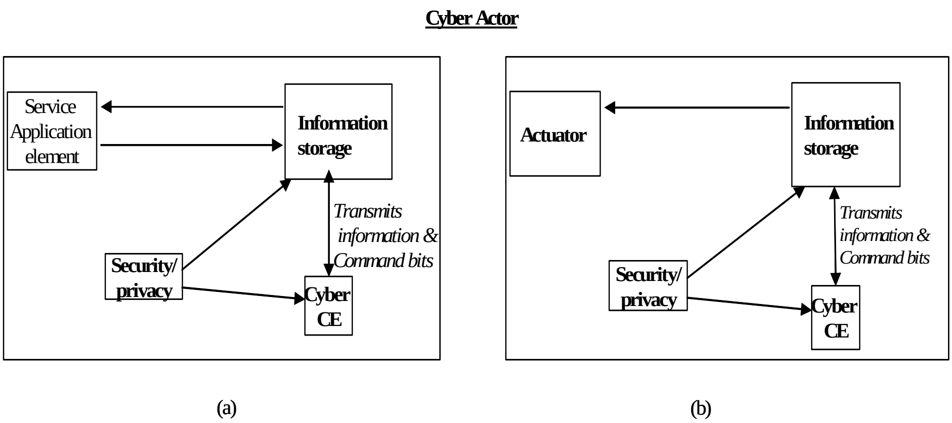


Fig. 26. Two versions of a Cyber actor. Each one has the construct for information storage, security/privacy and Communication element. a) A pure information only cyber actor, the service application element provides the behavioural model. b) An example of a cyber actor that turns on and off a light, there is no need for feedback.

The cyber actor can be of a far more complex construction. It can provide only pure information, perform a limited physical (e.g., an actuator that turns on/off the light) task or represent a full blown robot, in which case it would have many of the constructs of a physical proxy. Indeed, the cyber actor is essential for any physical interaction that the TeleReality user may wish to perform in the environment. The potential constructs are shown in figure 26. Irregardless, the cyber actor must have two primary constructs, the CE and the *information storage* construct. These are the minimum of what's necessary for a cyber actor to be seen or identified by other actors in the environment.

4.5 Technological comparison with other VR research fields

The logical constructs for a TeleReality system have been shown so far. This section compares the technological needs of a TeleReality system with respect to other VR technologies. It shows the common features they share, but will also show their differences. The technologies used as a comparison are Telepresence, Augmented Reality, and Collaborative Virtual Environments.

4.5.1 Navigation

The most important characteristics of navigation are shown in table 4. *Navigation means* focus on how the position of the proxy i.e., robot or virtual avatar is controlled. Some users navigate the environment from a fixed position, using controls to move their virtual avatar around CVE's, and TeleReality systems, or in the case of telepresence, a robot. AR users can just walk around the world normally without the need for any navigation control device.

The *navigation implementation* is important as they define how the system implements navigation, the chief means in which a connection is made to each system. Telepresence makes a normal telecommunication based connection, then proceeds to control the robot by visual feedback. In other words, it does not use any technology like position sensors to control its position. Augmented Reality relies upon a varied combination of positioning systems, sometimes used in conjunction, to determine its position and the specific orientation. Collaborative Virtual environments rely on the Cartesian plane used to generate the 3D polygon points and require an additional telecommunication address to access the system. TeleReality uses a combination of remote telecommunication address and position systems as in the case of Augmented Reality.

Table 4. A comparison on the use of navigation in different VR technologies.

Navigation	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
Navigation – means	Control device e.g., Joystick	Personal movement	Control device e.g., Joystick	Control device e.g., Joystick
Navigation implementation	Address – e.g., phone no and visual feedback	GPS/ RFID tags	IP address, Cartesian system	Address, GPS based location system

4.5.2 Interaction

The various forms of interaction possible are shown in table 5. Only Augmented Reality permits a true form of interaction with a physical manipulation of objects. Teleoperation can permit the manipulation of physical objects, if the physical proxy is suitably designed.

Telepresence and TeleReality use video as the base for their interaction with the visual field, i.e., the video images are a fundamental component in determining how they behave in the environment. Augmented Reality systems are a different proposition. Some AR systems rely on a closed HMD with mounted videos providing the visual information, while others utilise non-video see through HMDs. Therefore the need for video varies depending upon the HMD solution used.

Only telepresence does not use virtual objects as a primary interaction means, instead relying on physical interactions using a proxy. It could also be argued that the work of Milgram and Ballantyne (1997) and Lin and Kuo (1999) use virtual objects. However, these are only used locally on a virtual replica of a remote robot, and the end result is actually physical interaction with the environment, not virtual. Voice or audio, is an important form of interaction, but one that is not actively used on CVE's. Usually people who wish to have an additional voice channel while using CVE's establish a separate voice connection e.g., Teamspeak, so therefore, it is not a fundamental part of the system.

The last three categories examine the types of 3D graphics systems used in interaction. Not every technology uses a computer generated graphical system as part of their system and both AR and TeleReality only use it by combining the graphics with real images. Telepresence does not use graphics as a fundamental part of their design. A graphical user interface is essential for interaction for all of the technologies except telepresence. Only CVE's use a full graphical system, with AR and TeleReality using a partial system combined with real world images or video. Unlike AR though, TeleReality can use virtual avatars as an interaction base in the remote environment. Telepresence has no need for virtual avatars as it uses the physical proxy instead.

Table 5. A comparison of interactive techniques

Interaction	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
Physical objects	Yes	Yes	No	No
Video	Yes	yes/no – depends upon the HMD used	No	Yes
Virtual objects	No	Yes	Yes	Yes
Voice	Yes	Yes	No	Yes
Graphical Interface	No	Yes	Yes	Yes
Graphics	None	Partial	Full	partial
Personal avatars	No	No	Yes	Yes

4.5.3 Communication

Communication classifies the systems based on the type of communication systems and the type of data sent across it. A comparison of the technologies is shown in table 6. AR does not support communication across long distances. Telepresence relies more heavily on video and audio communication, with some support for collaborative tasks depending upon the type of system used. That means that telepresence systems do not include collaborative work as a basic function of the system. CVE's and TeleReality are heavily based on the need for collaboration and communication between participants. AR systems are usually individually based, and the network established is mainly between the different elements needed for the system. TeleReality systems need a combination of fixed and wireless systems, while this is only optional for the other systems. Both TeleReality and Augmented Reality make heavy use of ubiquitous computing networks to provide information flow between them and the environment.

Table 6. A comparison of communication between different VR technologies.

Communication	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
Telecommunication networks required	Yes	No	Yes	Yes
Data format	Audio/Video	Object state data	Object state change, TCP/IP	Audio/video/obje ct state data
Fixed networks	Yes	No	Yes	Yes
Wireless Networks	Possible	Yes	No	Yes
Ubiquitous Communication required	No	Yes	No	Yes

4.5.4 Privacy/Security

In terms of security, the systems vary. It is not an integral part of any of the other technologies with the exception of TeleReality. Only AR does not have access to 3rd party certification. AR needs security between their appliances. Telepresence has a need of support encryption of the video images for sensitive applications. A comparison between privacy and security concerns is shown in table 7.

Table 7. A comparison of security and privacy in VR technologies

Security	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
Security – required	Varies	Varies	Varies	Yes
3rd Party Certification	No	No	No	Yes
Security control	Administrator	Person	Administrator	Personal
Encryption	Varies	Yes	Varies	Yes
Privacy	Varies	No	varies	varies

4.5.5 Presence

The different aspects of presence in VR technologies are shown in table 8. All of the technologies support immersion to some degree or other. Telepresence uses video and AR uses computer graphics to enrich the environment, but must still maintain the sense of presence. CVEs try to create presence in a virtual remote environment. TeleReality attempts to create a presence in a remote real world environment. Where the chief differences lay is in the different types of VR technologies they need to achieve the experience of presence. None of the technologies require haptic devices for operation, this is normally an added bonus. Only AR and TeleReality require a HMD as a mandatory component for their systems, other systems normally suffice with a Video Display Unit (VDU) of some type. Likewise, AR and TeleReality also require hardware to continuously monitor their position and orientation, in order for other elements in the system to work. This is only an optional method for the other technologies.

Table 8. Comparison of presence equipment used in various VR technologies

Presence	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
Immersive technology	Yes	Yes	Yes	Yes
Haptic mandatory	No	No	No	No
HMD mandatory	No	Yes	No	Yes
Position sensor mandatory	No	Yes	No	Yes

4.5.6 Miscellaneous Related Issues

Table 9 shows a number of general related issues that also affect the different VR technologies. CVE systems are based on entirely virtual synthetic worlds, while the other technologies are based on the real world. Collaboration is possible with CVEs and

TeleReality. AR was normally designed to be a singular experience, although this is changing (Renevier & Nigay 2001). Of the technologies, only CVE's and TeleReality are designed to be truly multi-user, i.e., each person within the system has the possibility for totally independent actions and behaviour. Gaze awareness is possible in all of the different technologies. The need for accurate registration hardware is only required for AR and TeleReality. A form of technological collision detection solution is needed in all of the systems except AR, where it is the users' responsibility to avoid walking into things or people.

Table 9. Miscellaneous characteristics of various VR technologies.

General	Telepresence/ Teleoperation	Augmented Reality	Collaborative VE's	TeleReality
World –real	Real	Real	Virtual	Real
Collaborative Work supported	No	No	Yes	Yes
Multi-user	No	No	Yes	yes
Gaze Awareness	Yes	Yes	Yes	Yes
Registration	No	Yes	No	Yes
Collision detection	Yes	No	Yes	Yes

4.6 From constructs to practice

So far, this chapter has described the features that are valued for a collaborative telepresence system, which are a multi-user system, navigation, interactive and security. Basic constructs that would meet these goals were then described, but are still of a general nature. The following three chapters provide more detail on how these characteristics are implemented using these basic constructs, but some key points should be kept in mind.

The one question that needs to be considered is how interaction will work without physical proxies. The goal of these constructs is to support information flow between actors by representing information visually, which can be seen through a HMD and interacted with. Chapter 5 describes how interaction works and is supported by the communication and information storage constructs proposed here.

There is only one virtual proxy, and only one environment model which everyone uses for navigation and interaction. This environment model construct is identical in principle to completely synthetic computer generated world environments, except that it is based on the real world. The users of this model view a portion of this model, depending upon their position in it. This is what makes it a multi-user system, the fact that there is a common repository of information that everyone can access and where everyone can control their perspective. This is identical in principle to how people access collaborative virtual environments. The key question therefore is how does one go about building that environment model to achieve this aim, and how does one navigate that environment. This is dealt with in chapter six.

Lastly is the question of security and privacy, which unfortunately arise and affect the system because a large network of cameras in our environment sensors construct may give rise to privacy concerns. Also, the underlying communication system needed to support the interaction constructs have security issues which should be mentioned. These topics are described in chapter 7. Therefore, the information needed to support these constructs will be addressed in the following three chapters.

4.7 Summary

This Chapter described the basic constraints of existing telepresence systems, the reliance on physical proxies, lack of multi-user access and restricted interaction methods. A consequence of these restrictions is that telepresence systems are not deployed in collaborative work settings, but usually in hazardous environments. A further problem is that the work effort can be divided between getting the task done and controlling the proxy. To address this problem, a TeleReality system will replace the physical proxy with a virtual proxy. The characteristics of a TeleReality system were derived from the desire for the system to support presence and these characteristics are primarily navigation, multi-user support and interaction, with secondary characteristics of communication, privacy and security. From these primary characteristics, the basic constructs of a TeleReality system are derived. To support human-human-environment interaction, three actors, the remote TeleReality actor, AR actor and cyber actor are derived. The TeleReality and AR actor need constructs for interaction, that are *sensor display*, *control devices*, a *MR environment* construct to merge real and virtual objects, an *information storage* construct to manipulate and store data, an *environment sensor* construct to assist interaction and a *privacy/security* construct. The cyber actor needs an *information storage* construct. The virtual proxy needs an *environment sensor* construct to capture information of the environment, such as visual information, and an *environment model* construct to build a picture of that world. Each separate distributed entity requires *communication elements* for communication, as well as *privacy and security* constructs. A comparison of the basic technologies that define the different VR technologies was also shown.

5 Interaction

This chapter explores both the nature and character of interaction in a TeleReality environment. First, the general nature and behaviour of interaction within the TeleReality environment is described. Second, the interaction relationship between a TeleReality and Augmented Reality user is defined and the concept of the ‘TeleReality interactive space’ is proposed. Then, the three basic components necessary for interaction in TeleReality, the artefact, actor and transaction are explained. The organisation and relationship between these components is defined by using a modified geographically based multisphere framework.

5.1 Interaction in TeleReality

Telepresence applications offer three standard forms of interaction, proxy control, audio and visual. Collaborative work adds a fourth dimension, the ability to deal with data and information. In collaborative work, information and data must be disseminated, shared, modified, created and destroyed. In a normal collaborative work meeting, the participants can freely pass around objects to each other, such as paper, handouts, and business cards. They can arbitrarily get up and draw diagrams on a whiteboard, or use projectors. Currently, these are things that a remote user of a videoconferencing system may not do unless they are placed in a specially constructed environment. Bergmann and Mudge (1994) investigated a shared collaborative audio, visual and computer based information space. A specially constructed meeting room was designed that had teleconferencing systems and a large interactive board, which is quite similar to the Clearboard concept (Ishii & Kobayashi 1992). As a result, Bergmann and Mudge (1994) believed that the support and logistics burden is a major obstacle to the more widespread adoption of state of-the-art collaboration technology. Similarly, others tried to support remote awareness and collaboration by using specially designed CAVE environments (Raskar *et al.* 1998, Gibbs *et al.* 1999).

The TeleReality approach to interaction uses VR technology to abstract the physical equipment, services and location by using virtual representations of their objects and actions. Real objects and events are represented by corresponding virtual representations, displayed to a user on a HMD. Actions taken on the virtual objects are reflected in the real world and vice versa. More ever, objects that are entirely virtual may also be introduced. The objects communicate with each other using wireless links, so therefore, the only special equipment needed in the room itself is a radio link, negating the need for specially constructed physical proxies to support collaboration activities. The manner in which interaction in a TeleReality system can be shown, is best described by way of an example scenario.

5.1.1 Example Scenario

To describe how interaction within a TeleReality system works, a variation of a scenario that is based on a taxi-trip (Kuutti *et al.* 1999) is described. This scenario intended to show the expected benefits of using AR and remote Telepresence in operation for a future mobile phone user. A person arrives at an airport with an AR equipped phone. Having ordered a Taxi, the person travels towards his destination. As they are passing by a hotel billboard they are notified of an advertisement for a hotel by a short radio link, as shown in figure 27(a). Needing to get a hotel room, they contact the hotel using the contact details that was downloaded from the billboard to their phone by a wireless communication link. The phone now switches from an AR service to a remote Telepresence application. The following actions take place.

- The user is transported to the reception area and meets the receptionist. The receptionist sees the remote user using his/her HMD. A voice communication channel is established.
- The receptionist proposes a number of rooms, as shown in figure 27(b).
- The remote user requests to see a room. The receptionist shows him a room by transporting him to a vacant hotel room. The receptionist now acts as a remote participant as well. The remote user can then explore the room.
- In the room is the TV film channel service. This is also represented as a virtual object shown in figure 27(c). The user can select the object and view the movie schedule and costs.
- Happy with the room, the user returns to the reception area and reserves the room. The receptionist opens a virtual reservation form. The remote user takes his personal details, represented as a virtual object and transfers it to the virtual reservation form, automatically filling it in. The receptionist gives the remote user a virtual key object, which they can use to access their room both remotely and when they actually get there.
- Later, having arrived at the hotel room and left for some dinner, the remote mobile user remembers that he forgot to turn off the light. He contacts his room, selects the light virtual object and orders the object to turn off the room light.

(a) (b)

(c)

Fig. 27. (a) Receiving an advertisement for a hotel. (b) Using telepresence to view different hotel rooms. (c) Room with a TV menu object.

5.1.2 Analysing the Scenario

Analysing this scenario establishes a number of characteristics that are essential to the operation of the system. To clarify these characteristics

- Each object seen by one user, either real or virtual, is available to all of the others, in this case the receptionist.
- Each area, in this case the reception and hotel room, has a number of objects that signify some function relevant to that room. These artefacts can be manipulated by a virtual interface to acquire information (TV menu) or carry out a task (turn off the

light). Some of these objects can be manipulated physically by people located in the remote room.

- Another point of interest is that objects can be added and removed from the environment. The receptionist created a reservation form object and made it visible to the remote user. The remote user created a copy of their personal information, added it to the environment and then transferred it to the form. The personal information and reservation form was then removed from the environment, but of course stored in the hotel reservation system.
- Another point is that the users personal information card could be moved around, i.e., the objects are mobile. The present user was able to ‘hand’ the remote user a room key. Artefacts are then transferable from one owner to another.
- While some objects represented a physical reality (light switch), other objects were entirely virtual (room key).

Looking at this scenario it is possible to show that a TeleReality environment can contain a rich set of objects, both real and virtual and are largely autonomous and mobile. These objects are *available* to multiple users, can be *exchanged* and *modified* in real-time. Some objects representing real world objects are *coupled* to the actual real world objects. Objects can be *added* and *removed*. In addition, the status of the user present in the environment, whether real or virtual, does not restrict their ability to interact with the environment. The challenge of interaction in TeleReality is to make that location irrelevance a reality in as natural a manner as possible.

5.2 TeleReality and Augmented Reality user

TeleReality works by providing a user with video images of a remote location and augmenting it with synthetic objects. AR users are not remote, but receive information augmented over their current perspective. Both can be considered as being part of a mixed reality system, as the real world and virtual are mixed to some degree. The primary difference between them is one of spatial separation from the target environment. The AR user is present whereas the TeleReality is remote. Both should have the same synthetic objects that is augmenting the video for the TeleReality user and overlaying the perspective for the augmented Reality user. Both should also have the same facilities and capabilities to interact with the objects in the real world location as described in the taxi trip scenario.

In reality, an AR user is physically present in the environment and can interact with any physical object residing there. A TeleReality user, by virtue of the fact that they are not physically present, has a more limited set of possible interactions with the same physical objects. At the same time, access to purely virtual objects should be equal. Ideally, one would wish for both AR and TeleReality users to have the same set of interactions. In recognition of this issue the *TeleReality Interactive Space principle* is proposed.

5.2.1 TeleReality Interactive Space principal

The Cyber Interactive space defines the degree to which a Mixed Reality system reflects both the real world and the virtual world that the user inhabits. The principle operates thus: taking all of the potential interactions that may be performed by an AR user present in the environment, the sum of the indivisible, or atomic, interactions within that environment is acquired. These lists include a wide range of interactions, such as gestures, sound and real or virtual object manipulation. As an example, Manninen (2002) identified a subset of these interactions by breaking down various interaction forms in CVE's to their basic interaction type and this is shown in figure 28.

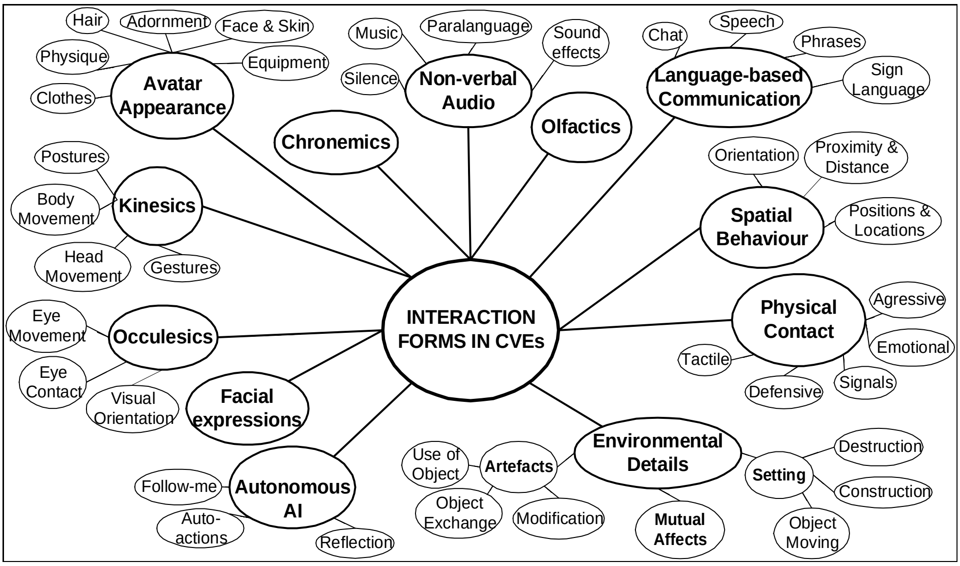


Fig. 28. Rich Interaction Model for interaction forms in CVE's

The sum of all the possible interactions is the total number of interactions possible in the environment, or the *Total Interaction Space*. For a TeleReality system, some of the interactions that may be possible within the environment are incapable of been executed because they are unable to be represented and influenced by the technological media. As an example, consider a chair in a room. A virtual model can represent the chair, and for a remote TeleReality user, the actual interactions are performed with this virtual model. The TeleReality user has limited actions with this model, they can e.g., sit down on it but they do not have the means to manipulate its position by physically moving it. An AR user would however have the ability to manipulate the chair's position as they can actually physically move it. Therefore, the full list of interactions possible with the chair is not fully available to the TeleReality user. From this we define two subsets of the total interaction space, the *Augmented Reality Interactive space* and the *TeleReality Interactive Space*, which are shown in figure 29. Each of these sets is defined as follows:

- The *Total Interaction Space* is defined as a *set containing the sum of all of the potential interactions possible in the environment*'. This comprises of a set of interactions available from the environment, but also available from the AR and TeleReality users within the environment.
- The *Augmented Reality Interactive* space is defined as *the subset of the total interactive space containing all of the potential interactions available to an AR user within the environment*. The AR Interactive space should contain the majority of the possible interactions available in the total interaction space, but they may not have access to some virtual interaction schemes reserved for others, e.g., some TeleReality user services.
- The *TeleReality Interaction Space* is defined as a subset of the total interactive space *containing all of the potential interactions available to a TeleReality user within the environment*.

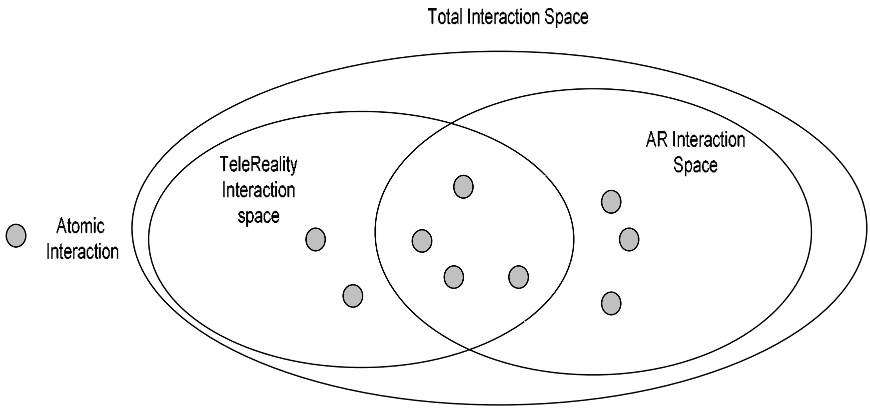


Fig. 29. Total Interaction Space contains all of the interactions in the environment. AR and TeleReality Interaction space contains their user's interactions, indicating the common shared atomic interactions.

Now, if we consider that the total sum of possible interactions is equal to 1, i.e., the total interactive space, each interaction is assigned a value greater than 0 but less than or equal to 1. A system that has only 1 possible interaction is assigned a value of 1. A system with five possible interactions of equal importance could be assigned a value of 0.2 to each interaction. Interactions of different levels of importance could be weighted differently. Once the values are assigned to each interaction and they are placed within their appropriate subsets, we can determine the value of each subset. The goal of the system is to have *cyber interaction space* with a value of 1, i.e., that all users, AR and TeleReality can perform as many of the potential interactions as possible. The sophistication, complexity and interactivity of our system can then be measured and evaluated to determine the completeness of the system. The goal therefore is to have a *Cyber Interaction Space* that expands so that it is equivalent to the *Total Interactive Space*.

5.2.2 Techniques to achieve a value equal to 1

TeleReality users cannot carry out many different potential interactions in the environment, as they lack a physical presence within the environment. To rectify this, sophisticated remote controlled actuators, robots would have to be included in the environment to carry out the interactions. This makes it difficult to accomplish the TeleReality interactive rating of 1. The more items that the TeleReality user cannot interact with, the worse the rating will be. The items that offer the greatest difficulty are those items that can or need to be physically moved or changed position. To improve the number, it is possible to adopt a ‘minimalist’ approach by removing items of clutter and unnecessary objects, thus reducing the items in the environment that the TeleReality user cannot use. For example, by removing chairs in a room, one automatically removes a source of limited interaction and improves the *TeleReality interactive space* value. Of course, there are limits to this approach. Alternatively, items can be designated as non-interactive, e.g., a painting in a room. Normal activity would not allow anyone to touch the painting, therefore it can be considered to be part of the scenery as opposed to an interactive object for the participants. All of these methods may improve the CIS value, but one should ensure that it is not done excessively, or one could damage the integrity of the environment.

5.2.3 Interactive breakdown

Based on the Heidegger view, interaction within the environment is a very important component of immersion and work. Interactive breakdown describes the situation where the interaction with a virtual representation of a physical object does not match the actual physical reality. As an example, consider a previous example where a TeleReality user’s avatar is sitting on a virtual representation of a real chair. If an AR user in the environment moves the physical chair, the virtual avatar also moves with it. This should not be possible in a real situation, as it would be far more difficult to move a chair with an actual real person sitting on it. The net effect is to damage the feeling of immersion by reminding the user that they are not operating as if they were physically present in the environment.

5.3 Basic Interaction Constructs

The basic interaction constructs used in TeleReality are based on the concepts of *actor*, *artefacts* and *transactions*. The relationships between these concepts are similar to that found in CVE’s between human controlled avatars, computer AI based avatars and non AI based computer objects. By substituting the AR user for the computer controlled AI gives the basic structure of the basic components needed to support interaction. These main components are artefacts, actors and transactions.

Artefacts: What is meant by artefact here is an independent service or application within the environment, both real and virtual. They can be manipulated, moved, copied or interacted with in some form. Each artefact resides in a location within the environment. This location applies to both the location of its source code and the position and orientation it occupies in the visual field.

Actors: Actors represent the entities in the environment who principally initiate interaction and avail of the benefits of the interaction with artefacts. There are three principal artefacts within a system, the TeleReality actor, the AR actor and the cyber actor. Cyber actors describe autonomous computer controlled elements of the environment that initiate, maintain and monitor the environment. In practice, cyber actors will be implemented as software agents. Typical examples of actors include real people, or inventory management systems. Each actor supplies artefacts to the system.

Transactions: Transactions describe the means in which messages or signals are exchanged between artefacts and between actors and artefacts. Transactions can come in many forms, e.g., gesture, service request, protocols. Interaction is only possible between actors and artefacts by using transactions.

5.3.1 The multisphere model

Interaction is mainly performed between the actors and the artefacts through transactions. It is very useful to describe a framework in which the relationship between these constructs is defined. A modified version of the ‘Multisphere model’ proposed in Wireless Strategic Initiative (WSI) ‘Book of Visions’ (IST 2000) is used to form the basis for establishing this relationship (Hickey & Pulli 2001). The modified multisphere model is shown in figure 30 and differs from the original model by sub-dividing the immediate environment into two separate parts, while combining the ‘instant partners’, radio access, interconnectivity and cyberworld presence layers into one global sphere. The reason for combining these spheres is that the author feels that they are strongly linked anyway, and offer no increased level of understanding to the interactive model.

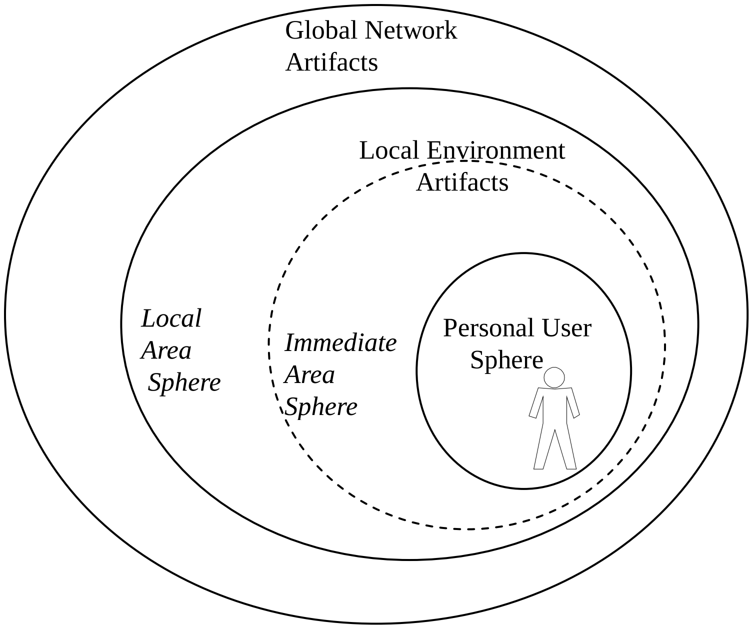


Fig. 30. A modified version of the WSI Multisphere model

The basic's of the multisphere model is focused on an I-centric design with a strong focus on a geographical based organisation of artefacts. The main reason for having an I-centric view for this grouping is that it centres more on each user's view of the system, be they a TeleReality or AR user. By focusing on the user, the design is more concentrated on achieving their sense of presence. The focus on the geographical model also reflects the reality that this is a real system, and hence, each user operates within and with the environment they inhabit. It makes some sense then to use a real world based model to define the interaction relationship. A summary of these spheres is presented in Table 10.

The first grouping of artefacts is to collect all of the artefacts that belong to each actor, i.e., the virtual and real actor. These artefacts are collected into the *Personal User Sphere*. These artefacts are initially only available and observable by their owner within this sphere. Although they exist in the environment as such, only their owner can access, see or are even aware of their existence. The owner may make these artefacts available to others or the system, depending upon their needs. The user may interact with these artefacts at will. The user may decide to make his artefacts available to all of the other participants, selective actors or to none at all. If for example, the user has a personal artefact called 'business card', he may hide this until he wishes to use it. Then he may show it publicly, whereby anyone can take it, or hand it directly to another specific person. Considering the wide range of virtual artefacts that may be owned by a user, the *personal user sphere* is an important means of controlling access to their artefacts and maintaining both privacy and security.

Table 10. Summary of layers in the multisphere model.

Name	Purpose	Range	Service Types
Personal Service Sphere	A network of objects owned by the user.	0-2 m	Watch, calendar, phone, Internet, music player
Immediate Area Sphere	A small area containing services specific to that area, e.g., bus stop	1-10m approx	timetables, printer services, shop advertisements
Local Area Sphere	larger area, providing general services to a number of people	e.g., 1-200m	Phone services, car navigation, traffic reports, directory services
Global Area Sphere	Access to the larger world, e.g., internet access	???	Web forums, TV, Web pages

The next means by which artefacts may be grouped is to assign some artefacts so that they are only available from a specific limited area within the environment. This *Immediate Area Sphere* is used to assign some artefact tasks that may not be needed widely. This is similar to the idea of having a print room, coffee room or meeting room. It will sometimes be convenient to limit availability of some artefacts to a specific area. This grouping is done also to prevent information overload in the environment and is also reflected by the reality of the real environment itself. An example of such a restriction is a meeting room with a projector. To have a meeting using the projector, a projector artefact is made available only in the meeting room, the virtual user may access the projector which displays a real slide show to the participants of the meeting. There may be no need for this artefact outside the meeting room, hence the artefact is restricted to that room.

The third level is that of a *Local Area Sphere*. This usually encompasses the entire environment or a large grouping of local area spheres based on a geographical basis. Artefacts here are available to all participants and cover things that are more general to the entire area. An example of this is a phone directory service in a building. This is a general service that can be accessed anywhere in a building and is not tied to one specific purpose room.

The final sphere is the *Global Area Sphere*, which gives the users access to artefacts that are available from the wider world, such as telephone artefacts, web access and databases.

There are a number of advantages to the multisphere model. The main advantage is that it helps to organise the availability of a wide range of artefacts by putting them into a geographical context. When one e.g., uses a VE, it is quite normal to limit some functionality to specific areas within the virtual world. When one goes to that area in the VE, one knows that one can shop, meet people. The second advantage is that in a large environment it can help to filter out the availability of artefacts. The third advantage is that it helps to increase the feeling of presence. Nothing can dampen the feeling like seeing an object that does not fit into the current context that one sees. As an example, imagine walking into a pet shop and receiving artefacts from the nearby butcher’s shop. The fourth advantage is that the multisphere model can help to form a means by which security and privacy can be organised, by preventing access to artefacts offered by certain areas.

5.3.2 Construct relationship

The multisphere model shows the relationship between a single user and the artefacts that they receive from the surrounding environments. In applying the basic constructs of actor, artefact and transaction using this model, the actor becomes the user at the centre of the model. The artefacts become the different artefacts and the transactions are the communication means in which information is shared between the actor and artefacts and between artefacts. To complete the model of the relationship between these constructs, the relationship between actors must also be shown and the effect this has on the artefacts. The picture of how this works is shown in figure 31.

In figure 31, each actor is given their own personal user sphere. The real virtual space contains all of the artefacts, real and virtual, available within the system. Each one of the actors contributes some artefacts, at a bare minimum, a communication artefact with its communication presence. The AR and TeleReality actors maintain a personal user sphere, where their artefacts reside. They may wish to make these public to the other actors if they wish. The cyber actors reside in the local area sphere, providing artefacts that are generally available to AR and TeleReality actors, as well as to the general upkeep of the environment.

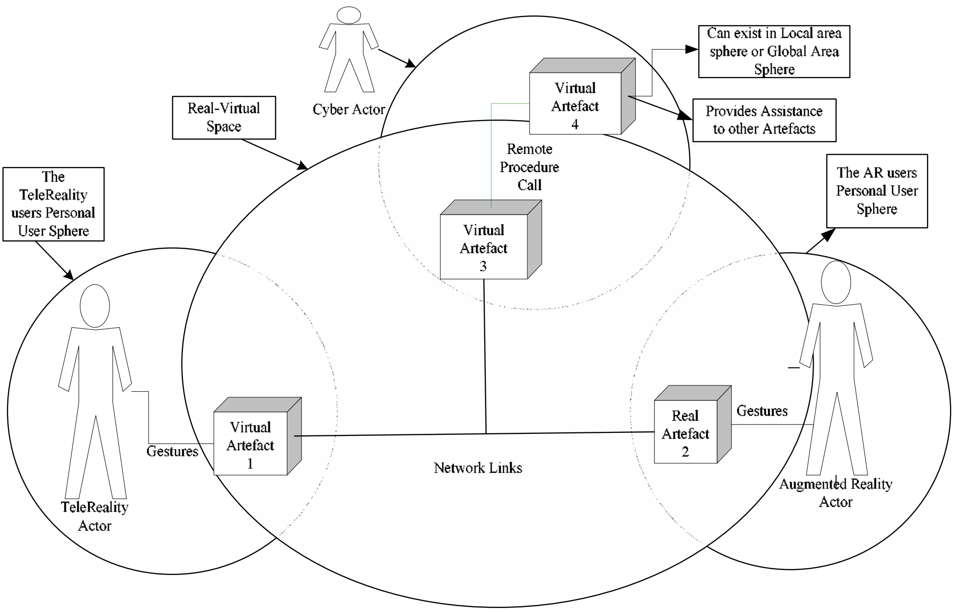


Fig. 31. The relationship between the different actors, artefacts and transactions.

In figure 31, the TeleReality and AR actor contribute artefacts 1 and 2. The cyber actor contributes artefact 3 and 4. Of these artefacts, the AR and TeleReality actor only see artefacts 1,2 and 3, as 4 is hidden to them although it is being used by artefact 3. In principle, artefacts that exist in the real/virtual space are not automatically shown to all and can be removed from view if necessary. There are different types of transactions between the artefacts and between actors and artefacts. Examples of these types of interaction include using gestures, network communication and remote procedure calls.

5.4 Artefact Mobilityand Interaction

Artefacts in the TeleReality system are not strictly static but can be highly mobile. Some of the artefacts are tangible, i.e., they are independent objects in their own right and can be moved around and manipulated, e.g., a business card. To support artefact mobility, objects can replicate and migrate, two features common in CVE's.

To understand what is meant by artefact mobility, the basic structure needed for the artefact will be described. Each artefact contains software code. This software code resides on a physical device. Therefore each artefact has a *location* on a computing device that can be accessed by other artefacts through software calls or over telecommunication links. The artefact needs an address, e.g., an IP and port address. This is the location of the base software code that contains the data and functionality that the artefact intends.

Each artefact can also have a *visual location*, i.e., each artefact can, at some point in time, occupy a place in the visual field. This visual location is used so that the artefact object can be rendered onto a HMD display.

When working with artefacts by moving, copying, deleting, interacting or creating them, one is working with either of these two locations. As an example, one may have an artefact called 'calendar' on one's personal mobile device. The physical location of the software is on the mobile device, while the visual location of the rendered object calendar is rendered on the HMD at a point relative to the user's location. The user can move the visual location of the calendar to a new more convenient location within their visual field, while keeping the physical location of the device intact. Likewise, the user may decide to move the physical location of the artefact to another device, say their home computer, but because the visual location has not changed, the user still has the same services as before. A reason for moving the physical location of the code may be that the mobile device lacks the desired resources and is not used often. The feature here is that the visual representation of the artefact to the user is entirely separate to the actual code location.

5.4.1 Main Artefact Components

To support this type of behaviour, each artefact consists of two parts, a client side, which contains the information on the rendered object and potential interaction events, and the server side which contains access to the artefact behaviour and data e.g., databases. It is this client side of the code that is downloaded to the user of the artefact and contains the visual parameters of the object and its location, i.e., the position within their field of view that they can observe the object. A client-server relationship is established with the server code, as shown in figure 32. The client contains little data, just enough functionality to send UI events to the server code and display any information that is returned.

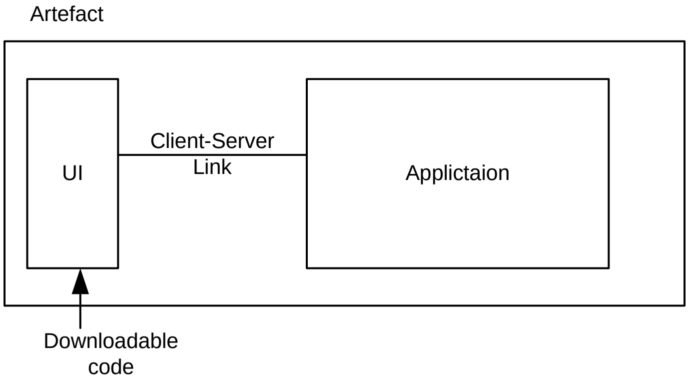


Fig. 32. Artefact with two parts. The UI element connected to the main code via a client server relationship.

5.4.2 Artefact Replication and Migration

Artefacts can be moved from one location to another, but will require the assistance of other artefacts that provide services. The basic service offered is replication, i.e., an artefact is copied. Replication can cover the copying of the artefact, including all of the data it currently holds at the time of replication, or replicating the structure of the artefact, i.e., with an empty set of data. The replicated artefact exists in the same device as the original artefact, but with a different address e.g., a different port number on the same IP address, and a different visual location address on the visual field.

Migration covers the service which transfers the artefact to another location, i.e., changing the address of the physical code and the visual location. The easiest way to do this is to first perform a full replication, transfer the copied artefact to its new destination and delete the original artefact. In performing these actions, the user may need to give new locations for both the code and visual code.

5.5 Summary

This chapter describes the constructs' components used to support interaction within a TeleReality system. An explanation of how interaction work in a TeleReality system was first presented which emphasises the way in which interaction is carried out between remotely distant people and with the remote environment. This interaction is carried out using virtual objects that act as a front end for the service on offer. The relationship between the TeleReality actor and AR actor is expanded upon to explain that they are very similar, with the primary difference between them being the location that they receive their visual information from. TeleReality actors get it from a remote location, while AR actors can see the area they are in. A second difference is that the AR actor can physically interact with the environment at will, whereas the TeleReality actor needs a cyber actor. The degree that a TeleReality user can interact with the environment is described by the TeleReality interactive principle. This is arrived at by counting all of the possible interactions that can be carried out in the remote environment, then estimating how many the TeleReality user can perform. Interactions are weighted in accordance with their importance and the goal is to achieve a value of one, i.e., all possible interactions are supported.

The principle parts of the components of the interaction scheme are introduced, the *actors*, which are the main protagonists in the system, the *artefacts* which represent the active objects in the system and the *transactions* which is the means in which information and actions are exchanged. Artefacts are owned by the different actors. Artefacts can also be replicated and migrate between locations. Artefacts are located according to a geographical based system, they reside in the personal user sphere, the local area sphere or the global area sphere. The purpose of this organisation is as a means to filter out the available objects according to their geographical relevance, as well as to closely link them to the partitions described in the next chapter.

6 Visual Representation and Movement

A key feature for a TeleReality user is the ability to navigate a remote real world location. The previous chapter discussed interaction within the environment. Here, the visual representation of the environment itself will be studied. The objective is to replace the normal physical proxy with a virtual proxy. The virtual proxy must support navigation around a large space, providing their owner with the correct visual perspective to match their current position and orientation in the system. Two constructs, the *environment model* and environment sensors are responsible for accomplishing this. The environment sensors constitute as a system of strategically placed cameras placed in the remote environment. These images are sent to the environment model which contains image processing capabilities to calculate the correct perspective image for each user. A complication of using this system is the necessity of a very high communication bandwidth to support a video based solution. The main problems and the solutions proposed in this chapter are shown in table 11.

Table 11. Issues affecting the visual field plus a description of the proposed solutions.

Problem	Solution	Description
Insufficient communication bandwidth	Visual Scenery and objects	Visual scenery is rarely updated while only visual object state information is updated
Support for virtual proxies	Fixed camera systems	Deploy a large number of cameras and use image processing technology to calculate the correct perspective images
System type	2D/3D and static/dynamic	Gives various models from a simple, less immersive system (2D/static) to a more realistic and immersive system (3D/dynamic)
Navigating large environments	Partitions + handovers	Subdivide large areas into subgroups called partitions. Support handover from one partition to the next. Handover is determined by FOV of device, movement and user position
Location system	Symbolic/geographic	Use symbolic address for the partition and tracking systems for exact location within system

6.1 Visual Components

The primary focus and most challenging aspect of TeleReality is the visual field, i.e., what the user sees, how they see it, how it is represented and how that representation is communicated over a telecommunication channel. There are essentially two different types of components within the visual field that have to be accounted for, the *visual scenery* and *visual objects* shown in figure 33.

6.1.1 Visual Scenery

The visual scenery component represents the static non-changing imagery in the visual field. These include images from objects such as walls, floors windows or any other non-mobile object. Objects can be added to the scenery, e.g., paintings, pictures, but they do not fundamentally change in appearance very often. Another aspect of the scenery includes objects that are not considered within the interactive frame, i.e., objects that do not have interactive primitives. As an example, a table can be an interactive object, it can be moved or its orientation changed, yet for the most part the participants do not interact with the table on a heavy basis. Therefore it is possible to designate the table as a scenery item. To be considered as a scenery item the object must

- Maintain a constant and fixed position and orientation
- Have no interactive constructs regarding position

The advantages of the visual scenery objects are that once they are known they need to be rarely updated. Once the system is started, the scenery objects are sent to the remote user, and there is little need to continuously update the scenery information with follow up communications. Many video-coding systems operate in this manner, using efficient codec's that only transmit changes in the images (Waggoner 2001).Some of the problems with the scenery objects are that while the objects themselves may not change considerably over a long period of time, the lighting conditions under which they are viewed will. The change in lighting conditions will have some noticeable affects, e.g., the projection of shadows. Using existing video codec's that transmit the new image to reflect the change in light in the room currently covers these changes. However, if the light sources within the target environment can be modelled and monitored, then it is possible to maintain the original image and only send the current state of the light source. That would reduce the data transmission significantly, as well as allowing the remote user to brighten up a place they are visiting for better visibility.

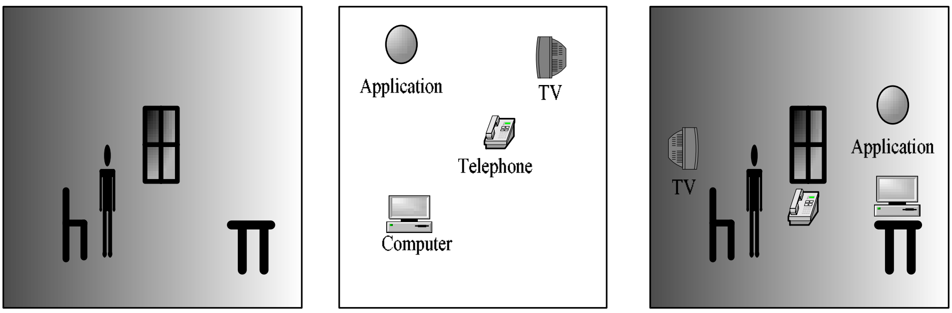
6.1.2 Visual objects

Visual objects represent objects within the environment that are interactive in nature, objects such as doors, humans, TV's, books. These particular objects shown in figure 33 (b), are at the disposal of the user to control and manipulate. They are essentially the visual front-end client interfaces of the artefacts described in chapter four. Unlike the scenery objects, it is expected that they will be altered regularly and are active in the environment. To be considered as a visual object an item must be

- Independent of the environment
- Interactive

Independent: Each object exists separately to the environment that they exist in. This means that the visual object itself is not an integral part of the environment. This allows the same object to be used in multiple locations, although the artefacts offered may vary.

Interactive: The object is interactive and exists to perform some function in the environment. It can be manipulated, moved and changed. The object is normally the front end to an artefact.



(a) (b) (c)

Fig. 33. (a) Three background visual scenery objects. (b) some visual objects (c) A composite picture containing visual scenery and visual objects

Each virtual object used in a TeleReality system, irrespective of whether it is 2D or 3D, must have its known *geometry* and appropriate textures sent to the visual display of each user. To exist in the environment, each object must be assigned a *position* in which to reside. The initial action in placing an object in the environment is to render the objects geometry in the position specified. A change in the object's position, as the result of an interaction event, causes a transform to be applied, changing the object position. It is these changes that are transmitted over the communication medium. This means that there is no updating of the object unless an event has taken place. In effect, visual objects are handled in a similar manner as objects in CVE's like MASSIVE (Greenhalgh & Benford 1995) and Bricknet (Singh *et al.* 1995).

Real objects can be used in conjunction with virtual objects that accurately model their characteristics. In this case, the virtual object reflects the actual state of the real object, i.e., they are tightly coupled. An alternative approach is to use a camera system to identify the objects in a room and build the model itself (Mohr *et al.* 1998). Alternatively, a system of active (laser) cameras to determine 3D object models and passive cameras to generate colour and texture can be used. However significant problems remain, including the merging of the information from the two cameras types to give an accurate model (Geist *et al.* 1997).

6.2 Presentation - 2D vs. 3D

A major classification of any system is whether it should be a 2D based system, a 3D based system or a combination of both. Which mode the system takes is of a major importance in deciding the complexity of a TeleReality system. A 2D system offers a window to the real world, the *Television paradigm* (Pereira 2000). It is by far the most common and cost effective form of representation in video conferencing and remote teleoperation. The 2D solution also offers cost effective video codec's at low to high bit rates for the transfer of visual information over low to high bandwidth media. Traditional video codec's only offered the picture view of the environment i.e., the visual scenery construct, whereas newer video encoding standards are based on a composite of interactive audiovisual objects, i.e., a combination of visual scenery and visual objects (Ebrahimi & Horne 2000). However, 2D systems by their nature do not capture any depth information on the scene being shown, making detailed navigation impossible. When depth information is captured, there are significant occlusion problems. In these cases, a 3D picture of the environment is built up, but information that is out of sight of the camera and depth capturing mechanism cannot be obtained. Attempts to navigate this space will show large visual blanks when the viewer looks in those directions.

3D world presentations offer more flexibility in terms of view and navigation, but at a more considerable cost in system complexity, communication and processing resources. There are generally two approaches to scene construction. The first approach involves the construction of an exact 3D-computer model of the environment. A video camera with 3D-depth information is used to capture the visual scenery and this scenery is textured onto the 3D system. The textures are regularly updated whenever there is a significant change to the system. People and objects within the system are represented either by audiovisual objects (either 2D or 3D) or by 3D avatar representations. In the latter case, these virtual objects are heavily coupled with the real world, so if a person talks, so does their avatar (Rubin & Vatikiostis-Bateson 1998, Olives *et al.* 1999).

The second approach is based on the provision of a continuous and live 3D-video stream of the environment. This is usually accomplished by using multiple cameras, e.g., two cameras providing a stereoscopic image to the remote users HMD. The drawback of this system is the need for heavy image processing and bandwidth. Lanier (2001) uses a seven camera system and to achieve a frame rate of 2fps, he required a bandwidth between 20Mb/s and 80 Mb/s.

6.3 World Types

The combination of visual scenery and visual object components should adequately represent the visual information provided to a user of the remote environment. However, what is the nature of that environment itself? For a TeleReality system, the remote environment must be inherently real, i.e., it must be tightly coupled to the events in the real environment to maintain its accuracy. There are two alternative systems in which this goal can be met, one focuses on a *static* world model and the other’s focus is on a *dynamic* world model.

6.3.1 Static Worlds

Static worlds describe systems whereby the real world is captured and stored, usually in an audiovisual database. TeleReality users can then access this database to navigate and interact with this world. Static worlds are not intrinsically connected to the real world, but some coupling can be achieved by linking objects within the environment to real world state events. Other questions then concern how often the static world should be updated. Static worlds can be either 2D or 3D based.

An example of a 2D static world is described by Hirose’s research team (Hirose *et al.* 1999). They described a system for building a virtual world based on the real world. They developed a system that used a system of 8 cameras arranged to give an omni directional picture and mounted it on a van. They then drove the van, using GPS to track their position, through a city. From the captured images they managed to recreate a panoramic view of the city. This enabled a user to 'walk through' the city looking in whichever direction they choose, but the user was confined to the path that the original capturing device took. A later design used layered-morphing techniques (Endo *et al.* 1999) to give a photorealistic version of the system.

There were a number of limitations to this approach, one limitation is that the user may only walk the path followed by the car. The system also had no interactive element and therefore no coupling with the real environment whatsoever. The problem with the environment from a TeleReality user's point of view is that much of the information contained was incorrect. As an example, when viewed at a later time, the world showed many pedestrians and cars that would have no longer been present in the environment. The Hirose system is based on a principle of ’pathing’, where one path through the environment is taken and the users are confined to that view.

To improve the freedom of movement through the static world, it should be feasible to create multiple paths in a particular environment. By creating a number of paths sufficiently close to each other, a user can move around a given space by effectively 'jumping' to a new path. Figure 34 shows this concept, a user is on path A. He is looking in the direction of a building. He attempts to move to the direction of the building. Path B is closer to the building so the user 'jumps' to path B and now receives the images available from path B's database.

Such a system is not without its disadvantages as well as its advantages. In terms of it's advantages, by using the layered morphing method and a video multi-resolution technique (Finkelstein *et al.* 1996), it should be possible in theory to provide seamless ‘path jumping’ to provide a feeling of freely moving about an actual space. The path separation factor would probably be entirely dependent upon the size of the space that the system should represent, small enclosed spaces may require a single path whereas larger spaces would require multiple paths as shown in figure 34.

With regard to the lack of coupling with the real world, a potential solution is that those dynamic objects within the real world location should send to this virtual world their updated location and status. As the original video was captured with GPS data for each point on the path, it should be possible to update the world to include virtual representations of these real life objects, e.g., a tram can continuously update it's location and the user would then seen the tram passing by him/her.

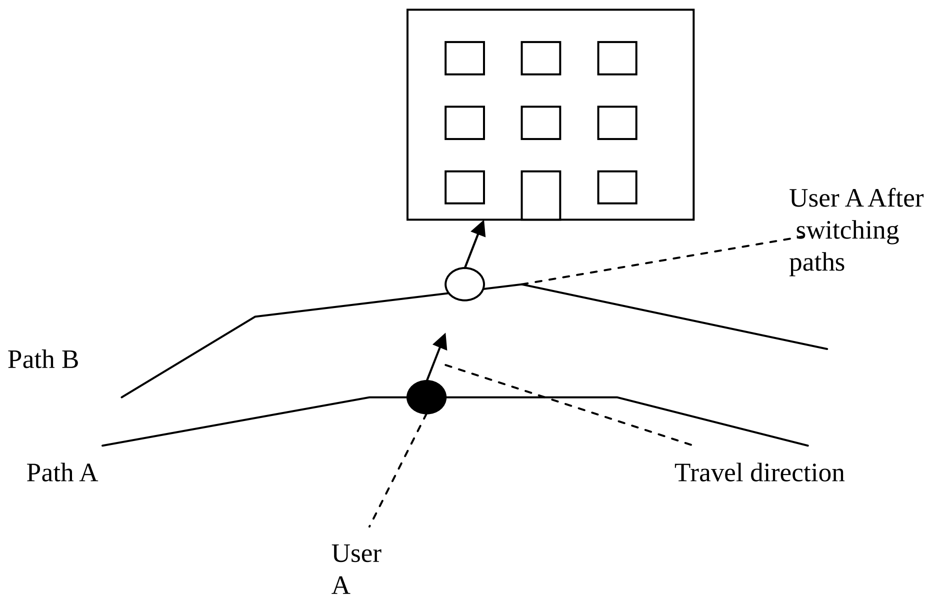


Fig. 34. Jumping to a different path

In figure 35, the general format of the static world model is shown. First, images of the real world are captured and stored in either an audio-visual or virtual database. This is updated either as a once off event or on a periodic, non real-time basis. When the TeleReality user wants to access the world, the front-end system constructs the world based on the information contained within the database. When the real world coupling is performed, a link is made to a visual object within the remote environment, and this link is continuously updated in real time by changing the state information (i.e., orientation, shape and location).

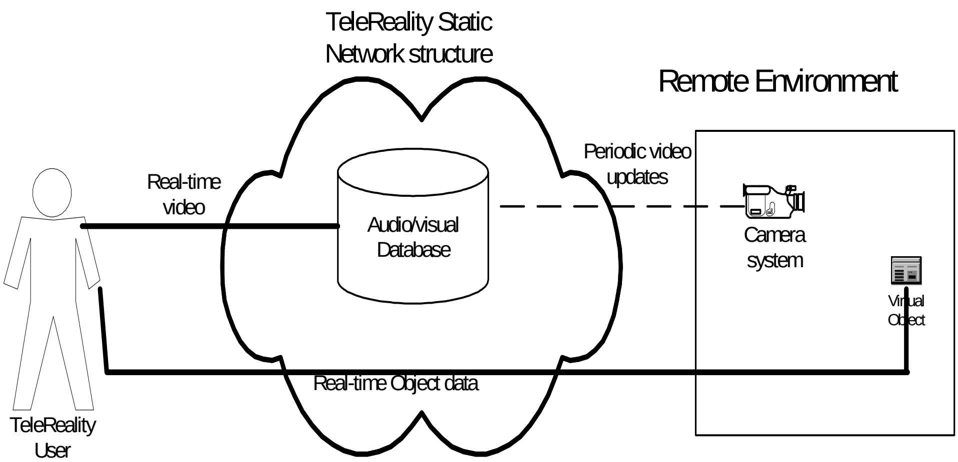


Fig. 35. Static System: Visual information is stored in an audiovisual database while artefacts are updated in real-time.

The main issue with a static system is to ensure the fidelity of information available to the TeleReality and AR user present in the environment. The user in the environment will see the reality of the environment, whereas the TeleReality user is relying on a stored memory of the system. There is a real potential for a mismatch in perceptions and information based on the old stored model. Supporting interaction with people in the remote environment is also difficult, unless the user can be represented as an object, e.g., an MPEG-4 audio-visual objects of users to support their presence in virtual environments (Rauthenberg *et al.* 2000).

6.3.2 Dynamic systems

Dynamic systems reflect changes in the real world location to the remote user instantaneously. Essentially, the difference then between a static system and a dynamic system is that the periodic update of visual information shown in figure 35 is in reality done at real time, as shown in figure 36. This change in the model removes the need for the audio-visual database. This relates to a very large change in implementation with the emphasis on the provision of a continuous stream of information. Most static systems will have a period of hours, whereas a dynamic system with high fidelity should act in real time, at the optimum update rate of 25-fps (Lanier 2001). The strength of this system is that the users can have total confidence that what they view is exactly the same as what the real user located in the environment sees, and that all of the information available is current and up to date. The difficulty however, is to capture and process the relevant visual scenery and objects in real time.

The current and most popular means of capturing the visual information is to operate a '*sea of cameras*' approach (Neumann & Fuchs 1993). This approach utilises a large number of cameras, whose position and orientation is known. A number of cameras are organised in such a manner as to capture all possible viewpoints in the system. Kanade's research team utilised 51 inward looking cameras mounted in a geodesic dome to capture the movements of two basketball players, although not in real time (Kanade *et al.* 1999). Fuchs and his colleagues worked with small puddles of 2-3 cameras to gather real world information (Fuchs *et al.* 1994). Lanier (2001) used a collection of seven cameras for their experiment with tele-immersion meetings. Cooke's research team used a 2-camera system for capturing arbitrarily shaped video objects in 3 Dimensional scenes and managed to achieve a frame rate of 10fps (Cooke *et al.* 2000).

The sea of camera approach has a number of constraints, but currently is the best strategy for the realisation of a 3D system. For a 2D system, a single camera should suffice. Regardless of whether a 3D or 2D system is deployed, the constraints on the system are similar. Camera systems can only present visual information that exists in their field of view. A total camera solution would require a FOV in the horizontal of 360 degrees and in the vertical of 180 degrees. Standard cameras mounted on the side of a wall can have a wide FOV but regardless there will be blind spots behind the camera mountings as shown in figure 37. Even some omni directional cameras with FOV of 360 degrees in the horizontal can have blind spots due to the optic system that they use.

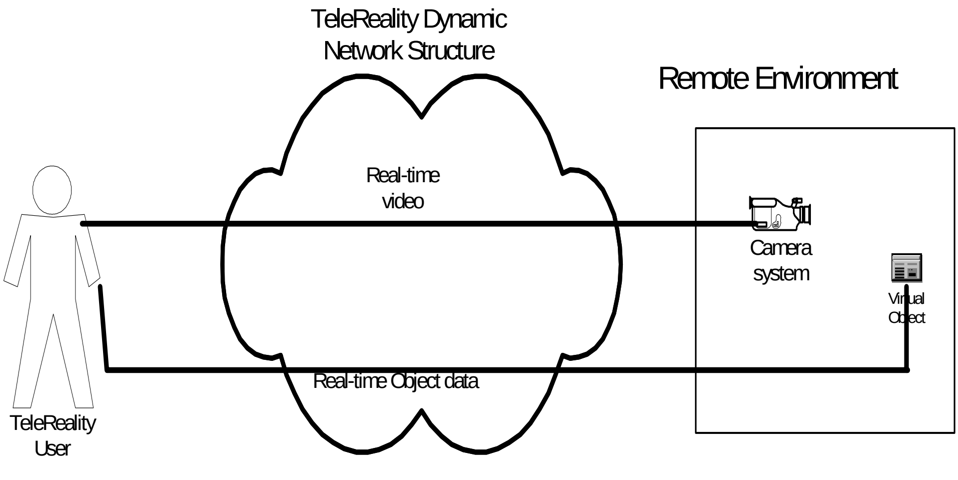


Fig. 36. TeleReality structure. Video and object information is sent to the remote user in real time.

A second issue is that the camera is usually mounted above a user's eye level, the virtual user has a look down view of the other participants as shown in figure 37. However, eye contact or gaze awareness is an important part of interaction with others in the environment. The users have a need to feel as though they are on an equivalent and equal level to the other participants in the environment. Therefore, the correct user perspective would need to be extrapolated from the camera images to acquire the correct perspective, as shown in figure 37. Therefore, one of the most important requirements is the correct user's image perspective.

The image perspective is the view that a user would have if they were physically at the room. This is challenging for the normal camera perspective. If the camera were to be configured to give an eye level view for a standing user, this would no longer apply if the user were to sit down. Also different users are of different heights, so the system would hardly be universal. Therefore calculating the perspective is not a trivial task. Systems that extrapolate 3D information from the scene are of a significant advantage for calculating the correct perspective. The second issue with single 2D camera systems is that they are only effective when the user is looking in the same direction of the camera. If the user turns to look towards the camera, they will see nothing there. Omni-directional cameras go a long way to solve this problem, usually at a cost of resolution.

A third problem with camera systems is that while a camera can look across a large space and present a satisfactory image, cameras are normally set up with a fixed configuration, e.g., a fixed focal length. The more focused a camera is at a far away point, the less information it shows of the near space area. The more focused it is at the near space, the less clear distant information becomes. Therefore camera systems are optimised to present information of the surroundings. What this means is that some information may not be covered by the optimum set-up. Therefore to present information close to the camera in good quality, distant information may not be as represented well, and vice versa. Normally, this problem is fixed by adopting remote control zoom and pan/tilt behaviour, however, for a multiple user system, it is not desirable to have someone control the camera as this affects others. Hence, when the system is configured, it is done so with the assumption that it covers a specific degree of space with satisfactory resolution.

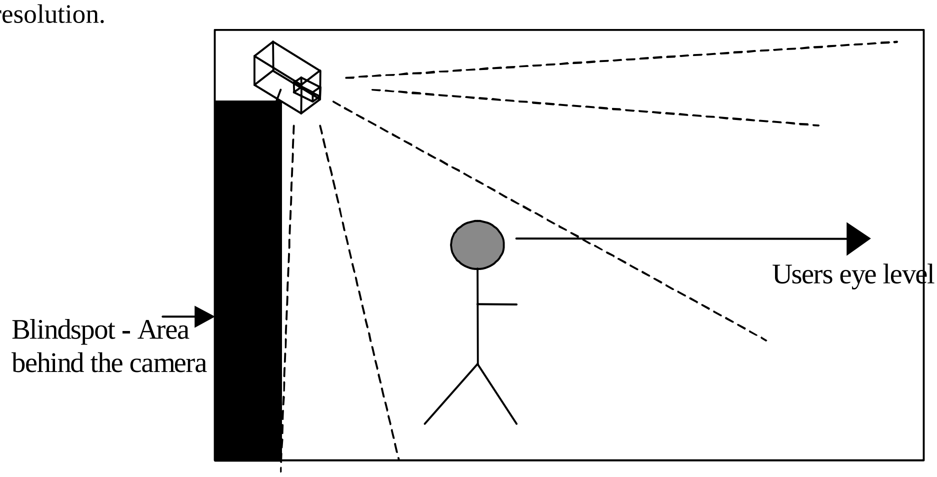


Fig. 37. The direction of the remote user's eye level is in comparison with the camera

6.3.3 Partitioning

The degree of space that the system is to cover may be far greater than the ability of our camera system to represent. For example, in an office complex with numerous rooms, one camera system will not work due to occlusion from the walls. To represent a space larger than one camera system can handle, one needs a different approach and that is to construct a number of camera systems and connect them together. This gives us another core concept, a large space is *partitioned* into a number of spaces where each space contains one camera system as shown in figure 38. Of course, the user of the system wishes to navigate the entire area in which they are present, and does not necessarily want, nor should need to know how that space is implemented. Therefore it is required to present a continuity of view in the large space, which means that the user should be able to move from one partition to the next as effortlessly and seamlessly as possible. This brings us to another requirement of our system, *visual handover*.

Visual handover is the continuation of the images presented to the user as they move from one partitioned space to the next. Visual handover is determined by the user's position and visual perspective. As an example, if a virtual user is located in one partitioned space as shown in figure 39(a), they receive their information from the cameras in that cell. When the user moves to partition B, but looks back into the space covered by partition A, they should still receive information from partition A, although they are located in partition B. To further complicate matters, part of the user view will also show information only available from partition B.

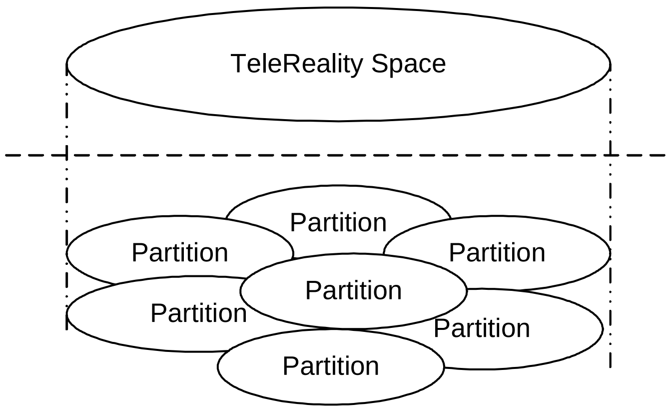
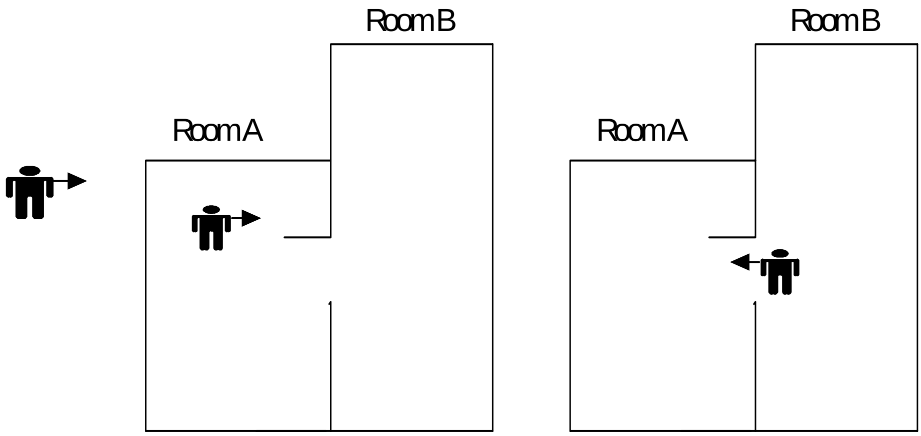


Fig. 38. The TeleReality space is divided into a large number of adjacent partitions

An additional problem is determining the point at which a person is in partition A or B. Lets say that the person is standing at the doorway between partition A and B. By looking into the room, the person is in partition A, whereas just turning around, the person is now in partition B. Therefore the perspective of the person is a contributing factor in determining the partition that they reside in.

In larger spaces, occlusion is less of a problem, but distance can be a more difficult issue. Each partition is set up so as to guarantee visual coverage of an area with a satisfactory degree of resolution. The inter-pixel distance determines the quality of this resolution. The farther away an object, the greater is the inter-pixel distance, meaning that the object is shown in less detail than an object closer to the camera. This is a similar problem to the level of detail problem in 3D graphics.

Therefore as one travels out of the partition, the resolution quality deteriorates. When the quality is too poor, it is necessary to switch over to another partition. Determining when the resolution is too poor is an extremely difficult task. The problem is that the partition itself cannot know when the resolution of an object is too bad. When the user is at the centre of the partition, the distant object's resolution is fine. At the edge of the partition, the resolution is not so good, but as it is beyond the view of the cameras in the partition, the partition does not know when it can no longer give a satisfactory resolution. To counteract this problem, the simplest solution is to take the known space in which the partition can strictly guarantee the best resolution and when the user leaves this area, the user switches to the new partition. The deciding factor here is the user's position in the system.



(a) (b)

Fig. 39. (a) A user in partition A (b) user is in partition B, but still receives information from partition A as that is where there perspective is.

6.3.4 Additional factors in determining handover

The perspective of the user is a very important component in determining handover between partitions. Figure 40(a) shows an example of the issues involved in FOV handover. In the first figure, the greatest portion of the information comes from partition A. In figure 40(b), the greatest portion of partition B, and some information from A. This would be the optimum time for handover information. A main factor in this calculation is the FOV of the HMD. Most HMD's will have FOV that are substantially narrower than that of the human eye. Therefore, for the TeleReality user, as they approach a handover point, the limited FOV of the HMD creates a more concentrated tunnel of vision than what the user could see if he were actually present in the room. Essentially, the wider the

FOV of the HMD, the closer the handover point will be to the door, while the narrower the FOV, the farther away the handover point can occur.

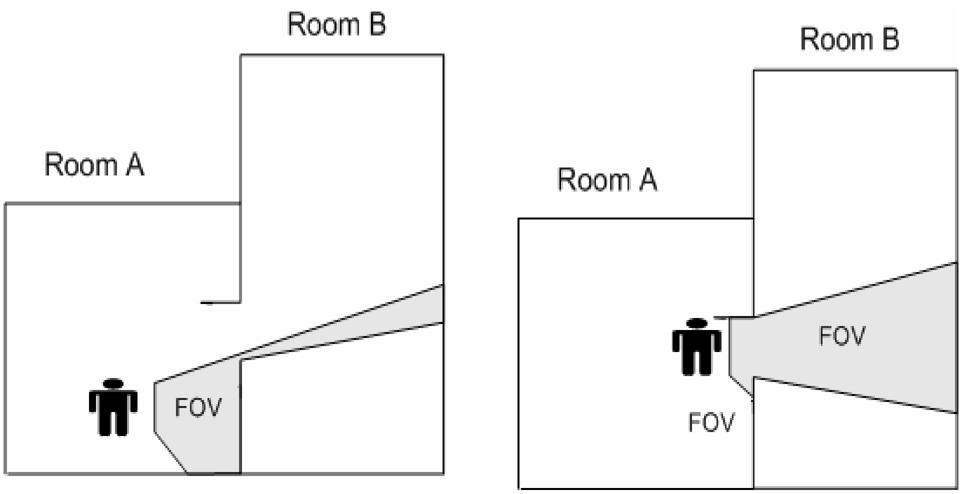


Fig. 40. (a) Shows the user's FOV in room A. (b) shows the FOV, mainly from room B and hence handover is considered.

An additional factor here is also the motion of the user. A stationary user can turn both their head and body in any direction. Therefore, small changes in position would definitely put them in partition A. However it would be safe to say that from a behavioural point of view, if the user is moving towards a portal and that portal increasingly encompass their FOV, then it is likely that they are actually moving through the portal and into partition B. Therefore, *perspective*, *location* and *movement* are factors in determining the need for handover.

6.3.5 Overlapping partitions

The partition system described so far operates with a strict partition boundary in a strict lattice structure as shown in figure 38. However, as previously discussed, this can cause some problems, especially when a person is standing at the boundaries. A more convenient solution is to have small overlapping partitions at these boundaries, which partly cover both cells. As an example, in the case of the person at the doorstep, an omni-directional camera located here can take over the image provision, as shown in figure 41. An omni-directional camera takes a 360 degree view of the world in the horizontal, and 180 degrees in the vertical. Image processing software is used to calculate a normal perspective image (Gaspar *et al.* 2002). These overlapping partitions are deployed when difficulties arise over camera orientation. Returning to our door border problem, users will normally move directly from one partition to the next, but sometimes stop to look into a room, or stop to talk with someone at that location. Overlapping partitions can then be deployed when the person's presence in a primary partition is uncertain for a short to prolonged period of time, or simply to provide more consistency when moving from one partition to another.

6.3.6 World construction – combing the partitions

Partitions are the essential means in which large dynamic scale worlds are constructed, the location and perspective of the user determines when the person moves from one partition to the next and overlapping partitions can assist in times of contention. Now, how the partitions are organised into a single system needs to be examined. For our user, their *perspective* and *location* within the system is essential. These therefore, become the basic building blocks. A third building block is the system *integrity*. By system integrity one means the degree to which the system reflects the reality on the ground. As an example, a building has a collection of rooms, each room containing a partition. The connection between these rooms is physical and an AR user can only move logically from one partition to the next as they walk through the building. However, with a TeleReality system, it is possible to arrange these rooms differently by moving the connections between the partitions around. Therefore, if a TeleReality user were to walk out from a room with an AR user, the AR user and TeleReality user could end up in entirely different locations.

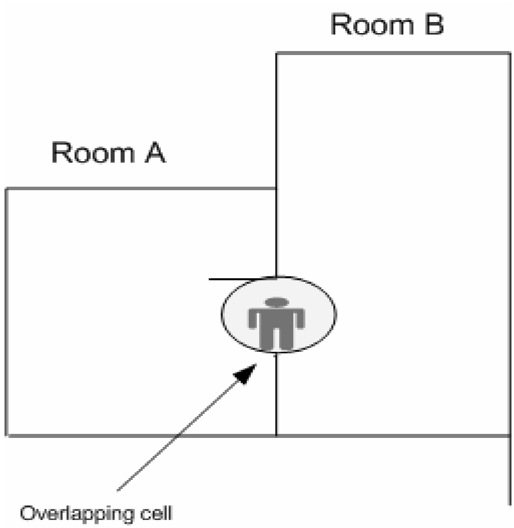


Fig. 41. An overlapping partition is used to solve problems of contention when moving from one partition to the next

6.3.7 Location

TeleReality systems are based on real world locations, so each partition covers a unique physical area in the real world. Real world locations are defined either by their geography e.g., longitude, latitude or symbolically e.g., an address (Torkin 2000). This differs from virtual environments, which normally use the Cartesian plane. Which system should a TeleReality system use?

The most common method for acquiring geographic location is to use the Global Positioning System (GPS). Differential GPS (DGPS) has an accuracy of 1-5 meters in the horizontal, while Wide Area Augmentation System provides accuracy of 3-5 meters in the horizontal and 3-7 meters in the vertical (GARMIN Corp. 2000). This generally works well for outdoor applications, however for indoor situations some difficulties arrive. First, more precise positioning is required as the rooms are normally less than the margin of error for GPS. Having the right co-ordinates for a partition is essential. The second problem with indoor locations is that even a small few meters error in the vertical position will place the cell on a different floor. A third issue with GPS in an indoor environment is that the signal attenuation and multipath fading makes it difficult for the GPS receiver to receive and decode the signal (Vittorini & Robinson 2003).For indoor systems, a secondary system is needed in combination with GPS, such as wireless systems (Van Diggelen & Abraham 2001) or an infrared network (Harter & Hopper 1994).

Symbolic locations systems use a unique address to represent each location within the system. As an example, an IP address represents the location of a single computer on the internet. This system is comprised by establishing a relationship between these symbolic address’s to form a single system. Leonhardt (1998) proposed combining the geographical and symbolic models into a single model. GPS is used to find the general location of objects, or in our case, partitions, while symbolic links would be used to define their exact geographical location.

While this system would work just fine, the reality is that each partition does not need a GPS position to define its location for the TeleReality user, the symbolic connection is enough to uniquely address each partition. This is also better as the address to enter a partition would be easier than entering the precise geographical location data of the partition. However, while supporting GPS is not mandatory, the AR user physically present in the system is likely to be using a geographical model, using GPS and an indoor tracking system. Therefore a remote TeleReality user and a local AR user are likely to be using different models to describe the same unique space. Added to this, the geographical location of the AR user will be needed to determine the handover of artefacts and will be helpful for a TeleReality user to navigate the environment. Hence, for completeness, each partition is defined by its symbolic address and geographical location.

6.3.8 Portals

Earlier, the need for handover from one partition to another was described. For the most part, these handovers occur at points that result from physical obstructions. For example, a room with four walls and one door has only one possible exit. Therefore, our partitions normally have only a few possible locations in which handover can occur. These locations for possible handover are called '*portals*'. Portals act as entry/exit points to the partitions.

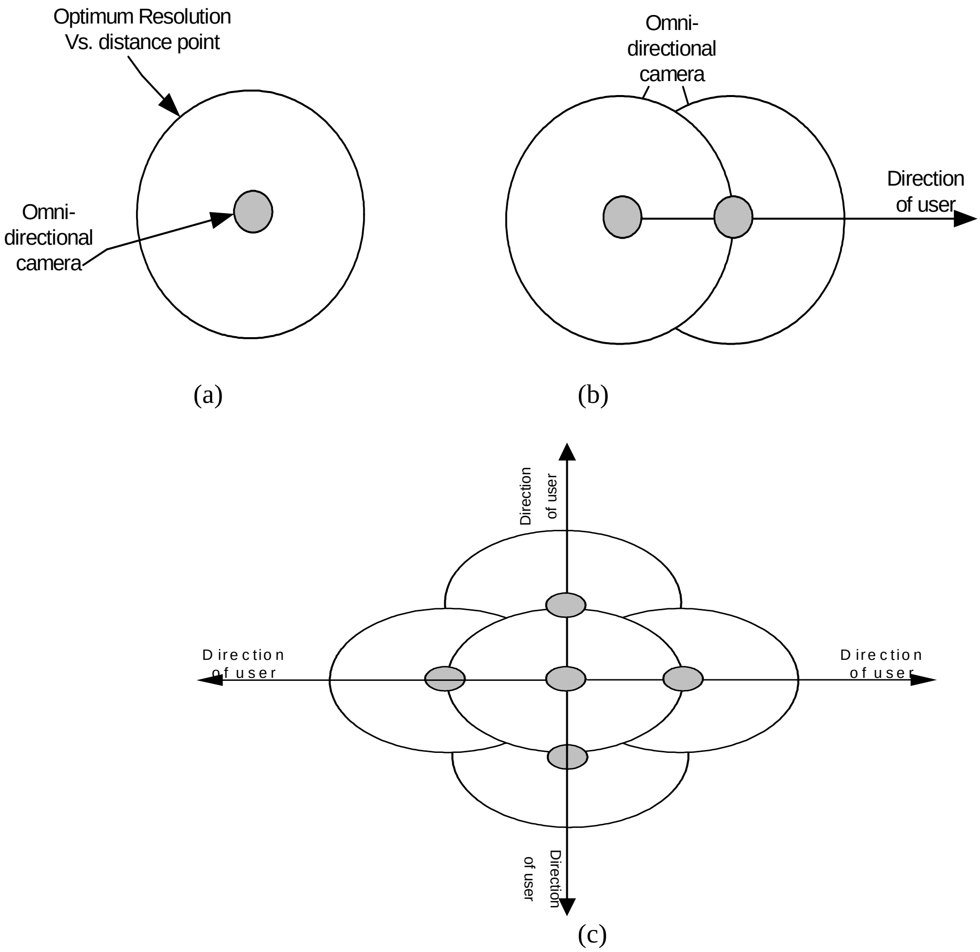


Fig. 42. (a) Shows the range of a single omni directional camera. (b) Shows the strategic placement of the second omni-directional camera to maintain satisfactory resolution. (c) Shows the general layout needed to maintain an even field in outdoor areas.

The advantage of portals is that they define a narrow location sensitive area in which handover can occur. Furthermore, they act as a two way conduit. Each portal connects two separate partitions only. Therefore, when one moves through a portal, the entity controlling handover is provided with the symbolic destination address by the portal itself. The portals from one partition to another must match in order to maintain fidelity and consistency, i.e., if portal A points to portal B, then portal B must point to portal A. It is only when one is close to a portal that handover can occur, therefore the definition of the portals location is necessary, but somewhat difficult to accomplish.

6.3.9 Large Outdoor Area's

Previously, it was stated that a person could leave one partition for another when they reached a point where the resolution is no longer satisfactory. This is based on determining the distance from the camera when the resolution deteriorates due to increasingly large inter-pixel distance values. Therefore, each camera will have some coverage area, as shown by the omni-directional camera in figure 42(a). Using Location based systems, it is possible to anticipate when this point is reached and order a new handover to the next relevant partition. However, this is not a straightforward task. In an indoor system, this is not a significant problem as the room size is usually well within this limit, there the problems are more related to the camera FOV and occlusion. In large spaces, especially outdoor environments or large warehouses, the difficulty is of a different nature. In figure 42 (b), the camera for the 2nd cell handover must be positioned at the edge of the limits from the first partition to maintain a continuing view. This creates a line along the central perspective of the cameras and can be used to have seamless transitions, in effect, a similar 'pathing' experience found in static systems. For multiple paths, the actual required configuration is shown in figure 42(c). This system would need to be maintained for each possible direction. For outdoor systems, multiple camera systems that ideally have high resolution and wide FOV's are preferred.

6.4 Movement

So far, the construction of the environment and the primary issues relating to the jumping from one partition to the next have been covered. Also, acquiring the 3D information of the system in order to navigate the environment has been discussed. Knowing that a system can support movement through the system, one must ask how that movement is conducted and what its effects are. For an AR user, movement through the environment is easy, they just walk around. For TeleReality users, the situation is different, walking around their actual environment while viewing a remote environment is difficult, if not dangerous. Generally, the TeleReality user will be required to adopt a stationary position while controlling his virtual avatar's movements in the remote environment. This control can be done by using a device like a joystick, keypad or mouse. The author has been unable to find any information pertaining to the effect of a person wearing a HMD and navigating a remote real world with a virtual avatar. Indeed for teleoperation, the problems usually revolve around gaining knowledge of the terrain and restrictions in the camera perspectives. But most of the effort is in actually controlling the proxy, avoiding hitting people's objects, as was the case with the remote controlled car experiment and the blimp experiment (Paulos & Canny 1997).

Some experience has been acquired from controlling virtual avatars in Virtual Environments. Movement is normally done by using a keyboard, joystick or mouse, although some researchers use treadmills (Pratt 1997), which is probably not the most practically widely available solution. Roy Ruddle concludes that using HMD's is a more efficient means than using a desktop system as subjects walked through the buildings more quickly and got a more accurate sense of relative straight-line distance (Ruddle *et al.* 1998, 1999). Others use illusions of self-motion, called '*vection*', to create a set of tracked actions that can give the user a feeling of movement and hence presence. Slater's research group had the user walk in place and compares the effects to a person using hand movements to signal forward movement (Slater *et al.* 1995). Fairchild his research team required the user to lean forward to acquire the same effect (Fairchild *et al.* 1993).

The effect these techniques will have for a TeleReality user is unknown. In general, a cognitive map of the area helps the user navigate and work in the environment and having as much peripheral vision as possible helps, i.e., wide FOV (Darken *et al.* 1998). Also, Witmer and Kline (1998) noted that users do tend to underestimate the perceived distance in VE's and the real world, but that underestimation is greatest in VE's, so it is not certain as to what the actual effect on TeleReality users will be, but it will probably be a combination of these research findings.

6.5 Summary

This chapter describes the components needed to build the constructs for the environmental sensors and environmental model constructs. The system makes use of '*visual scenery*' and '*visual objects*' to represent the visual field. The advantage of these is that it reduces the bandwidth for sending video, as that changes in the visual scenery, which accounts for most of the information and hence bandwidth, happen rarely and as a result only the scenery objects are updated in real-time. Depending upon the sophistication of the video codec's, only the visual object state information could be sent, further reducing the data bandwidth. The world can be represented by either a 2D or 3D visual system. 2D systems provide better bandwidth and simplicity over 3D systems, but at a cost of subjective presence. The world itself is of two types, a *static* world where the world information is temporarily stored and a *dynamic* world, where the images are displayed in real time. The world is subdivided into a collection of *partitions*, which are connected together by *portals*. Partitions are needed due as video cameras are constrained by phenomenon such as occlusion and degrading resolution over long distances. These partitions are linked together to form one continuous world. To maintain that continuous nature of the system, *handover* between each partition is performed, using parameters based on the users location, Field of View and movement. The location of the partitions is based on a geographical-symbolic model.

7 Privacy and Security

Providing privacy and security within a TeleReality system is a complex task that impacts heavily upon both the design and the constructs of the system. The TeleReality system is based on using cameras to capture information from the environment, either in real time or delayed time. This can create an Orwellian ‘Big Brother’ system, which is open to severe abuse if left unchecked. Additionally, the current reality is that the system of ad-hoc networks used in AR has security issues based on the lack of an independent trusted 3rd party certification bodies. This has an impact on TeleReality users, as part of their system will be based on ad-hoc technologies, while otherwise standard distributed security procedures would apply. This chapter examines these two issues, how to establish trust within the system and how to provide secure connections. The solutions will be both organisational and technological, and are summarised in table 12.

Table 12. Security issues raised from the use of camera systems, ad-hoc networks and their proposed solutions

Problem	Solution	Description
Invasive nature of Cameras	Organisation – Trust contract	A legal agreement outlining the users rights within the system
Establishing trust	Awareness	Make users aware of who or what inhabits the system with them
	Control	Provide control and power to the person in the camera environment
	Access	Restrict access rights of the visitor
	Integrity	Certify the identification of other users in the system
Security problems in ad-hoc networks	Network key, Ad-hoc key, Device profile	Only allow one person on an ad-hoc network access to the network. Use a network key to contact network, use separate ad-hoc keys to talk to other networks. Maintain a device profile in the network to check for unknown devices.

7.1 Invasive protection

Under any system which comprises of a set of cameras recording an environment in which people work or live and where access is provided to others known and unknown, issues are bound to arise concerning the invasive nature of the technology, and how it affects work and everyday activity. In 'the Truman show' the main actor was constantly monitored by cameras without his knowledge, the issue of invasiveness was not a problem as he was unaware of their presence. In what is referred to as 'Reality TV shows' such as 'Big Brother', the contestants volunteer to give up their privacy in the hope of winning a prize, fame or a more lucrative job opening. While the contestants are aware of the monitoring system, if not the position of all of the cameras, they have agreed to the set-up before hand, i.e., they knowingly and voluntarily inhabit the world in the knowledge that the general public will see their actions. In both of these cases, the constant monitoring was not an issue as the subjects were either totally unaware of them or consented to their presence. The critical issue is when people who are aware that the system is in operation and do not consent to the invasions of privacy that result from it.

In media spaces, which are systems set-up to create a large sense of awareness in the work environment, some workers were opposed to the scheme (Mantei *et al.* 1991). One of the principle fears was that their behaviour would be monitored. Experiments showed that it was not the big brother of the company who monitored their behaviour as that of their fellow co-workers (Fish *et al.* 1992). A typical use of monitoring was to check to see if a person was in their office in order to precede a conversation. The observed user was unable then to avoid the conversation, as they were known to be there at the time. The second use was the monitoring of a person's presence at his work place to check regular time keeping, which was not entirely accurate as they may be in other relevant work places. The third form of invasiveness was the ability for third parties to observe whom that user was talking to in their office at any particular point of time. These problems existed within a test environment, but one can only imagine that these problems could be magnified in a work environment that had stronger political rivalry or where workers were facing redundancies.

Media spaces set out to develop a system of total workplace awareness that was too invasive for people's comfort, to such an extent that it was in practice unacceptable and unworkable. Numerous researchers tried to limit this intrusive behaviour by greying out different participants (Hudson & Smith 1996, Lee *et al.* 1997, Zhao & Stasko 1998), but this also had the effect of working against what the media space concept was about, open shared awareness (Gaver *et al.* 1992) and little practical progress has been made upon this subject.

Most instances of a TeleReality system will involve periodic image updates and this could lead to similar problems as experienced by media systems. Before proceeding farther, it should be stated that the purpose of the media systems and TeleReality system are different. Media spaces tried to construct a shared awareness and workspace by monitoring multiple remote areas with cameras. TeleReality systems are meant to be highly functional, the person is in a partition and interacting with the system for a specific reason. It is this different purpose of the TeleReality system that will need to be exploited to provide a solution to the fears described by the media space systems.

7.2 Establishing Trust

What a TeleReality system needs to do is to propose a pre-agreed degree of trust between those who reside within the system. This gives us another main construct of the system, the '*Trust Contract*'. A trust contract is an agreement made between the actors in the environment concerning the nature of their presence in the environment. This agreement is based on an organisational contract, which can be supported by technology. The function of this agreement is to establish clear lines of control within the system. By control one means the ability to have the power to dictate whom is allowed to be present within the system and under what conditions. These conditions may vary, depending upon the TeleReality system in place. The following types of trust agreements may be proposed

- *Public* The most invasive form, everyone has uncontrolled access to the system
- *Partial* Some people are given more privacy than others are. For example, in a company, the workers may have highly restricted access within the system, while security personnel have total access.
- *Private* Each person within the system has control over who sees them or who can communicate with them.

In dealing with each of these trust relationships it may be advisable to look at the means in which they may be realised on an organisational level. Ideally, the more control a participant has over the system, the better they are able to ensure their own privacy. By this one means that access and control of the system should be divested to the actual users, both present and remote. It is this level of control that each relationship should determine before one agrees to be present within the system. The four constructs needed to provide this trust relationship in a TeleReality system are

- Awareness
- Control
- Access
- Integrity

7.2.1 Awareness

The concept of awareness covers the necessity for the user to know with whom and with what they share a space. In a real world situation, a person in a room can easily see every person or object within the room, so they are usually aware of the most important elements of information with which to complete their tasks. In TeleReality systems, some objects are entirely virtual. This means that they have the ability to hide their presence from other users while gaining information without the knowledge of the other participants in the environment. To make a user ‘aware’ of the space is to let them know who cohabits the space they are in, i.e., people and both virtual and real objects.

It is necessary for each participant in the system to be aware of who else is present, and the form of their presence, i.e., what capabilities and privileges that they have been granted. If there are two people located in a room and a third person enters as a virtual avatar, then all three participants should be aware of each other’s presence. If for example, one of the real actors initiates a communication with the virtual actor without the other actor being aware of this, the virtual actor may be able to listen in on a conversation without the knowledge of the other person. In such a three-person relationship, any one of the actors can listen in and observe behaviour, so it is essential that the other actors are aware that this is taking place so they can modify their behaviour if necessary.

This rule also applies to the use of a company to monitor their workforce, it should be necessary for the company to make that presence known to the workforce using avatars in a manner separate to that of the camera system. By making the users aware of that presence, one can try and replace the uncertainty that they might be observed with that of the certainty that they are being observed. By representing the observing parties, one offers a more substantive form than just the camera, where the users do not know for sure who is watching them and when. Awareness in a TeleReality system is simply presenting to each user the active parties and objects within the space they inhabit, even if they cannot make use of all of the objects or communicate with all of the people. By simply knowing who is around the users can modify their behaviour accordingly to maintain a degree of privacy.

7.2.2 Control

While awareness lets a user know who is present within their environment, so that they can modify their behaviour, control decides who should be present within the environment. The issue of control is to get agreement between the actors in a partition to share the space and interact. Control can be given to each individual person, the owner of the space or the owner of the system. The central aspect of control is getting an actor to agree on the presence of a TeleReality user. This is similar to the analogy of a telephone call. A typical scenario has a TeleReality user contacting an actor, requesting their presence within the system. The actor sees who is attempting to join them and can then decide if they agree to their presence or refuse it. In this case, the contacted actor determines whether a person is allowed to be represented in the environment. This is also not that unlike a person knocking on your office door, one can choose to answer it or ignore it. In this scenario it is the person in the room who decides who is present, thus they can have full confidence in the system because nothing takes place without their knowledge or consent. The second part of this contract is that the controlling person can also determine the end of the visitor’s presence, by terminating their privileges, i.e., the same as hanging up the phone.

This negotiation of point to point contacts is also an important aspect for audio communication. As mentioned earlier, one problem with multiple presences within the environment is the possibility of eaves dropping. Unlike video, it is more difficult to filter out extraneous voice. By using point-to-point negotiation, one can also establish a point-to-point audio link. This also ensures that everyone participating in a conversation has agreed to it and there is less possibility for eavesdropping.

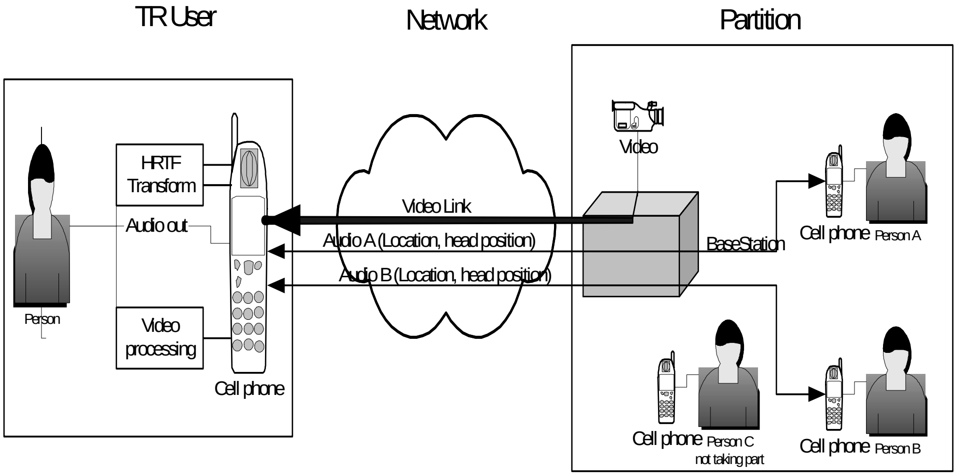


Fig. 43. A session is set up whereby the remote user establishes a separate point to point connection with each relevant user. Person C is excluded, but is aware that the remote user is present, just cannot hear what he says.

In practice our system will look somewhat like that shown in figure 43. The TeleReality user establishes a conference call by contacting two people in the partition, while ignoring the 3rd person. Both people accept the call and a 3-way communication ensues. The video stream is constructed and sent to the user. Two audio channels are established to the remote user, one for each of the participants in the conversation. The real actors establish only one audio link to the TeleReality user as they can hear what each other says in the room anyway. The audio information is transmitted in the links with information on each person's relative location and head position. These are then fed into a Head Related Transfer Function (HRTF) (Kapralos *et al.* 2003), that generates a 3D sound information profile for the TeleReality user and likewise for the real actors. The other person in the room is only able to eavesdrop on the conversations of those present in the room, and not the entire conversation. Another way to initiate communication is for the remote user to directly access a partition after obtaining the permission of the partition owner and while in the partition, the remote user can then initiate contact with people there by interacting with their avatars. Communication is established not by a pre-known address of the remote user, but by spontaneous interaction.

This need for control should highlight an important aspect of TeleReality systems. For a TeleReality user to interact with the environment, they must have the consent of the owner of that environment, to interact with other people, they must have the consent of that person. Control is always therefore placed in the power of the requested party on a negotiated contractual basis.

7.2.3 Access

Access works by determining the limits to the visitor's behaviour and activities in the system. As an example of access, a TeleReality user contacts a real person and is invited into their partition. Once in the partition, the TeleReality user can move to another room, which may not be what the person who invited them wanted. One may not want complete strangers exploring all aspects of the building you inhabit. Therefore, the real user would restrict their access to that specific partition upon accepting their presence.

To make the system work more smoothly, each partition should be allocated a security clearance code. The people who inhabit those partitions will have clearance for their own partition and for other common areas in the building. These people may grant temporary visiting rights, with the appropriate security clearance for the duration of the remote user's presence in the system. This is cancelled when the presence ends. In figure 44, Mary has three clearance levels. One is just for her room, the second is for her co-workers rooms and the third is for the general visiting areas (e.g., lobby, corridors, meeting rooms, and cafeteria). When a person requests a presence, Mary, can give them access to her own room. If she leaves the room and wishes the visitor to come with her, then she can 'invite' them, i.e., give them a temporary clearance for those areas. She cannot give the remote user access to other people's rooms, which is up to each of those occupants to decide.

A similar issue is also at work with the interactive objects within the environment. The optimum way to manage these objects is to collect a list of objects within various security groups, e.g., general information could be combined into one security group as almost anyone can have them. Each person is assigned a number of security groups depending upon their needs. They may also be given rights to elements of a group on an element by element basis. As well as having access to different groups, some user may be permitted to grant temporary security clearances to others. This is to allow visitors to use some artefacts they do not have permission to use in order to carry out a shared work activity. As one example, a visitor in a meeting room, may wish to use a projector, but would not automatically have the rights to use it. The others in the room would have the rights, and can assign the visitor the rights to use it. This would disappear whenever the grantee of the rights decides, or when their presence is terminated.

7.2.4 Integrity

Integrity is essentially the component of the trust relationship that focuses on guaranteeing that the user is not tricked or fooled by means of deception. One of the main aspects of establishing trust between the actors in the system is that they should be able to establish a relationship based on a level of trust that they are both dealing with each other in good faith and that the actor they are dealing with is exactly who they say they are. Therefore it is necessary to determine and confirm the identity of the other actors and that they are communicating with. It should also cover the protection from nuisance attacks such as spam and resource draining attacks also.

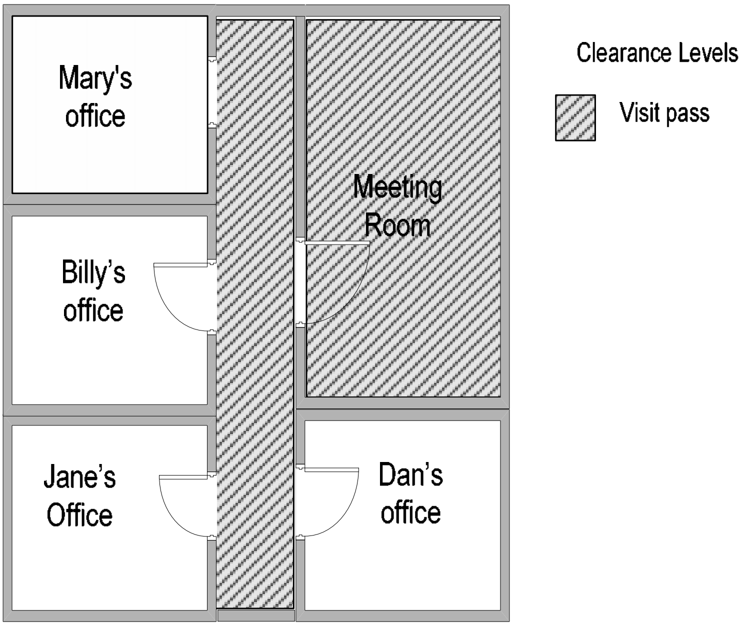


Fig. 44. Office space. Mary has the right to give others access to her room and public area's, but not to other peoples rooms.

In cases where two AR users are communicating, trust can not only be established by technological means, such as biometric certification, but also by the common locality of the users. It is not a strictly anonymous situation. For TeleReality users in the system, there is the possibility for deception, as the video image could be doctored or tampered with. Earlier it was mentioned that connections should be formed on a telephone based model. With current mobile telephony models, each person is given an id (their telephone no.). The phone companies can provide services such as ‘call identification’, and act as a form of 3rd party independent certification, as the phone number is linked to a name in their database. Of course, a person may phone and say give a different name to that in the database, but the use of their number is enough to determine their real name. Deception is still possible however. Generally parallel methods are used to associate the person identity with the phone id., e.g., face to face meetings, introduction by trusted 3rd parties.

Integrity of the system also means guaranteeing to the user, that the communication and technologies used to support the organisational security systems is safe and secure.

Essentially, this means that when a questionable communication is taking place, e.g., between people with two different sets of security standards, that the user is at least informed of the risks that they are taking. A more detailed analysis of the problems involved is discussed later in this chapter

7.2.5 Importance of trust relationship contracts

In an example of a workplace where the workers are being monitored by the owners, by being made aware of that presence, the workforce may decide that it is unhappy with the situation and distrusts the owners purpose for the system. The workers may then attempt to negotiate a deal establishing a degree of trust with the owners, that may result in the system been removed, or prohibited for use in firing, promoting, directing or discriminating against the workforce. In essence, the workforce attempts to negotiate a system with parameters and guidelines which it can have a relationship of trust with. Trust between users is normally established based on a history of past contacts and easier to maintain when the relationship is a real face to face meeting. One trusts the other actor, not by technological means but by a parallel relationship. Trust, in this manner, is more of a subjective process than by an absolute objective measure.

7.3 Security

Security is concerned with providing the practical physical and technological security concerns for the TeleReality system. So far, the organisational problems have been discussed, but now it is necessary to understand and provide the physical level of security. The TeleReality system is based on a complex, multi-user telecommunication system. This communication system includes fixed, mobile and ad-hoc networks. Of these elements, the ad-hoc network element has a number of inherent weaknesses, e.g., a lack of 3rd party certification (Hickey 2002).

7.3.1 Addressing the security weakness

Trying to cope with the weakness of ad-hoc security in the TeleReality system is a difficult problem. However, ad-hoc security systems are developed based on the assumption that there is no 3rd party certification system possible as there is no guaranteed connection to a constant and known network. However, in a TeleReality system, there is a 3rd party element available, as part of the network is a fixed network. Therefore, when an ad-hoc user is not in a TeleReality system, they may use ad-hoc security procedures, while within a TeleReality system, they should be forced to use the 3rd party certification through the fixed network.

The procedure that should be followed as a result of this situation is to separate the system into two parts, one dealing with the network and the other dealing with other ad-hoc networks. These two parts are kept separate. To achieve this, a number of steps have to be taken.

– *Network Key.* The connection between the fixed and ad-hoc network should be secured by a *network key*. This key works from the personal user sphere, meaning that once the connection is made, all of the artefacts within the personal user sphere are certified and operate under the security of encryption on the fixed network link. This is separate to the encryption used on the ad-hoc radio links.

– *Ad-Hoc Key*. An ad-hoc key is used to operate with other ad-hoc elements. The fixed network can be used to validate and certify the other user's ad-hoc key.

– *Integrity*. The user guarantees the integrity of the artefacts in his personal user sphere.

The latter point places the burden of responsibility on the ad-hoc user to ensure that their network is safe and no one else is a part of it. This is not always possible though, as 3rd party eavesdropping is still possible. Two ways, when combined, may counter this problem. One is to have a device network profile of the user's personal ad-hoc network stored within the fixed network. In this scheme, when a new device is added, its profile is stored in a safe place on the network, either in the user account or in a 3rd party site. When a connection is made, the user's current device network profile is checked against their stored profile. Unknown devices in the network are identified. Using these schemes, a 3rd party in the example in figure 45 can only impersonate the device, but cannot get or use the network identity to get access to the fixed network.

Another element of a TeleReality system that helps address the security weakness is that the users have access to a visual picture of the other users through a real time video link. This can help in the identification of the other users in the system either by familiarity, or checking their video image against a stored picture of the user held by an independent 3rd party.

7.3.2 Artefact security

So far, both trust and device security has been established. Lastly, comes the need for artefact security. Each artefact has the capacity to act entirely independently, or *migrate* locations. Therefore its location can always vary. Each artefact also has the ability to move part of its code to a different location, e.g., in a code download to execute an artefact interface. The artefacts work by sending information that executes code on the machine that requested it. In many cases, this code may not actually be requested by the receiver, but as part of an advertising code. This code e.g., shows a 3D object that advertises some object. By downloading executable code, one risks receiving viruses or

Trojan horses. Without any established software environment standard, the need for downloadable code is unavoidable for systems.

Virus protection is essential with a TeleReality system. These types of virus may be malign, deleting information or act as denial of service attacks. Often there will be occasions where the attacks are not intended to be malign, but have consequences on the system by absorbing precious resources such as memory, battery power and processing power.

Artefact access: Each artefact has an access profile. These profiles determine who has access rights and the conditions upon which that access are granted. Public access means that anyone who wishes to use the artefact may access it. Protected access means that the access is limited to groups or individuals. As stated earlier, the access profile can be changed dynamically by the owners of the artefact, but only for a finite period of time where necessary. Therefore, when each service is accessed, a record is kept of the service. In a client server relationship, this security profile can ensure that the connection is terminated when permissions have been revoked or other conditions are met.

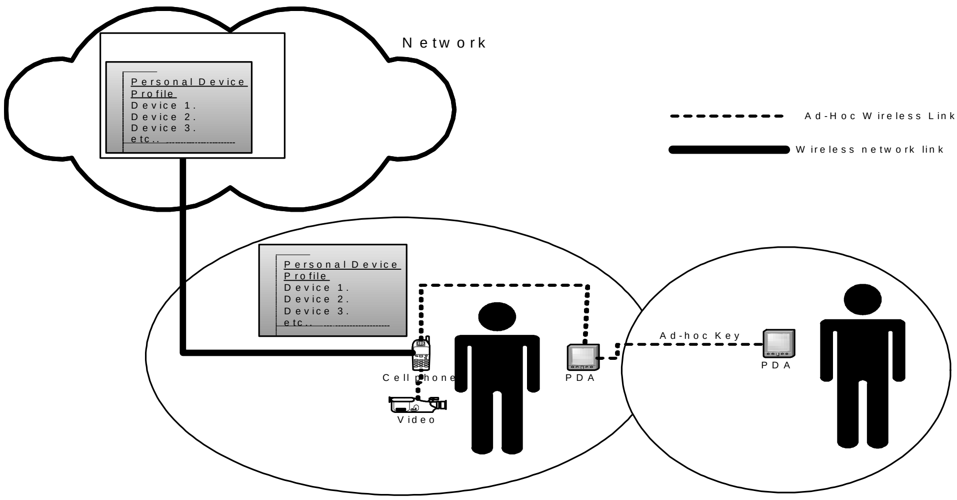


Fig. 45. The user with an ad-hoc network keeps a copy of his device configuration in the network. He also uses separate secret keys for accessing the network and for other ad-hoc users.

7.4 Summary

This chapter described some of the problems related to establishing privacy and security with a TeleReality environment. The basic problem focuses on the fear of a big brother system. The basic means for establishing privacy is to negotiate a *Trust Contract* between the participants within the system, which is an organisational agreement so that everyone is aware of what is allowed and not. Privacy can also be helped by requiring that the users be *aware* of who occupies their environment as this ensures that they know who is there, and hence they can modify their behaviour. Those being seen by the camera should be given *control* when and how someone enters the environment, by using a paradigm such as a telephone. *Access* rights must be established for the system, the partitions and the objects in the environment and visitors can be assigned temporary access rights, as well as establishing *integrity* in the system to give the participants confidence in the system. Connections are established by '*point-to point negotiation*'. Ad-hoc networks are the weakest security link within the system. To counter this, each user in the system with an ad-hoc network must establish a link to the fixed network to access 3rd Party certification services. Each user maintains a *device network profile*, to detect suspicious network elements. A *network key* is used to establish a connection with the network, and a separate *ad-hoc key* is used to establish communication with other ad-hoc networks.

8 Visual Cell system

The visual cell model supports an enhanced telepresence system by implementing the constructs discussed so far. The model is first explained and then it is shown how the elements of the model are mapped onto the constructs.

8.1 Visual Cell Model

The visual cell model is one potential solution that could be used to implement a TeleReality system. This model is based on similar principles to that of mobile cellular telephony networks. The basic premise is to create small areas called a *visual cell*. A visual cell can act as a totally self-sustained TeleReality environment. The visual cell model is defined as '*A limited space containing an image capturing device, where the space contains interactive real/virtual objects. Visual cells can be combined in a matrix to form one space that can be navigated and interacted with*'

This definition maps the multisphere model described in section 5.3.1 on top of the partition system so that each partition corresponds to one immediate area sphere. The model also contains all of the TeleReality constructs identified in Section 4.3 and the impact they are described individually in the following sections..

8.1.1 Environment Model

The environment model construct is implemented as a 2D/3D dynamic system that is based on visual information received from the environment sensors, i.e., video footage. The environment model then streams the visual scenery to the end user. The model is created by grouping cameras (environment sensors) into different visual cells, which is the equivalent of a partition in the world system. To create a larger area, *visual cells* are connected to each other into a lattice type structure in the environment model. The model then consists of a large number of visual cells that are combined together to offer a seamless environment to the users. Different techniques are used to perform handover, but teleport and location based handover methods are the easiest to produce. This is because the initial systems are likely to be 2D systems, with highly fixed camera paths. The location system is based on a GPS/symbolic address system, with each visual cell assigned a unique IP-address and port number. Portals point to the next cell address that the current cell is connected to.

Artefacts in the environment model must be obtained from the various actors within the environment. These artefacts are organised geographically within the environment model and is based on the multisphere model. Artefacts located in the immediate local area sphere are owned by individual visual cells. The system also provides a more general set of artefacts that span a number of cells. These include all general artefacts and migrating artefacts.

During the cell handover, it is important that not only is a continuous image maintained, but also that artefacts are handed over from one place to the next. That is, the challenge for a visual cell system is to ensure that the virtual proxy can move from one cell to another without losing visual information. During a cell handover, the virtual proxy must receive all of the information from the new cell, while discontinuing the artefacts from the last cell.

8.1.2 Environment sensors

The environment sensors consist of an array of fixed cameras arranged into cells, which record visual information that is then presented to the environment model for processing. The sophistication of the cameras can vary, but ideally stereoscopic cameras are used for 3D image capture. The model can also supports depth information sensors which are embedded in the video frames.

8.1.3 Security

Security access to various artefacts is restricted by access rights to individual cells. Each artefact can also be given different access rights. The system can include access to 3^{rd} party authentication and certification artefacts.

8.1.4 MR Environment

The MR environment obtains the video streaming information it needs by querying the environment model with its position and orientation. Based on this information, it receives a video stream as well as a list of artefacts and their position. These are then rendered, in addition to artefacts from the users own personal user sphere, for the user's viewing and interaction.

8.1.5 Information storage

The information storage holds the users own artefacts, as well as any information they receive as a consequence of interaction. Other models may support the use of information stored across a wide selection of devices, and perhaps these devices can offer parallel services to the user. However, the minimum requirement is that there be one information source to support the system and this particular model represents only one device holding that information.

8.1.6 Communication elements

The communication elements are a combination of fixed networks, primarily Ethernet/Internet systems, plus ad-hoc technologies such as Bluetooth. In the system then, the AR actor connects to the system using an ad-hoc connection, while the TeleReality actor uses the cells ad-hoc communication basestation to talk to the AR user. This basestation acts as the communication element for the virtual proxy during all interaction and communication carried out by the remote TeleReality actor while in that cell. The ad-hoc radio element that represents the system in the visual cell must be able to cover the same area that the camera system covers. What this means, is that in the visual cell system, there is one radio connection point for each partition. This is not necessarily so for other models where one radio link could cover many partitions or one partition may have many radio elements. In practice, every actor in the visual cell model will utilise an ad-hoc network to connect and communicate with the system. The connection itself is still based on a single autonomous link, with no guest users allowed on that person's network. This is an arbitrary restriction on the system. The other effect of this is to say that a person may not connect to the system by making a connection to a radio element in a visual cell they do not inhabit.

8.2 Visual Cell System architecture

The visual cell system architecture consists of a number of integral components. These components are shown in figure 46. Figure 46 shows a two visual cell system, each containing a VC base station. Each room is considered to be a visual cell and they are independent of each other. The VC server is responsible for combining both cells into a larger navigable space. The service database keeps a list of all of the artefacts available within the system. Copies of these artefacts are sent to each base station when needed. The TeleReality user connects to the system by using an external network to access the gateway and hence the system. The TeleReality user is represented in the system by a virtual avatar. The AR user communicates directly with the VC base station using a short range radio link. Objects in the room that also contain interactive artefacts, register with the system, also through a short range radio link. A more detailed description of the system components follows.

8.2.1 VC Base Station

The basic element in these visual cells is a VC base station. Each visual cell has a single base station. The VC base station acts as the main control point for the cell. The base station manages the image capturing devices responsible for capturing the visual information from the room and transferring that image information to any remote users. As discussed in Chapter 5, the form the image capturing device can take may vary, from a single camera to a complex sea of cameras system. These images are processed into a visual scenery object and they undergo a compression algorithm and encryption before been sent to the remote user.

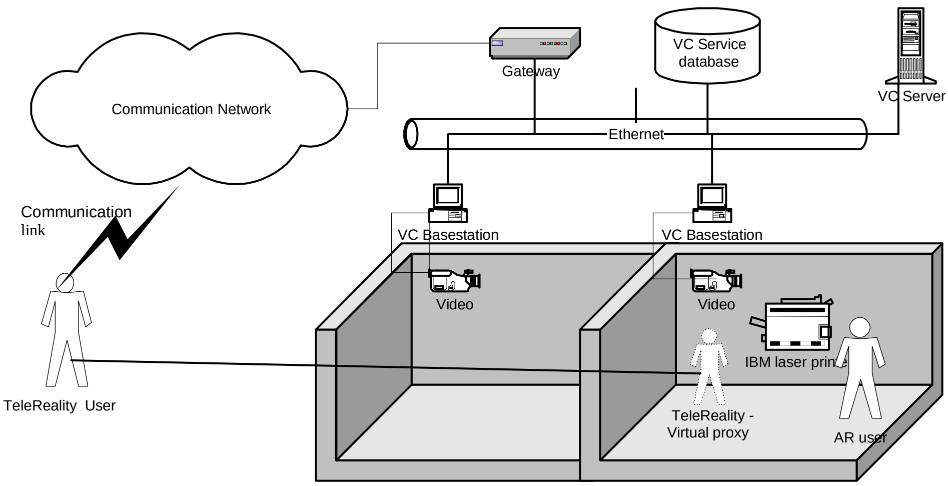


Fig. 46. A visual cell system with two cells, a visiting TeleReality avatar, an AR user, and a virtual object representing a printer

The VC base station also acts as an ad-hoc network node and hence contains a radio transmitter/receiver. By acting as an ad-hoc network node, the VC base station can form connections with other elements in an ad-hoc network. This ad-hoc network forms the basic communication means by which artefacts can be discovered and transactions performed. The VC base station also contains a communication link, usually in the form of a high speed data link, to other elements in the visual cell system and externally to the remote user. The VC Base station acts in many ways as an agent for the remote user. The remote user does not have a direct ad-hoc network connection to other members of the environment, but rather the VC base station acts as this ad-hoc node.

The other function of the VC base station is to act as a point of access to artefacts outside the individual cell. It is through the base station, that real actors may access local area artefacts, i.e., artefacts that do not specifically belong to a single visual cell but to the whole area.

The VC basestation also allows access to trusted 3rd party certification authorities. The base station then can have multiple roles as a central co-ordinating point for the system. Each base station is given a unique address, which is used to identify its location and position within the system and the world.

8.2.2 Visual Cell server

The visual cell server has a number of functions. The chief purpose of the VC Server is to act as the central point for ensuring co-ordination and communication between multiple visual cells. This is most significant when one actor within the system, e.g., a TeleReality, AR or cyber actor, moves from one visual cell to another.

When the TeleReality actor moves from one cell to another, the server must first detect the need for the transition. Once the VC Server has determined that a cell transition is needed, it executes a handover between cells. Depending upon the sophistication of the system, the visual information from the new cell must be communicated to the TeleReality user. The more sophisticated the visual processing of the system is, the more smoothly this process is. Ideally, the user should be unaware that any change of area has occurred, but this is not easily achieved within the visual field. Second, the remote user must be assigned to the new communication network node represented by the new visual cell base station and the old network information should be closed. The third action is that, in conjunction with the services database, access to the old artefacts from the old visual cell should be removed and a new list of available artefacts from the new list should be established.

For AR users of the system, there is little need to worry about image transfer, but it will be necessary to remove old artefacts and perform a new artefacts discovery procedure to find out what new artefacts are available. Hence, the server still needs to enact handover, but only on the artefacts and communication levels for the real user. Of course, if a communication link had been established to a remote user in the previous cell, this is also lost as well as the visual avatar. The final act is to update other participants of the new status of the user within the system.

Occasionally artefacts move from one cell to another. In this case the behaviour is similar to that of the AR user, its old artefacts connections need to be removed and new artefacts acquired. The link to the network through the new VC base station is also changed.

8.2.3 Service database

The service database operates in conjunction with the visual server. It contains a record of each artefact that exists in the system. The record for each artefact contains information about which cells each artefact is found in, as well as access and security rights. The service database also contains, in conjunction with the VC server and base station, the process by which users discover new artefacts. Various Service Discovery Protocols (SDP) are in existence and can be used with the VC system, e.g., Jini, Bluetooth SDP (Avancha *et al.* 2002).

8.2.4 Gateway

The gateway is the access point of the system to external users. The gateway acts as a go between and conduit point for all of the information, visual, audio and artefacts orientated to and from the external user. The gateway is also the first point of security in the system, acting as a firewall, only permitting authorised external people to initiate access to the system. The gateway can also act to limit the exporting of unauthorised information from the system, to ensure that sensitive information or artefacts are not being stolen or used.

8.3 Location of constructs within visual cell architecture

Figure 46 shows the system architecture for the visual cell model but does not make clear where the constructs are located in the visual cell model, so to avoid potential confusion, the general relation of the constructs to the architecture in this particular model needs to be explained. Figure 47 shows how the main virtual proxy constructs are mapped to the different nodes in the visual cell. The VC model is used to primarily describe the support of the virtual proxy and hence the internal implementation of constructs for the TeleReality, AR and cyber actors is not covered here.

The cameras in each room operate as the environment sensors, providing information to the VC server. The VC server, using the sensor information from the basestation, and the service database operate together to build the environment model. The environment sensors are not responsible for locating artefacts from the various actors, these are communicated directly to the server and service database by ad-hoc radio elements, i.e., the various actor and proxy communication elements. The TeleReality user receives the information locally and renders the MR environment for viewing with a HMD. The proxy CE is divided in functionality between the VC Basestation and the gateway, with the gateway providing a continuous link to the TeleReality actor and the VC basestation providing local radio communication to the AR actors present. A more detailed breakdown of the construct components and how they are realised in this model is shown in table 13.

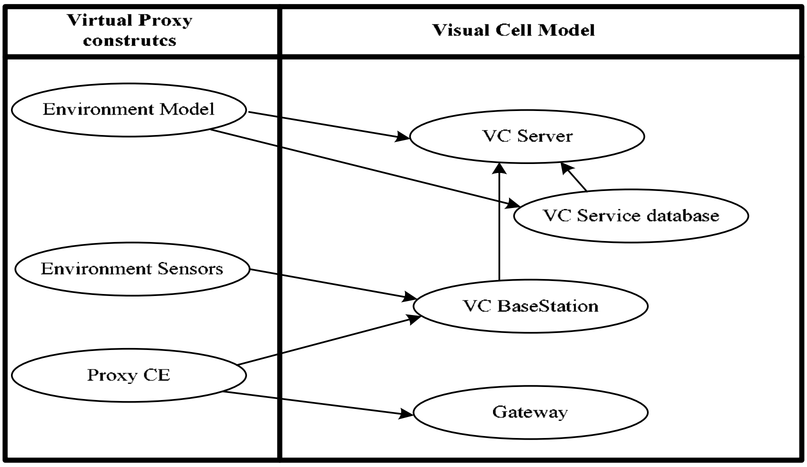


Fig. 47. Mapping of Virtual Proxy constructs to the Visual Cell architecture. The model does not include the different actors. Environment sensors are handled by the VC basestation, as is the proxy CE. Environment model is constructed in the VC server by utilising information from the service database and sensors. Proxy CE is split between the gateway and the radio connection in the VC basestation.

Table 13. Shows the list of constructs used in the visual cell system and the options taken.

TeleReality Construct components	Proposed Solutions
Visual Scenery	Rectangular Video
Visual Object	3D Graphical objects - Java3D/OpenGL
Presentation	2D (3D systems not available yet)
World Structure	Dynamic
Partition	Visual Cells
Portal	3D artefact
Handover	Forward movement, Portal interaction
Location System	Symbolic-geographical (IP-address, port no and GPS).
Transactions	Gestures, keyboard input, TCP/IP, Bluetooth
Artefacts	Applications, migration and replication supported
Personal User Sphere	AR/TeleReality actor
Immediate Area Sphere	Visual Cell Basestation
Local area sphere	Visual Cell Server
Global network sphere	Internet, World Wide Web.
Trust Contracts	Personal
Awareness	Total
Control	Called AR actor
Access	Varied
Integrity	3rd part certification
Ad-hoc security	Bluetooth security, WEP

8.4 Summary

This chapter introduced an example of a model using TeleReality constructs called the Visual Cell system. The visual cell system primarily focuses on the virtual proxy element, as the actors are established by the use of AR constructs. The visual cell model is made up of a group of visual cells, that are equivalent to partitions, and these are connected together. The main architectural components include the VC base station, which is responsible for capturing information in the environment, as well as acting as an ad-hoc proxy element to the environment, and for the TeleReality actor. The VC Service database contains the details of all of the artefacts in use in the system. The VC server is responsible for linking all of the visual cells together to support handover and the organisation of artefacts in the multisphere model structure. The gateway acts as the communication element that connects the VC Server to the remote TeleReality user. These architectural components are mapped in some ways to the TeleReality constructs. The VC Basestation maps to the *environment sensors.* The *virtual proxy CE* can be found in two places, the gateway and VC Basestation, as it needs both to complete the communication link to the AR actors. The VC Server, in conjunction with the VC Service database, work together to build the environment model. A description of how the construct components are implemented is provided in table 13.

9 Construct Evaluation through experiment

This chapter evaluates the constructs proposed for a TeleReality system by describing the results of two experimental prototype systems used to enhance telepresence. The constructs represent essential components that are needed for these systems. The evaluation will also describe the logical identification of the constructs and their cohesiveness.

The evaluation of the proposed constructs underwent two iterations, as described in the research method, figure 2. The first prototype implemented what is called the 'Mobile VR Meeting' (Hickey 1999). This implementation examined the early constructs of an enhanced telepresence system, and lacked a number of key constructs. Based on these results, an iteration was performed and a second prototype called the 'Visual Cell Model' (Hickey *et al.* 2000) was proposed and which uses the constructs outlined in this thesis. A partial implementation of this VC model was built, but did not include support for an AR actor or security. The effects of the constructs on these systems are evaluated.

9.1 The Mobile VR meeting Experiment

It was anticipated that third generation Mobile phone systems, such as UMTS (Munro *et al.* 1998), would offer data bandwidths of up to 2 Mb/s for indoor and 384 kb/s for outdoors networks (Berruto *et al.* 1998). This compared favourably to existing GSM Phase 2+ systems, such as General Packet Radio Service (GPRS), which offer in theory 170 kb/s (Brasche & Walke 1997). The increased bandwidth meant that there was the possibility of realising heavy data downloading from the Internet, as well as support for good quality video. Research interest started to focus on the possibility of applying VR technologies to mobile systems. There were a number of reasons for expecting that VR technologies, primarily AR based, could be used with mobile phone systems. First, it was expected that HMDs the size of sunglasses could be available in the early part of the 21st century (MacIntyre & Feiner 1996, Starner *et al.* 1997). The availability of such devices shown in Figure 48, would be the key enabling technology for VR mobile services, and work commenced on developing the technological know how and determining what kind of services, interfaces and additional requirements would be required in projects such as CyPhone (Pulli *et al.* 1998). The second reason why such HMD's would be of major importance to the mobile sector is the thorny question of the problem that is mainly known as the small user interface problem (Kuutti 1999). With 2nd generation phones, the design criteria focused on building the smallest phone possible, with the longest talk-time possible. The advent of the Internet boom and 3G services with their high multimedia content requirements have changed that paradigm. What was needed now, were larger screens to showcase these new data services, while still keeping the device as small as possible. HMD's could potentially erase this problem by separating the display from the device, giving a larger display resolution than is possible with smaller phones. The essential premise is that small lightweight HMD's have the potential to revolutionise the mobile communication industry.

(a)

(b)

Fig. 48. (a) An example of a CyPhone with stereoscopic display. (b) A proposed future version of the CyPhone with support for a HMD.

The PAULA project intended to study and anticipate the consequences of this change (PAULA 2003). One part focused on the effect of the user in his local environment using AR, while the other part focused on the mobile user in a remote location, using telepresence. The author concluded at an early stage that to support a sense of presence in the remote environment for the mobile user, one would have to provide him with a freedom of view, i.e., the ability to explore or navigate the remote environment. The author also wanted a multi-user system, and the greatest limitation to this was the nature of the physical proxy itself. This experiment then tested out the importance of being able to view multiple viewpoints with a remote environment and what were the consequences on behaviour and the requirements for the needed technologies.

9.1.1 Experimental Rationale

The rationale of this experiment was to examine the possible use of VR technology for a mobile user who wishes to take part in a meeting that is been held at a different location.

The concept is based on a mobile user who is in transit from one place to another. During this time, he is expected to take part in a meeting at another location. The normal method, by which this would happen, is for the mobile user to establish a phone conference, with the mobile user's presence being restricted to a voice only connection. Assuming that future mobile users may have the use of a HMD, it is conceivable that the mobile user could take part in this meeting by establishing a one way videoconferencing connection.

This experiment led to the first attempt to create a virtual proxy. To create a virtual proxy effect, an omni directional camera was introduced as the camera in the meeting room. An omni directional camera takes a 360 degree panoramic view of the surrounding area. Special software is then utilised to convert that image into a perspective image. By attaching a head tracking system to the mobile user, the software could determine the viewing perspective of the mobile user and calculate a new perspective image each time the head moved. This would create the effect of been able to look around the room as one moved ones head, providing the user with a limited ability to navigate the environment. This should help to increase the sense of presence within the environment. The additional advantage of this system is that the same panoramic image is sent to multiple mobile users taking part, and each mobile user calculates the appropriate perspective image based on their state within the system. The purpose of the experiment then was to test out these theories to see what problems would arise and how beneficial the system might be.

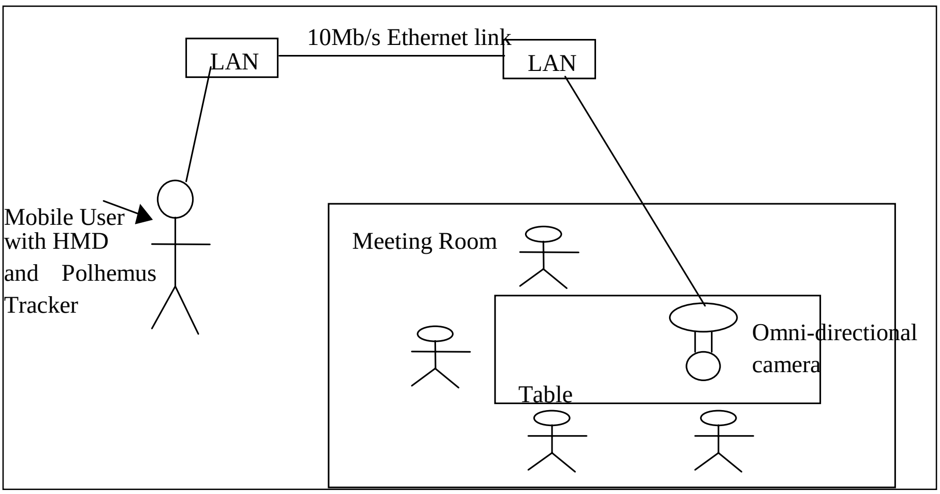


Fig. 49. Experimental Mobile Meeting set-up.

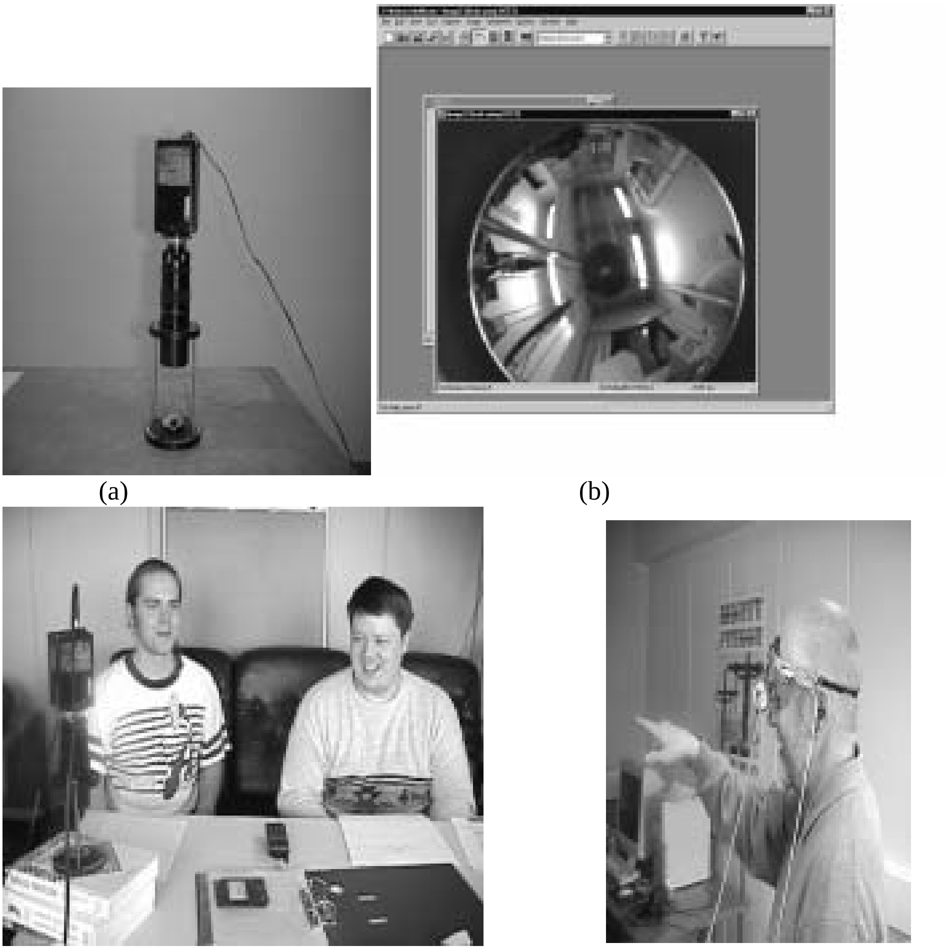
9.1.2 System set-up

The test was carried out within the University of Oulu, Computer Engineering laboratory and the set-up is shown in figure 49. The mobile phone user was located within one room, while the other participants in the project were located in a separate room. The rooms were some 20 meters apart and the room doors were closed to prevent any sound or visual cross pollution. The omni-directional camera was placed in the meeting room. It was located on the meeting table, at the edge of the table, to best try and simulate the position that someone would take if they were sitting around a table. The omni-directional camera was a CycloVision ParaCamera which is connected to a Matrox Meteor II frame grabber that continuously captures a constant stream of video images. The CycloVision ParaCamera shown in figure 50(a), is comprised of the Panasonic colour digital camera GP-KR222 and an optical system comprising of special lens and a parabolic mirror that captures a 360-degree view of the surrounding area, as shown in figure 50(b). The Field of View of the ParaCamera is 360 degrees horizontally and 180 degrees vertically. This captured image is called an ‘obstructed image’, as the image is partially obstructed by the camera itself and the metal holding the mirror. Each recorded image is called a ParaFrame and shows the entire 360-degree in one frame. This is difficult to interpret with the normal eye.

Additional software is required to convert the ParaFrame into a normal perspective view. This perspective view has a narrower FOV, more appropriate for a normal camera system, in this case 60 degrees horizontally. To change the perspective image, sensory information from the Polhemus tracker or mouse is used. CycloVision provides ParaPlayer software to convert the 360-degree video stream into a perspective frame. However this software did not suit our purposes, as it did not support a full screen mode, which is essential for use with the HMD. Therefore new proprietary software was developed to perform the perspective transforms in full screen mode. One side effect of converting the images to full screen mode is that it has a degrading effect on the image resolution.

The communication link was a 10 Mb/s Ethernet link. The communication software was written in C++ and used the UDP protocol stack for data transmission. The panoramic-to-perspective algorithm in the mobile user's computer was written in C for faster performance. A voice connection was established over a normal phone line using a meeting phone system in the meeting room. There was no data compression algorithms used over the transmission lines, which severely affected the data rate of the system.

The mobile user was given a Sony Glasstron HMD, with a screen resolution of 640*480 and supports a RGB colour scheme. The HMD was connected to a computer. A Polhemus tracker is used to monitor the users exact head position. The head position co-ordinates are transmitted to the computer where they are used to calculate the perspective view displayed on the HMD.



(a) (b) (c) (d)

Fig. 50. (a) ParaCamera (b) 360 Degree ParaFrame (c) Meeting room with ParaCamera in the table corner, (d) Remote participant with HMD and Polhemus Tracker.

There were 2-5 people located in the meeting room shown in figure 50(c). The meeting room itself was quite small. The mobile user was asked to stand for the experiment and unable to move around as shown in figure 50(d). This lack of movement was a necessary constraint due to the short range of the Polhemus tracking system and also due to the cables attached from the HMD to the computer. The meeting lasted for 3-4 hours. Different meeting attendants switched places with the mobile user in order to develop and contribute opinions on the system.

The ParaCamera was placed on the corner of the table. Originally it had been intended to invert the camera and suspend it from the ceiling, i.e., with the parabolic lens facing the floor. However, finding the optimal height to suspend the camera was problematic. The optimal distance for viewing participants sitting down had the side effect of cutting off their bodies if they stood up. Manually changing the height was not considered a good option, as it would necessitate constant fidgeting with the camera position. Consequently, the camera was placed with the parabolic lens lying on the table as shown in figure 50(c). The camera has the effect of being the spatial representation of the mobile user in the meeting room. Initially, it had been intended to place it in the centre of the table, but to provide a more natural feeling, it now approximates the position a person would take were they physically present at the meeting.

9.1.3 Experiment Results

The intent of this experiment was to examine a multi-user telepresence system supporting presence. The focus then was to see how the remote users reacted when using the system. Each person who tried the system liked the ability to look around the remote world location. The concept was not alien to them, and they used the system as was expected. When a person in the meeting room was talking, the remote user turned to look in their direction, when a different person spoke, the remote user usually looked towards them. When people in the meeting room were talking to the remote user, they addressed the video camera system. It was not determined if this was also affected by the closeness of the phone system to the camera. However, a number of problems arose that reduced the experience. These can be categorised into two separate problems, the first relating to technical limitations of the experiment, the second related to the expected feeling of presence.

9.1.3.1 Technical Issues

Video quality: The quality of the video feed to the mobile user was slow and the video resolution was poor. The main reason for this is that a very large video image was used, 640*480 RGB. Therefore each frame was 7.3724 Mb’s in size, so over a 10Mb/s LAN connection we should get about 1 frame per second. Added into this factor, is that an Ethernet connection is not usually operating at 100% capacity, more like 60-70%. Hence there were usually only 35-40 frame updates per minute, rather than the preferred 25-30 frames per second. Video compression would have assisted in this task greatly, however none was implemented as part of this experiment. This low framerate detracted from the user’s experience. This was noticed less when the remote user looked in one direction for considerable periods of time, but the induced latency effect from the slow framerate was distracting when they changed their head position quickly.

The second issue was that the video resolution was also a problem. The image captured by the omni directional video camera was a 360-degree video image at resolution 640*480 with RGB colour scheme. The video image shown to the mobile user was at the same resolution, 640*480 with RGB colour scheme. This was a key requirement that the image shown to the user was in full screen mode for immersion reasons. In practice, the panoramic-to-perspective conversion software could only use a subset of the panoramic image to generate the perspective. For the FOV of 60 degrees used, that was approximately one sixth of the panoramic image or 67,000 pixels. With these 67,000 pixels, the conversion software filled out the pixels by extrapolating information to give a perspective image of 640*480 resolution image. The effect, however was to seriously blur the actual visual data. In practice, beyond a distance of approximately 3 meters, it was quite impossible to distinguish between different people. Better images could have been achieved using lower resolutions for the perspective image, such as 320*240. However, it was felt that this would have been too small an image to try and give a sense of immersion to the user. The other alternative would be to increase the resolution at the panoramic camera side, but 640 * 480 is the max resolution of the Panasonic GP-KR222.

Therefore, the poor performance of the video stream in terms of framerate and resolution quality, was then the most significant factor that subtracted from the subjective sense of presence for the mobile user.

Lighting. No consideration was given to the effect of lighting prior to the experiment and the effect of the oversight was to become apparent quite quickly. To the right side of the room in relation to the camera, there was a window through which the sun was shining directly. While the mobile user was looking in the other direction, there was no problem. Once, however, they looked to the person to their right and towards the sunlight, the image became too bright and they could not distinguish anything. The reason for this is that the light distribution through out the room was uneven. The type of parabolic reflective surface used on the camera set-up made this worse. The light fell with far greater intensity on one part of the camera, while other parts were less intense due to the curvature of the reflective surface. The only means available to rectify this situation was to close all of the curtains, and relying only on the room's light source.

9.1.3.2 Usability and Presence

Audio: The camera system used in the experiment was an omni-directional camera providing the mobile user with a 360-degree view of the room. The mobile user made use of that ability to the maximum when conversing with the participants. As each person around the meeting room talked, the mobile user usually turned to look at who was currently talking. The system however did not have a 3D sound system. This meant that the users had to try and determine who was speaking before they could look at the person speaking. This was not helped by the poor video quality, both resolution and framerate, which meant that the mobile users could not properly determine who was speaking by their lip movements. In many cases, the mobile user relied on their own voice recognition, i.e., they were able to associate a voice with a person, and knowing where they were sitting, they could then turn to look at them. In cases were the person did not recognise a strange voice, it was quite difficult to know who to look at. This confusion was quite disorientating to the mobile user, who struggled to quickly determine who to look at. A 3D sound system should be used in conjunction with the video system in order to increase the sense of presence and mitigate the mobile user's confusion.

Gaze Awareness: Despite the attempt to create a virtual proxy, there were still strong limits on the user's perspective as a result of the omni-directional camera system. One quite telling incident during the experiment concerned the placement of the video camera. The mobile user was in a standing position, yet the position of the camera within the room was placed at table level. The horizontal view of the mobile user was then at a table level height. The other user's horizontal eye level was somewhat higher than this level. The result was that the mobile user felt that he was looking up at the other participants. In other words, the eye to eye level contact was necessary for a feeling of communication, equality and respect with the other participants. This is in line with research carried out by others such as (Okada *et al.* 1994). For this reason, the camera position had to be raised by some cm's (using books as shown in figure 50(c)) to provide a sufficient eye level contact for the mobile user.

The fact that the position in the remote room (sitting) contrasted with the standing position of the mobile user did not raise any questions. It seems the mobile user accepted his position as shown to him, even though it contradicted his actual position in reality. It was only when some restrictions to the users sense of equality was challenged that the user demanded equality of view and a change of camera position.

The lack of a two way video feed can sometimes cause problems for the people located in the meeting room. At one point during the meeting, a passerby saw the mobile user with the HMD and came over to talk to him. The mobile user started to show the device to the passer by and in doing so, removed his presence from the meeting room. The people in the meeting room tried to talk to the mobile user, still believing that he was taking part in the meeting, to no avail. If the people in the meeting room had some knowledge of the mobile user's state either by video or verbal cues, they would not have wasted time trying to communicate with the mobile user.

Interaction: The meeting itself was conducted in an informal manner and simply discussed different ideas, mainly pertaining to the experiment in hand and related research factors. At one point, the remote user asked for the email address of a person in the meeting room. As the person in the meeting room was foreign, the mobile user was not sure how to spell that person's name. At this request, the person in the meeting room wrote down his email address and showed it to the camera. The mobile user could barely decipher it due to the poor quality of the video, so the text had to be made larger.

This little incident highlighted the fact that in the meeting set-up we had factored in two mediums for interaction, voice and visual. What had not been factored in was the need during the meeting for passing around information other than voice or visual display. There was little to no capability to exchange information within the context of the system. In contrast to CVE's where all of the data artefacts used are freely available and exchanged with other participants, no such facility was available within the context of our meeting room experiment. To do so, one would have had to use an external means of communication in parallel to the video and voice connection. There simply is no means for someone to pass around objects such as slides to a person who is not physically present. It is then noticeable that our system was severely found wanting for any interaction that did not involve either voice or visual.

9.1.4 Discussion of Experimental Impact

This experiment was conducted at an early stage (Spring, 1999) of the development of the TeleReality constructs. At this point, it was expected that a virtual proxy that supported multiple perspectives would contribute to a feeling of presence. That this was achieved while supporting a multi-user system was one accomplishment.

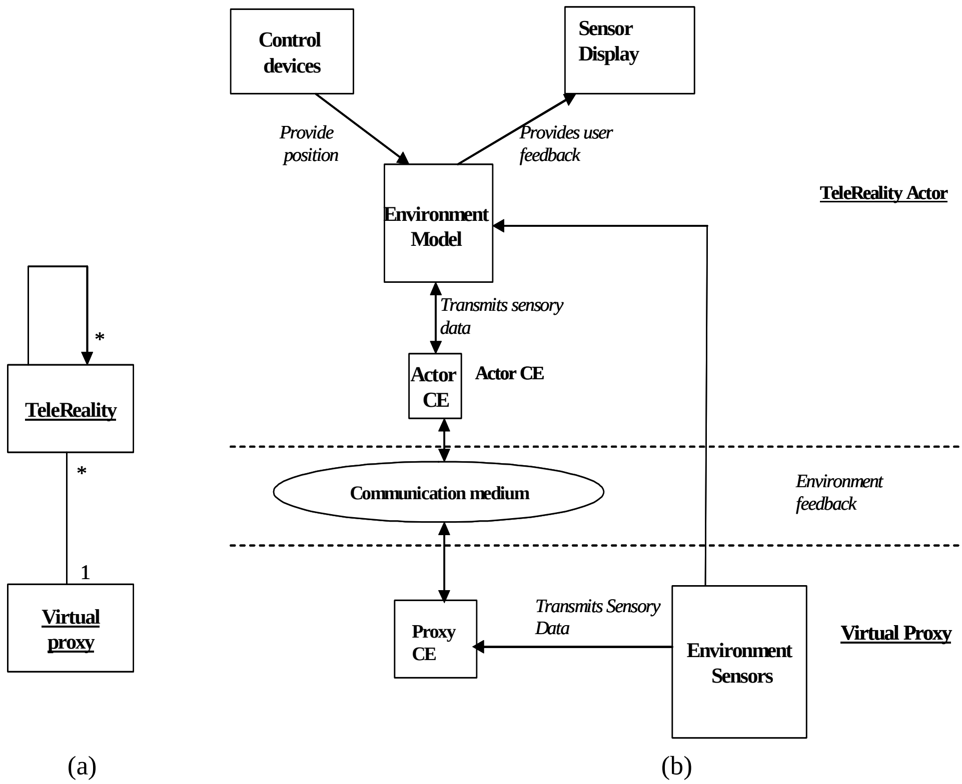


Fig. 51. a) The virtual proxy can be used by many telepresence users. b) The constructs for this system are shown. The system already makes use of environment sensors and a MR environment

The constructs that made up this experiment are shown in figure 51. In figure 51(a) the multi-valued role of the user and proxy (in this case the omni-directional camera) are shown. Figure 51(b) shows the overall constructs. There are two important points about the constructs in this system. First, there is the introduction of the environment model. This is located with the TeleReality user here because it is here that the user calculates the correct perspective that is shown to the sensor display construct (in this case the HMD). This works well for this case, as the environment sensor has only a single camera which broadcasts the same video to everyone. The second thing which makes this case different from other telepresence systems is that there are no sensors which have a direct logical connection to the sensor display constructs. The information that comes from the remote environment cannot be directly sent to the users display as it will make no coherent sense. The user sensory feedback only comes after processing through the environment model.

Another point that should be made is in regards to the absence of actuator related constructs. Had our system contained a non-omnidirectional camera, but one whose orientation could be controlled, e.g., pan/tilt/zoom, then only one person could use the system at anytime and this would effectively take one of the most important aspects of a telepresence system as set out by Sheridan and others, the ability for multiple users to exert control and work in the remote environment.

It was anticipated that gaze awareness would be an important factor in supporting presence, although the degree of importance was not fully realized. The consequence of gaze awareness was to realize that, while the tested system provided a multi-user system with some freedom of view, that this perspective was still severely restricted by the single camera, i.e., the user could not ‘stand up’, ‘sit down’ or move around. This complexity was even worse if one considered using a physical proxy like a robot, which would mean having a rather large device that could not be controlled by more than one user. The focus then was to examine how one could have a greater ability to navigate the remote environment without the need for a robot.

The importance of interaction had not been fully considered when the experiment was undertaken. However, the importance of interaction became noticeable during the experiment and further research showed that other researchers in other fields considered it to be not only essential for a system to be useful, but also an integral part of supporting immersion (Zahorik & Jenison 1998). The implementation of interaction would have meant two things, having a common framework where the remote user and those present in the environment could use to work together, i.e., support a collaborative work environment. This presents a problem in how do they communicate. The solution was that the users in the environment should have AR equipment and ad-hoc network elements to connect them to the fixed network and from here, a common interactive environment using synthetic objects can be formed with the remote user. With the HMD’s, all of the users could see and share a common set of information represented by synthetic objects.

The anticipated introduction of this camera system and ad-hoc network technology meant having to deal with problems brought about by the fear of a big brother system, and the inherent security risks associated with ad-hoc networks.

Taken all of these factors into account, it started to appear that what was evolving was something different, something that was not currently implemented as part of telepresence systems and contained many differences. This experiment helped confirm that new building blocks would have to be proposed and studied.

9.2 Evaluating the Visual Cell system

The visual cell system was the next version of the mobile VR meeting experiment. The visual cell system was developed as a model of a TeleReality system, in order to test out some basic ideas about the constructs. It is identical to the construct chart shown in figure 23, although the security/privacy constructs were added in much later. It differs from the mobile VR meeting room experiment in that it supports limited interaction and supports the ability for the virtual proxy to move from one room to another by using a system of fixed cameras.

The visual cell model described in chapter 8 was partly implemented at the University of Oulu as part of the PAULA project, which ran between the years 1998-2001 (PAULA 2003). Many of the core elements for a TeleReality system were not satisfactory or technologically available. With these restrictions in mind, a simple TeleReality system was implemented to test out the basic constructs. The purpose of this system was to try and determine the logical structure of the constructs and their cohesiveness with each other. This experimental system then concentrated on software architecture, the environment model (simple cell handovers, artefacts provision), the MR Environment constructs (user interface for the system) and the cyber actor. As no proper ad-hoc network was available, a wireless LAN was used to hold some artefacts.

9.2.1 Visual Cell Implementation

A simple two-cell system was constructed and the configuration was as shown in figure 46. The software architecture for the system is shown in figure 52. The system is scalable and extra cells can be added if necessary. The VC server and one base station are combined into one computer, a Pentium 1.4 GHz. Also the remote user's client application was merged with the VC server software for reasons of convenience during development, although they can be easily separated. The entire system operated within a single network, and the software could operate from any node within the network. Communication between components was managed by using TCP/IP port connections. The second base station used a Pentium 400 MHz computer and only contained the basic image capturing device software. The artefact's database resided in a 3rd computer, which was accessed by the client software.

Camera system Sophisticated image capturing systems were not used in this project. Implementing a real time sophisticated camera system would have taken a good deal of resources. A single camera system per base station was used instead. Each camera had controlling software. A Matrox Meteor II Frame grabber captured the images. Each captured image was a 640*480 RGB image. The image was sent whole to the central server. A video codec would have been useful for sending the images but none were implemented. The cameras used were Panasonic GP-KR222's. These could be configured into 2 versions, one presenting a perspective with a FOV of 60 degrees, and the other as an omni directional camera with a FOV of 360 degrees. In the omni directional mode, the client uses a Polhemus tracker to discern the orientation of the user head and a new perspective image is calculated before it is rendered to the client user. The panoramic mode provides a greater flexibility of view at reduced image quality over distance. The perspective image gives a sharper image but at the expense of a narrower FOV.

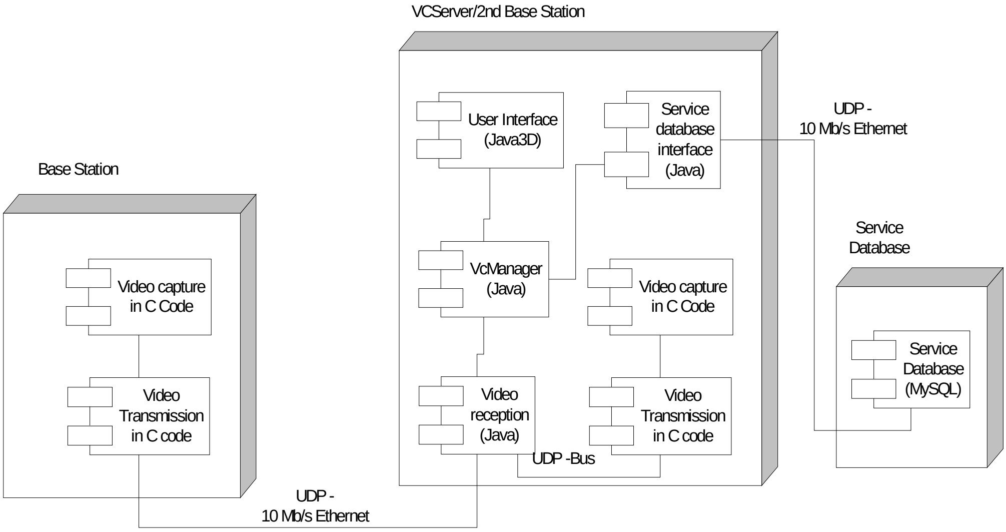


Fig. 52. The software component model for the visual cell system.

The base station software was written in C, which controls the interface to the Matrox Meteor card, while the rest of the software was written in Java/Java3D.

9.2.1.1 VCserver software

The visual cell server manages the network and it determines when a handover is required for users. When the user initiates a forward movement, this is detected by the VC Server and the correct handover to the new cell is initiated. This means that the video feed from the old cell is discontinued and the remote user is provided with the video feed from the new cell. The VC server also works in conjunction with the artefacts database management software to determine the new artefacts that are available to the user and what old artefacts are discontinued. The code for the VC server is written in Java.

The service database contains two parts, the database itself, which uses MySQL, and software to manage the artefacts. The service management software keeps a track of the location of each artefact and provides a list of all available artefacts to a remote user based on their location. The Management software offers an interface to the database and also to the VC server. If necessary, user artefacts can be registered and removed from here also. The management software is written in C.

The user interface software is written in Java and Java3D. It offers the front end to the system, and is responsible for connecting the user to the server, displaying the video images and artefacts to the user. The type of interface that is generated is shown in figure 53. The user has two options to move between cells, to press the up-arrow key on a keyboard to indicate forward movement, which will cause a cell handover in the server. In the fixed camera perspective set-up, this means that only one possible inter-cell handover occurs. The camera itself points to the next cell. In an omni camera set-up, multiple cell handovers can exist, depending upon the user perspective.

The other means to enact handover is to use a teleport option. By selecting an interactive object displayed to the user in a cell, the user is transported to the next cell that is connected to the current room. In the perspective camera set-up, this system enables many handover objects to different cells to be included in the system.

The interaction system is based on a simple cycle/select scheme, which is similar to the alt-tab action on windows. In a cycle/select scheme, a list of all of the artefacts available to the user is obtained from the artefacts database. By pressing a button on their interface design the user can move from one object to the next until they have the artefacts they desire. Then selecting this object causes the artefacts to launch a more detailed interface for interaction with that application. The reason for using a cycle/select scheme is that it does not require the users to use a gesture based system (Zimmerman *et al.* 1987, Antoniac *et al.* 2002). Gesture based systems have problems differentiating between objects that are partly occluded by other objects. Also, the cycle/select scheme is easier and simpler to implement.

The interaction system made use of three artefacts objects, these were an information object, '*I*' in each cell showing the owner of the room and any relevant information mentioned by them. A '*door*' object was used to effect an action in the environment, which was to teleport to the next cell. The '@' artefacts object was a message board, where the remote user could leave a message for the person who owned the room. These can be seen in figure 53.

(a) (b)

Fig. 53. (a) A view from the visual cell with 3 synthetic Objects, the @ object is highlighted. (b) An application tied to the (i) spherical object is launched.

9.2.2 Discussion on the Visual Cell model

The visual cell model had a similar problem to the VR meeting experiment in that it does not make use of a video codec. As a result it only had a transmitted framerate of approximately 20-30 frames per minute. There was no scope for collaborative actions by multiple TeleReality users, although it was planned for future actions.

The cameras were then set up strategically for a 2D system. The use of a single 2D camera system meant that the direction of the camera basically became the equivalent to a path, users could only move along this path with no capability of looking around them.

A main aspect of this implementation of the VC model was the application of the modified multisphere model for artefact distribution, the movement from one visual cell to another, the construction of the user interface and an increased knowledge of the requirements for the software architecture

9.2.2.1 Software Architecture

The principle difficulties in the software architecture arose when trying to implement the visual interface as seen by the virtual proxy for the mobile user. There were two choices for implementing the 3D artefacts, OpenGL and Java3D. Additionally, the software would also have to support a video streaming. As part of the software was already written in C, OpenGL was first examined as a possible solution. However, it became apparent that it was difficult to use OpenGL with video, as there was no inherent support for video. Then Java3D was examined as a possible solution and it was felt that video could be easily incorporated into a Java3D environment. This choice however had a serious consequence that was discovered later. When the Java3D scenegraph is compiled, no extra information can be added to the scenegraph (as of version JAVA 3D 1.2). Different nodes may be modified or changed, but no new nodes could be added.

The consequence of this is that in a TeleReality system one is continuously downloading new graphical 3D objects from people who come and go from the system or when moving from one cell to another. These objects have to be rendered by the system, but these objects cannot be added to an existing executing Java3D scenegraph. This is not the case for OpenGL, where new nodes can be added in or deleted at any time. So, for Java3D, each possible interactive object that can be used by the system must be pre-rendered and known before the system is started. While this could be managed for this project, it severely restricted some of the concepts that it was planned to do, mainly, examine the involvement of independent objects.

A second problem with using Java3D with the video object was the slow rendering rate. The video was rendered into a background object and this took a considerable degree of processing time to update one frame, adding to the problem of the slow framerate occurring due to the lack of a video codec. It had been attempted to texture map the images to create a 3D space before hand, but there was no support for updating or changing these texture maps once the program started executing. This is an improvement over OpenGL however, as it had no means for supporting video that I could find, or even after consulting experts in the field. OpenGL simply was not designed with the use of video in mind.

The other problem with the Java3D system was the performance. The system ran with a number of threads, e.g., one handling the communication, another handling the server or rendering. All of these threads consumed a great deal of resources and while all had equal priority, the system was sluggish. When priorities were changed, the system failed to work satisfactorily. Allocating higher priority to the rendering process caused all of the spinning virtual objects to slow down considerably and open up very slowly (which was a problem when trying to teleport to the next room). It also slowed downed the communication software, hence reducing the framerate further. Increasing priority to the communication system caused the other process to slow down so that the received images were rendered very slowly. Assigning priorities to the virtual objects also slowed down the communication and rendering actions. To improve the situation, an improved Java VM with better task handling would be needed, but was not implemented in this project. The consequence for this seems to be that any client software architecture and software for TeleReality users needs to have integral support for video and the additional downloading and running of 3D graphical objects.

9.2.2.2 Interaction

The second purpose of the environment was to test out the issues involved with interaction in the visual cell environment. Two aspects were explored, the user interface technique and the attitude of users to the virtual objects themselves. The user Interface used a cycle/select scheme to interact with three objects available within the environment. There were two reasons why a cycle/select scheme was taken over a gesture-based system. The first was limited resources available for the implementation of a gesture-based scheme. The second reason is that it is expected that the TeleReality and AR user will have a mobile phone as part of their BAN, and hence may use the interface paradigms familiar to mobile phone use’s i.e., cycle/select. The idea was that the user would wear an HMD and view a live video of the remote room and interact with the visual objects. Interaction would occur by using the cycle/select scheme with a keyboard or mouse. The idea being that the user would use their sense of touch to control the cycle/select buttons.

The purpose was to see how natural it was for the user to interact with the environment, without looking at the keyboard/mouse. A limited usability test was carried out to determine the concept and generally, the subjects managed reasonably well on a short test, which would seem to indicate that it could be used, although some subjects said that they would have liked to use a gesture system instead.

Other interesting points were that many of the subjects were quickly able to associate objects with expected actions. The one confusing object was the ‘@’ object, as no one understood what it meant until they tried it. Attaching textual explanations would have helped them. None of the users made a connection between the intended strategic placements of the objects in the environment. While the video image was 2D, the objects inhabited a 3D space, and different objects were placed at different distances from the viewing plane. However, this was not noticeable to the users, as they had no way of knowing what the actual size of the objects should have been. Therefore there was no need to strategically place the objects in the environment. This may behave differently in a 3D-video environment. The benefit of this is that accurate registration may not be essential for many collaborative work tasks from the TeleReality user's point of view, but it is likely that mutual gaze awareness would be.

9.3 Discussion of VC model constructs

The outcome of the VC implementation was to highlight a number of important points for the implementation of the system. It demonstrated the ability to connect two partitions together. It highlighted the problems in the software architecture that need to be overcome before a properly working system can be used. It also seems to highlight that strict registration as used by AR users is not necessary for TeleReality users, and that for many applications the strategic positioning of objects may not matter as much as the object itself. The organisations of the artefacts also seem satisfactory and are not a major obstacle to the system. Extra information on the nature of the UI was also gained. This indicated that there needs to be a limited number of objects in the space, even relatively small objects were responsible for significant occlusion in the environment.

The constructs for this system also seem to be located in the most logical places. An issue could be made to locate the environment model in the TeleReality user's plane as was the case with the VR meeting experiment. However the VR experiment meeting only used one camera, whereas this system uses more than one covering more than one partition, hence it is not logical to send all of the information to the remote user as most of it is redundant and it takes up bandwidth. A second reason is that different visual cell systems can have different image capturing devices. In our system, one had an omni directional camera and a normal camera. The omni directional camera would have required image proceeding software that the TeleReality user would need to have. Therefore the TeleReality user would need to know information about the camera system used in the remote place and have the appropriate software. The more systems you have with different camera configurations, the more intelligence and resources the user would need to possess. The position of the environment model in the remote environment on the other hand would be responsible for the environment processing and only provide the user with a video stream. The issue here is one of transferring the processing power to the infrastructure and not onto the users processing devices. Even at its simplest, which is the case in this experimental model, the environmental model must make basic decisions about which images from which camera should be sent to the remote user. Giving all of this information to the client side would pose a security risk, as well as being a strain on bandwidth and computer resources. An additional reason for having the environmental model located in the virtual proxy, is that it ensures that everyone has access to a common source, otherwise if others had this in their clients, they may be viewing slightly different views than what others are seeing because of differences in processing power, latency etc. Another reason for the environmental model here is that it can have cyber actors that offer services covering a number of visual cells, i.e., the local area sphere in the multisphere model. If it did not exist, then the cyber actor would be responsible for establishing communication to all of the basestations in the visual cell. To summarise then, the reason for moving the environment model to the virtual proxy is one of management, brought about by the additional cameras, the need to cover larger spaces and additional cyber and AR actors.

The MR environment construct is an additional construct located in the TeleReality actor side. The need for this is different than the environment model construct, as it must construct an interface to the remote world by mixing received video and sound with synthetic objects that it owns or obtains from the remote environment. This is a standard construct from AR systems and has more of a transformational role than a decision making role.

The information storage construct is added here as it provides information on the artefacts that will form the basis of interaction. This information is represented by synthetic objects and is linked to applications and data sources. These are not often a part of traditional telepresence systems as those systems were usually designed with the aim of doing physical tasks, whereas for collaborative work, the TeleReality system focuses also on information processing, although it can also be used to take control of individual devices that have actuators. This construct is valuable also for AR users, and one should keep in mind that a person can be a TeleReality user in one case, and an AR user later if someone enters their current real world space. Therefore, the more similarities there are in constructs between AR actors and TeleReality actors, the better.

AR actors were not included in this experiment, but cyber actors were. One difference from the construct model shown in figure 26(b) was that they did not have a separate Communication element but were a component in the system. The cyber actors were component independent (in java) and could have been connected to the system by a wireless communication element or TCP/IP connection without any problems for the system.

9.4 Evaluation of Constructs components

The TeleReality constructs are evaluated based on their importance as building blocks for a TeleReality model and implementation. Table 14 shows an overall summary of how the components of these constructs were addressed in the mobile VR experiment and the visual cell model. The deficiencies found in the mobile VR experiment were based on the lack of support for a number of these important components. The visual cell model was a more comprehensive system, encapsulating all of the basic constructs. The implementation of this model did not meet all the goals of a TeleReality system because neither an AR user nor security/privacy constructs were implemented. Hence, the deficiencies in the system were due to a lack of implementation of all of the requisite constructs.

9.4.1 Visual scenery and objects

The VC model and implementation separated the general scenery from the virtual interactive objects in the system. Had an AR user been present, they would also have had access to these objects. These objects were seldom updated. For the VC model, the visual scenery did not need to be updated as often as it was, because the room was mostly empty. It only needed higher update rates when other people were present. The importance of these constructs is in maintaining low data bandwidth and a front end for interaction with the system artefacts.

Table 14. List of constructs components addressed by experiments

TeleReality Construct components	Mobile VR experiment	VC Model Experiment	Description
Visual Scenery	Yes	Yes	Used video images to provide scenery.
Visual Object	No	Yes	Used Java3D graphical objects to represent artefacts.
Presentation	Yes	Yes	Both used a 2D system (video), needed an omni-directional camera to support multiple perspectives.
World Structure	Yes	Yes	Dynamic, world updated in real time.
Partition	No	Yes	The VC model divided the world into two linked cells to create one space.
Portal	No	Yes	A 3D object was used as a portal.
Handover	No	Yes	Handover achieved by a) interacting with a portal, b) pressing a move forward key.
Location System	Yes	Yes	Both used a location system based on a symbolic address only. Had an AR system been used, geographical information would also have been used.
Actors	TeleReality	TeleReality and Cyber	AR actors were not implemented in the experiment. Cyber actors presented services to the TeleReality actor.
Transactions	Yes	Yes	E.g. keyboard keys for interaction, UDP/IP for video transfer, SQL for database access.
Artefacts	No	Yes	Only application interaction supported, no migration or replication.
Personal User Sphere	No	Yes	The VC expt. had some artefacts that came from the users defined Personal User Sphere.
Immediate Area Sphere	No	Yes	Each VC basestation acted as a cyber actor with their own artefacts.
Local area sphere	No	Yes	The VC system also acted as a cyber actor with one artefact.
Global network	No	No	This was not implemented.

sphere			
Trust Contracts	No	No	An organisational contract, not supported by experiments.
Awareness	No	No	As no AR user was implemented, users did not have means of becoming aware.
Control	No	No	Not implemented.
Access	No	No	Not implemented.
Integrity	No	No	Not implemented.
Ad-hoc security	No	No	Network key, ad-hoc key and device profile. Not implemented as no AR user present.

The second important reason for these constructs is that an AR user would have no need for the visual scenery, as they inhabit the space, but does need the visual objects for interaction purposes. This also points to the need for dividing the visual field into these two components to support coordination between the AR user and the TeleReality user.

9.4.2 Presentation and World structure

Both experiments were dynamic systems that needed to operate in real-time to support communication between TeleReality users and those in the remote room. The visual cell model could be modified to support a static system, but that was not examined here. Both systems used 2D video images, however the use of an omni-directional camera gave some flexibility of view to the users.

9.4.3 Partitions, handover and Location systems

The visual cell model implemented partitions in what it called visual cells. This enabled the users to move around a larger space, in this case between two rooms. This is something that could not be done, for example, in the mobile VR experiment. While the camera system used was very simple with no proper support for extensive navigation, it does demonstrate the potential.

The moving from one cell to the next, or handover, was carried out by using the up-arrow movement key on the keyboard in the normal camera perspective mode. Normally, the position, perspective and movement of the user would be needed. In the VC experiment, while using the normal fixed camera perspective there was no ability to move around the space within the visual cell and the camera was pointing at the door, effectively providing a fixed perspective. Therefore the perspective part of the handover was fixed, as was the position of the user (the camera position). This only allowed movement as a factor in determining handover. When an omni-directional camera is used, the handover is more complex, requiring the user to look in the direction of the door first, to get the correct perspective, and then using the movement key to cause handover. An alternative approach was also used, by interacting with an object representing a portal, in this case a door object, one could teleport to the next cell. To reinforce this concept the object was placed adjacent to the door.

The location of the cells was based on their IP-address and port number. These, in effect, combined to give their symbolic address. Had an AR user been included, the geographical data would also have been used in the world construction.

9.4.4 Actors, Artefacts and Transactions

The lack of an AR user meant that the AR actor was not implemented. The loss of this actor meant that collaborative work activity between the TeleReality user and people in the environment were not possible. This was a severe restriction on the system as it excluded the ability to interact with people located in the remote room, an important conclusion from the VR meeting experiment. The lack of this component to the system indicates its importance as a basic construct of the system. The cyber actor existed in the VC model, supplying message board services, while the TeleReality actor provided personal information services. A number of artefacts, namely Java applications, were included to provide some functionality and tasks that can be carried out within the environment. These provide the basis in which work is performed by the TeleReality user in the environment. Finally transactions linked the actors to the artefacts, by sending user interface commands with the keyboards and transmitting video/audio and data between actors and artefacts.

9.4.5 Multisphere model

The artefacts in the visual cell model were arranged according to the multisphere model and their placement is shown in figure 54. Artefacts that belong to the user are included in the personal user sphere. The '*door*' object is included in the immediate sphere as it is heavily tied to the physical location of the door. The '*message board*' object is located at the local area sphere as it can be seen by people over a much larger area.

The advantage of using the multisphere model approach for the organisation of objects is that it also acts as a means to filter out a large amount of objects based upon the location of the user. Objects closest to the user are displayed to him, while objects further away are not. Julier's research team also effectively use a geographical based system based on distance to filter out information in mobile AR applications (Julier *et al.* 2002). If an AR user was present within the experiment, the dynamic between two personal user spheres could have been studied. The global sphere was not implemented, as no external services were used. Alternative means of arranging artefacts could be considered, but this is the type of model most often used in CVE's and proposed for user centric design. Additionally, it is the most appropriate for use with the system of partitions expounded in this thesis.

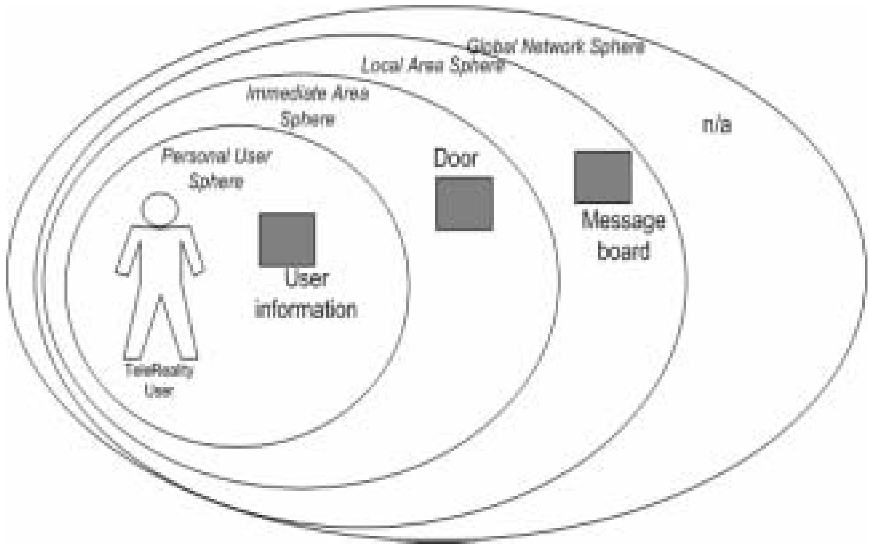


Fig. 54. Deployment of objects within the multisphere reference model.

9.4.6 Trust and security

The final set of constructs, the 'trust contract', awareness, control, access, integrity and ad-hoc security techniques were not fully implemented in either experiment and hence could only be evaluated to a limited extent. The basis for proposing the trust contract therefore comes from the belief that people will not embrace a system that is strongly invasive on their privacy. The reason for believing this comes from the research performed by others into *media spaces*. Unlike media spaces, where the goal was to support general awareness of others, whether they wanted it or not, TeleReality focuses more on a telephone based model of direct call connections. The actual connection in the VR meeting experiment uses the telephone model as it would be inconceivable that someone should take part in a meeting without been invited. The users in the room with the camera were always able to refuse the connection by refusing to pick up the phone and establishing the video link, giving them some control over the system. This provided them with some degree of control over that environment. In that experiment, general awareness of who was taking part could be somewhat confusing a there was no visual feedback for the meeting room participants of the remote user. In general, without the constructs identified here, one would find it very difficult to believe that such a system as TeleReality could be widely accepted, therefore indicating that they are of extreme importance.

The proposed means for organising the restrictions on using ad-hoc network elements to support security were also not tested due to the lack of an AR user. The current state of ad-hoc security protocols is not at a sufficiently advanced state to inspire a great deal of confidence. The proposal to force individual ad-hoc users to connect to a fixed network for 3rd party authentication is purely meant at this stage as a potential quick fix and the author does not offer it as the definitive solution.

9.5 Summary

This chapter describes two experimental systems and compares them against the construct components. The first experiment is called the Mobile VR experiment and was conducted in March 1999. This experimental system looked at using VR technologies in a videoconferencing environment to determine existing problems and benefits, as well as anticipating future lines of research. The experiment had technical limitations, but highlighted the importance and limitations of interaction and navigation in videoconferencing systems. This experiment strongly influenced the development of the research question addressed in this thesis.

The second experiment built on the results of the mobile VR experiment and led to the creation of the visual cell model. The visual cell model incorporated stronger support for navigation across wider areas, a fuller interaction model and multi-user support. The experimental system however suffered from the lack of an ad-hoc network element, and the 2D camera system and available resources resulted in a restricted level of navigation. Consequently, the bulk of the experiment ended up focusing on the user interface issues with a cycle/select scheme being used to interact with virtual objects. The objects were organised on a multisphere model. This second experimental system comprised many of the basic constructs identified for a TeleReality system and a discussion on the effect of these two experiments on the reason for selecting and placing these constructs is provided.

10 Discussion on an enhanced telepresence system

This chapter discuss the overall consequences of this work. First, a discussion on the impact the system has had on the research question and how the constructs were formed and proposed, in accordance with the constructive research method. A consequence of the place of this work within existing theoretical theories is also discussed and, finally, the existing fundamental technological obstacles to the future realisation of this technology are discussed.

10.1 Discussion: Addressing the Research Question

The research question set out to see what form an enhanced telepresence system using virtual proxies would take so that it is multi-user, interactive and navigable. These characteristics are important, and to support these requires modifying the constructs of existing telepresence systems. The key to a solution for this problem is to replace the most limiting aspect of a telepresence system, the physical proxy, with a virtual proxy. This virtual proxy acts as the basis for providing multi-user support, interaction and navigation.

10.1.1 The Virtual Proxy

The virtual proxy in the experiments works by giving the user a floating camera free perspective. Taking the omni-directional camera as an example, it captures more information than the user needs. The user can then select a subset of that information to view specific data. To the user, it is the perspective that is important and needs to be controlled, not the camera itself. This control and image processing is carried out in software controlled by the user.

In the visual cell model, a similar principle applies, except that multiple cameras are used. Again, the perspective image the user receives is determined by software. The user does not say, I want to move from room A to room B, so I must move a physical proxy. Rather the user indicates that they wish to move to another room and the software decides on what is the most appropriate camera perspective that should be chosen for that. The virtual proxy in both cases works by the user expressing intent, and the system works to provide the correct perspective to match that intent. For the other people in the system, who look upon a TeleReality user, then that user may be represented by an actual computer generated avatar.

10.1.2 Multi-user

Supporting a multi-user system is addressed by the means in which the virtual proxy is constructed. The perspective image that the virtual proxy chooses is selected from a set of common broadcasted video images. This means that a number of users can take the same video images and construct a perspective that is relevant to them. No one user can control what another user wishes to see. The limits on the number of remote participants are restricted by the technology of the broadcast video and available bandwidth, not by the actions of the user. The object with which the user interacts with can be accessible to multiple users in much the same way as multiple users can simultaneously access a web server. As a result, the system is multi-user.

10.1.3 Interaction

A concern about interaction was that if one did not use a physical proxy, how could one interact with the environment? The solution to this problem was to extend the telecommunication network to include all of the objects that can be interacted with. For many cases, this extension was to provide an ad-hoc network element that would act as a servant to the user within the remote environment. The network system is then comprised of a fixed network to connect the TeleReality user to the remote site and an ad-hoc network element to dynamically connect to various objects and people who exist in that environment. Interaction is also supported by the use of computer generated objects that represent the various objects in the environment. This user interface technique presents the user with a means for carrying out complex actions in a simplified manner.

10.1.4 Navigation

Navigation is supported by the construction of an environment which consists of numerous video cameras strategically placed throughout the environment. The basic technique one needs is to create a matrix of fixed cameras that provide video images to remote users. Control solutions are then required to determine which video images are relevant to a user's desired location and perspective. This organisation is done by creating a set of partitions, where each partition contains some form of image capturing devices. A large covered area is then established by connecting these partitions together. The decision process on which camera system should be used to provide the correct perspective means that decisions must be made while the user is within the partition and when the move to another system in a different partition. To support the moving from one partition to the other requires a complex decision process. Therefore, handover decisions are a fundamental part of any system.

10.2 Addressing the constructs

The constructive research method was chosen as the most suitable form for addressing the research question. The rationale for this is that existing systems did not have the basic constructs that one could answer the research question. The new constructs proposed must fit together logically and cohesively in order to be accepted. The constructs were general building blocks. Sheridan (1991) provided the basic constructs needed for a traditional telepresence system, while one needed to evaluate VR based telepresence system (Milgram & Ballantyne 1997, Lin & Kuo 1999) in order to identify the constructs. What occurred next was a process of modifying these constructs models through the VR experiment and the Visual Cell model in order to obtain constructs that specifically related to the research question. The evolution of these constructs from traditional telepresence systems to TeleReality constructs is shown in figure 55.

There are a number of significant differences between the TeleReality construct model and those other telepresence systems. The first difference is the virtual proxy that is introduced in the VR meeting experiment. This proxy only carries over the proxy CE construct as well as the environment sensors that were important in VR based telepresence systems. Gone are many of the constructs related to the control of the physical proxy. This virtual proxy had no capability for physical or data interaction. The VR experiment also saw a significant change to the user of the system. The user became the TeleReality actor, and the main difference is that it did not use the virtual model construct, but retained the environment model construct. The reason for discarding the virtual model was that the VR based telepresence systems used a stored 3D graphical picture of the remote area to build the environment model while the VR experiment relied upon a live video feed to build the environment model.

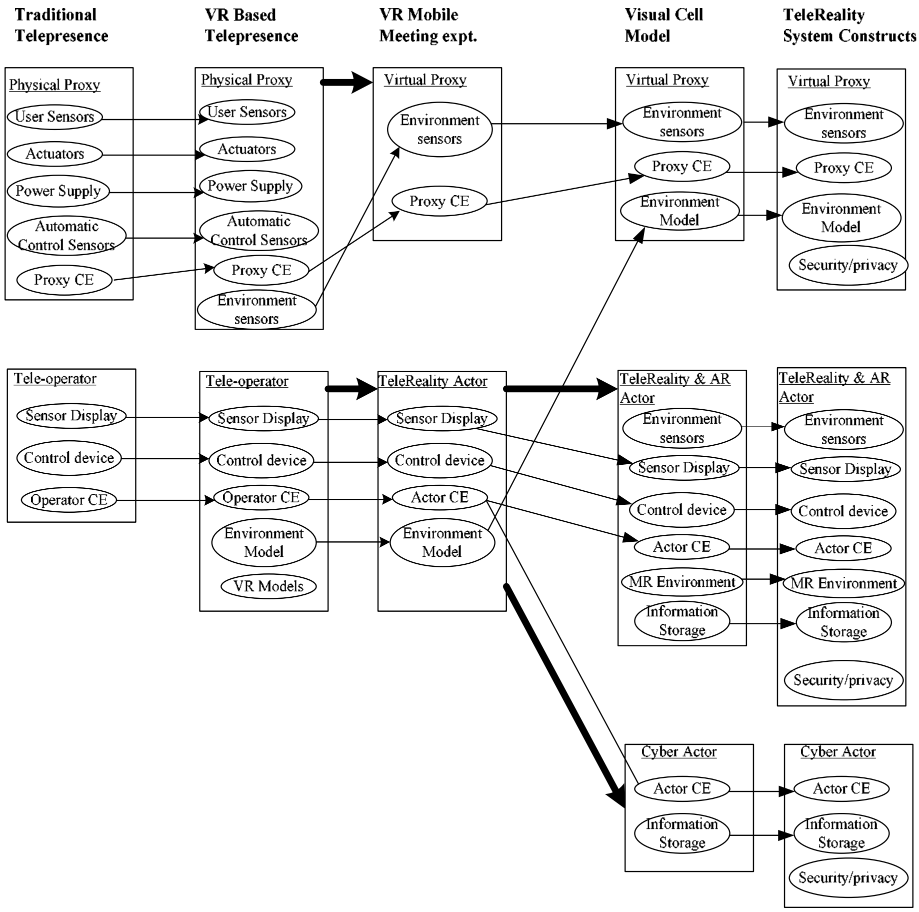


Fig. 55. Evolution of Constructs from traditional teleoperation systems to the TeleReality constructs.

The progression of the construct development from the VR meeting experiment to the visual cell model came as the result of the need to address the deficiencies of that system. In particular, there was a stronger desire for better interaction between multiple people and improved navigation capabilities. A need was seen for remote and local participants to interact together, primarily to exchange information such as data files. This lead to the incorporation of AR actors in the system. As the experiment also identified the need for autonomous agents to provide services, such as video projectors and even the occasional robot, these agents became cyber actors. For practical reasons, it was decided that TeleReality actors were to be as similar as possible to AR actors. The reason for this is that a person who can be a TeleReality actor in one session, could be an AR actor another time and it is better if they both have the same equipment. Three new constructs for the TeleReality actor are introduced as a result (they already exist for AR actors), the *environment sensors*, *MR environment* and *information storage*. As interaction is primarily about the exchange of information, although physical interaction is still possible through the control of a cyber actor, capabilities to support it can be found in the information *storage construct*. *Environment sensors* are important to detect what the actor is attempting to do, in order to make the experience easier and user-friendly. These sensors differ from *control devices* as they are persistent automatic actions, whereas *control devices* have to be given explicit commands by user intervention. The arrangement of the *MR environment* and *environment model* needs to be clarified. The *environment model* is moved to the visual cell model for two reasons. The first is that it is not a natural part of the AR actor. The second is, as described in section 9.3, the placement of the environment model was due to it being the most efficient place to hold it for all of the three actors. It was not originally placed in the virtual proxy for the mobile VR experiment, but that was a far simpler system. The current location is better as it reduces the complexity of the TeleReality actor, making him interchangeable with the AR actor and reducing the amount of data transmitted over telecommunication medium, as well as acting as a point where AR and cyber actors can access them publicly.

Another important point is the role played by the proxy CE elements. Whereas, for traditional telepresence systems, these were just used to transmit data to the actuators and from the sensors in the proxy, here they act as the fundamental means by which information is exchanged with the AR actors and the cyber actors. The CE not only starts to become a point to get access to the proxy, but also to get access to other entities in the system. This role may actually be performed by multiple CE's, as e.g., in the VC model where there is at least one CE per visual cell.

The final version of the construct model has only one change, the addition of privacy and security constructs. This was felt to be necessary, as some thought had to go into concerns over a '*big brother'* type system, as well as from concerns about communication security, especially in ad-hoc networks.

10.3 Restrictions on the system

The TeleReality system proposed here has a number of restrictions. One restriction on this system is that the remote proxy has very limited ability to carry out many physical interactions. If the TeleReality user sees a shovel in the remote environment, their virtual proxy cannot pick it up and start to use it. Therefore many missions currently carried out by teleoperation systems such as deep space exploration cannot be carried out with a TeleReality system. When interacting with the environment, the goal of the TeleReality user is to exchange information with the environment and people located in that environment. This also requires that those people have an avenue open to them in which they can participate. In other words, they must also have equipment (e.g., using ad-hoc network connections) capable of talking to the TeleReality user. Other current restrictions on the system are also technological, such as the need for efficient video codec's and the problems concerning ad-hoc security networks and the immaturity of light see-through head mounted displays.

10.3.1 Communication

The TeleReality system heavily depends on the sending of information across communication networks, both wireless and fixed. The remote user in the TeleReality system must receive video, audio and object status, plus any interactive data exchanges, i.e., file transfers. The AR user receives audio, interactive data transfers and may receive video from the remote user as well. The communication channel is then an integral part of the system and one must evaluate the requirements of that channel in implementing the system. The most likely bottleneck in any TeleReality is the wireless part of the network, as it usually has significantly more limited bandwidth than that available from a fixed network. The wireless network is most likely to be an ad-hoc network, of which Bluetooth is the only available system at the moment. These requirements will be compared to Bluetooth.

Video constitutes the heaviest load on the communication link. The optimum video image for a HMD has a resolution of 640* 480 with RGB colour scheme. Lanier (2001) says that the minimum frame-rate necessary for a smooth image is 25 fps. Without compression, the bandwidth required to transmit 2D video images would be 184.32 Mbps (640*480*3*8*25). To indicate the problem that this presents to the wireless media, Bluetooth has a bandwidth of 1Mbps (Burkhardt *et al.* 2002) and WLAN's usually have 10 Mbps, requiring a minimum compression ratio of 230:1 and 18:1 respectively. In practice, this wireless bandwidth is not always fully available, especially in WLAN, so the compression rates need to be somewhat higher.

For 3D systems, the most common approach taken is to use multiple cameras. Additionally depending upon the number of cameras in operation, the image must be captured, encoded/decoded, transmitted and processed. This creates a time lag between the capture of the image to its display to the remote users. Lanier (2001) states that the delay cannot be longer 30 to 50 milliseconds as longer delays result in fatigue, disorientation and in the worst case nausea, which does cause a problem as the transmission of images alone in his system across the continental USA had delays of 25 to 50 milliseconds. With a simple two camera system that is used to provide a stereoscopic picture, the information data rate that has to be transmitted before encoding is multiplied by two. Stereoscopic vision contains a high degree of image redundancy, so it is possible to code the images by exploiting the video redundancy (Jia *et al.* 2003). The result is still a higher degree of bandwidth than a normal 2D system. Another approach for acquiring 3D video is to attach a laser range finding device to add a depth component to each pixel and map this information onto a deformable 3D graphical mesh. With a support for a depth of 256 levels, i.e., 8 bits, and a single 2D camera system, the required bandwidth would be 1.473 Gbps (184.23 * 8).

Judging from the required bandwidth, extremely efficient video codec's would be needed that are capable of operating in real time. The best solution is to operate based on the ability to reduce data redundancy. The most efficient video codec's currently available is provided by MPEG-4 (Rauthenberg *et al.* 2000). MPEG-4 allows the codec to divide the scene into constituent parts, called audio video objects. These AVO's are encoded separately and transmitted at different bit rates. This fits in with our separation of the visual field into visual scenery objects and visual objects. The visual scenery, i.e., the background non-changing field is transmitted rarely. Localised parts of the scenery that change often, and would normally require a complete update of the scenery, are identified and represented as separate AVO objects, resulting in a smaller data rate been sent. These AVOs are usually important objects with the scenery anyway, e.g., people. Essentially then, the main background scenery is only sent once for the duration of the connection and updated rarely depending upon lighting conditions. For an empty room, then the main data is sent at the beginning of the connection and not updated afterwards, requiring negligible data rates. The same system would apply for 3D systems, however, in this case there is an added complication, if navigation or a change of view is required, and the scenery object must be updated continuously throughout the movement. If the user is looking on the same direction and moving in that direction, there is the possibility that there is a high degree of redundant data, but if they turn their head, new data must be sent in its entirety. The best case scenario for even a simple 3D system supporting navigation will have very high bandwidth requirements, hence the best available codec's will be needed.

Current research with the use of MPEG-4 and tele-immersion would seem to suggest that a workable solution to this problem is not available at the moment. In Lanier's experiment (Lanier 2001) they only achieved 2fps while requiring a bandwidth of 20-80Mbs. Rauthenberg's research group has a composite of three shaped videos into a 3D scene, but with a LAN, could only get a framerate of 12.5 fps (Rauthenberg *et al.* 2000). Cooke and his associates (Cooke *et al.* 2000), used a two-camera system to represent the video shaped object and could only achieve a framerate of 10fps over a wireless LAN. All used software encoders and decoders and note that hardware implementations may improve the performance.

In the simplest systems, only decoders would be needed. However many new multimedia handsets are supporting camera systems, so it may only be a matter of time before they will be included as well. Establishing the necessary resources needed is difficult to ascertain. For their MPEG-4 encoder, Pearmain and his colleagues needed a 200 MHz Pentium with two additional DSP chips to encode video (Pearmain *et al.* 1999). While 44 fps was possible, they state that Quarter Common Intermediate Format (QCIF) 'images at a frame rate of 8fps were used in the field trials as these could be sustained under all circumstances and were acceptable to users'. For their decoder, few details on performance were given, other than that it was carried out using a desktop computer, as no suitable laptop video card was available for it. One point should be noted about these experiments. In all of the experiments as much bandwidth as needed was provided. This is not a situation that would be available in a real application where one would have many users, not just one.

10.3.2 Registration

Registration is a very important component for Augmented Reality users and only slightly less so for TeleReality users. As mentioned earlier, the best means in which registration can be achieved is to use a hybrid system combining visual tracking with other methods such as magnetic trackers. The difficulty here is that this means placing a considerable amount of markers around each physical location both the AR and TeleReality user occupies. This is simply not practical as one can rapidly become irritated by the less than aesthetic quality of the markers. In practical terms, alternative methods are needed but no one of these provide the accuracy needed to have a successful technology, which is why so many researchers switched to visual marking systems in the first place. Otherwise, practical, accurate commercial registration systems remain unavailable to date. Although my research suggests that very accurate registration may not be necessary for carrying out every type of interaction, I would suggest that this is an area that needs to be studied more carefully.

10.4 TeleReality constructs and telepresence theory

The beginning of this thesis set out the different theories on telepresence, it is therefore useful to see how the proposed constructs compare against these theories. The basic theories on presence fall between the rationalistic approaches of Descartes, and against the Heidegger/Gibson views. In which of these categories does the system fall? The system was developed with the mind that those using it should be able to work easily with those in the remote environment, so less design focus was placed on the subjective feeling of presence. In saying that, the subject’s work will be more productive if they are quickly able to learn how to navigate and interact in the remote environment. Subjective presence is better for reducing the time it takes a person to learn how to operate in the remote environment, as the more natural it is, the shorter will be the training and the more efficient will the task accomplishment be.

As mentioned before, the theories often followed the technological systems, so having created what should be an enhanced telepresence system, do these theories still apply? The belief of the author is that the current theories still stand and a brief explanation of the various theories against the model will show why. One criticism that could be made of these theories is that they tend to look at presence from a singular user experience and quite commonly ignore the needs for a collaborative work experience. It can be argued that they do provide this in an implied manner, but Robinetts classification, which is the most detailed approach, makes no allowances at all. Additionally, the I-centric approach revolves too much around the single user experience when a great deal of telecommunication is geared towards multiple user communication.

10.4.1 Sheridan’s model

Sheridan (1991) identified three constituent parts of a telepresence system; the extent of sensory information, the control of sensors and the ability to modify environment. This research generally follows along these lines. The set of sensory information that a user can receive depends upon the complexity of the system used, but then this is also the case for traditional telepresence systems. The ability to modify the environment is different in scope, rather than principle. The TeleReality system is more strongly based on modifying the *information* environment by working with others. It can change the physical environment but only by taking control of other devices, hence reverting back to a more traditional telepresence construct model. In other words, this system can do what traditional systems can do, as well as much more. The control of sensors is probably the greatest path of divergence, but only by intent, not practice. Rather than the user directly controlling the sensors to get information, they query a central point which provides them with information collected by a set of passive sensors. With this one minor modification, the model set out by Sheridan can be applied perfectly to the constructs for an enhanced telepresence system.

10.4.2 Steurs model

Steurs (1992) descriptions of telepresence strongly match the system described in this thesis. The degree of vividness and interaction can vary considerably depending upon the complexity of the system. Vividness is constrained by the type of camera systems used (2D or 3D) and the depth is constrained by the telecommunication bandwidth, rather than any inherent limitation of the concept.

Interactivity is important in Steurs model. The system is designed so as to be realtime, although some non-realtime variations, such as static systems, can also be used. Both range and mapping are difficult, and if interaction is concerned with the modification of information, then the system is very strong. If its a case of physical manipulation of the remote environment by the use of robots, then the system is somewhat weaker. In the classification of a TeleReality system along the lines of vividness and interactivity, then it should belong towards the higher end as shown in figure 56.

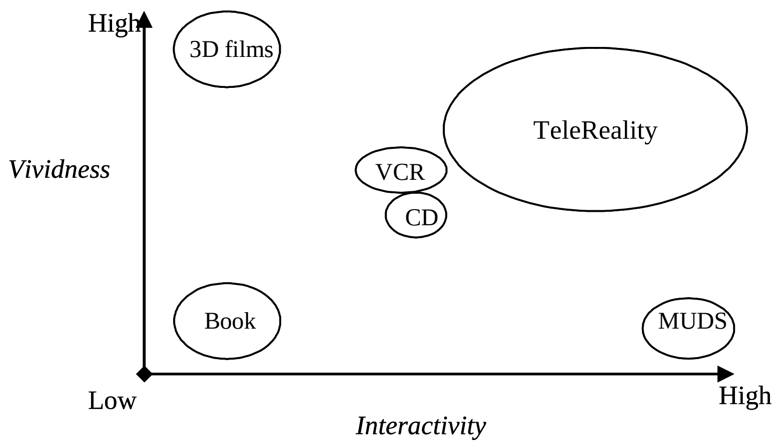


Fig. 56. Classing TeleReality in Steurs vividness and interactivity chart.

10.4.3 Zeltzers AIP Cube

Zeltzers (1992) Autonomy, Interaction and Presence cube, while conceived more for VE and CVE systems, work very well for the TeleReality system, if one realises that the model being composed is based on the real world and not a synthetic one. The remote environment is, by its nature, strongly autonomous. The introduction of cyber actors makes this particularly true, as does the comings and goings of AR users. These actors can be considered as agents that react to stimuli or have independent thoughts, as described by Zeltzer in his definition of autonomy.

Zeltzers view on interaction determined the degree with which the user can interact with the model at run-time. As this model meant more about the interaction of information about the model, it fits closer to the TeleReality system as it does not consider physical interaction.

Presence determines the degree in which the user receives and can control sensory information, which is very similar to Sheridan's approach. This view of presence is in how the technology can help the user interact with the remote environment. This point also matches with the goal of creating a system that assists the user in working in the remote environment. Therefore in the AIP cube, the goal of the TeleReality system is to have a high degree of interaction and presence, while realising that the remote environment, by virtue of been real, is strongly autonomous. Therefore the values are likely to range from any value of (1,0,0) to (1,1,1). The value of (1,1,1) would represent the point of perfect telepresence.

10.4.4 Robinetts classification

Robinetts (1992a, 1992b) classification is a useful means for looking at a TeleReality system, but a deficiency in the classification is that it does not differentiate systems that are clearly single or multi user. Table 15 shows how a TeleReality system would fit into Robinetts classification. There are two situations here, one whereby the cyber actors have some actuators to affect the physical environment and one where it is strictly informational exchange.

Table 15. Robinetts classification as applied to TeleReality systems and compared with telepresence systems.

	TeleReality (actuator support)	TeleReality (no actuators)	VR based excavator	Tele-operator	Remote pilot aircraft
Causality	transmit	transmit	transmit	transmit	transmit
Model Source	reconstruct	reconstruct	reconstruct	scan	scan
Time	1-to-1	1-to-1	1-to-1	1-to-1	1-to-1
Space	remote	remote	remote	remote	remote
Super-position	merge	merge	isolated	isolated	isolated
Display	HMD/graphics	HMD/graphics	screen/ graphics	HMD, force feedback	screen
Sensor	cameras	cameras	cameras and radar	camera	camera
Action measure	gesture, mouse	gesture, mouse, joystick	joystick	force feedback arm	joystick
Actuator	none	actuators in environment.	excavator	robot arm	flap, actuators in aircraft

The main differences here are that the model source is scan based. That means the information is taken from the environment and processed before display to the user. TeleReality has active superposition that merges the synthetic objects representing artefacts in the environment with real time video. This works to facilitate interaction for the users. The other main difference is that the interface to the system is different. A gesture based system may be more appropriate for interacting with these synthetic objects, whereas others need a joystick in order to physically control a device. Lastly, whether a TeleReality system has an actuator in the environment is entirely up to the implementation of the cyber actor, so one can have systems that have zero to many actuators.

10.4.5 Milgram s RV Continuum

In terms of the RV continuum, TeleReality is located to the far left of the continuum as shown in figure 57. If the environment contains no graphical content, then, within this definition, it is considered as the real environment. Once, however, graphical data is added to the telepresence system, then the system, according to the RV continuum, would reside in the area to the left on the Continuum. It is this part of the continuum that contains the region of interest in TeleReality. The RV Continuum was devised with the idea of display systems, e.g., monitors and closed HMD. See through HMD are not considered. However, it does form a nice classification measure whereby the TeleReality concept should fit in relation to the general research area.

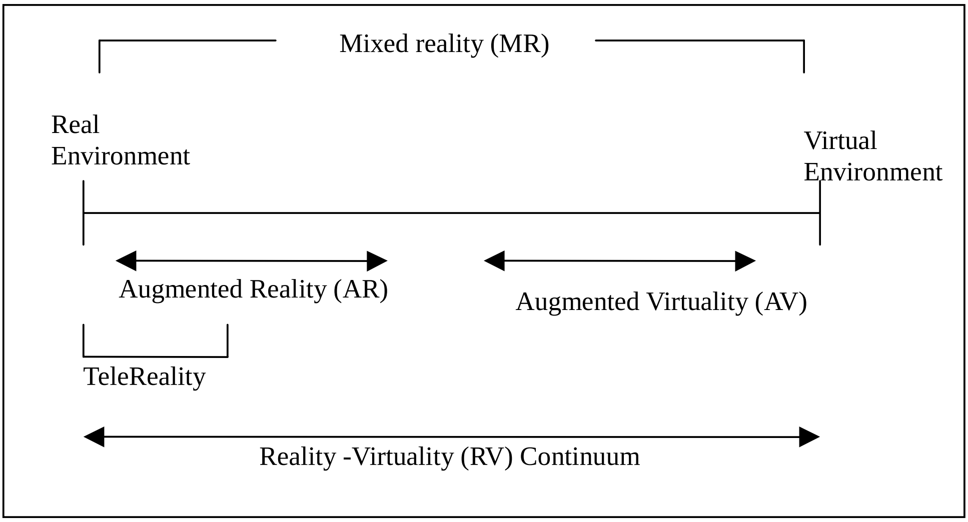


Fig. 57. TeleReality in the Reality -Virtuality Continuum

10.4.6 WSI Reference model

The WSI reference model is still under construction. The authors accept that there is still considerable work needed on ironing out a formal description of the semantics of the reference points, a methodology to define communications between via the reference points, a functional distribution and allocation to building blocks, dynamic representations of major transactions, and more details of possible business models. (Arbanowski *et al.* 2002).

These are valid points, as on reviewing the deliverables to the WWRF, one can notice numerous duplications of concepts across the reference points, with almost each layer having applications, services and presence services. Presence in the WSI model should not be confused with presence as discussed here. It is a telecommunication term that acts as a simple one/off state. One is either present in the system or can work there, or one is not. Despite these problems, it will likely be used as the reference model for future 4^{th} generation ubiquitous computing, due to its backing from the WSI and WWRF communities.

The greatest difficulty in using the WSI reference model in this work has been the I-centric approach taken. While it seemed a popular and reasonable idea at its inception, when trying to apply it to this work, problems started to emerge. The I-centric model is perhaps the most popular telecommunication theory that does not intrinsically incorporate communication between people. It is geared towards a single user communicating with the surrounding environment and developing services and applications that help them. All previous telecommunication models expressly support communication between people.

The WSI reference model diagrams are, in this way, unique that they only show a single person as the focus in a communication system. While I, therefore, use the WSI model to represent the arrangement of artefacts to the user, I had to introduce the concept of TeleReality, AR and Cyber actor shown in figure 22 and 23 to reflect the importance and support for collaboration, communication and cooperation between users.

The author accepts that the WSI reference model will set the stage for future thinking on communications in terms of design, teaching, research and development, but also feels that it is likely to unsatisfactorily address the primary need of telecommunications systems, the need to communicate with our peers.

10.5 Critical Analysis

This thesis addressed the research question through the use of the constructive research method. Under that method, it is sufficient that the constructs are both logical and cohesive in order to be acceptable, as well as being different from current systems. This thesis has shown how a telepresence system can be implemented with a virtual proxy, and shows how the constructs need to be arranged for this to occur. There are a number of points of criticism that can be made on this work.

The first of these is that this system is not enhanced as physical interaction is not an inherent part of the system, but rather, it is up to the whim of the local environment in providing a sufficient cyber actor. This is indeed the case. However, it is important to point out that it is task completion that is essential. Any number of the task's that are carried out in the information society reflect the flow of information, and for this, physical interaction is not necessary if it can be supplanted with a combination of intuitive user interfaces and communication systems. Where physical interaction is needed, then an appropriate cyber actor should be provided.

Some questions may arise from the identification and arrangement of the constructs. While the constructive research approach describes at some length the research process, it does not describe how one identifies the different constructs. Constructs are building blocks, but the question is, what level does one consider to be building blocks? Is it at the system level, architecture and object level? The author was unable to find a satisfactory guide in this process. Hence the system level view was taken. The author looked at the system components that are essential, disregarding the intricacies of each design, to find what was common and not common. The works of Sheridan, Kuzuoka and Milgram were essential for this. It can be argued, that in doing so, other constructs were ignored, but I believe that the constructs identified here are irreplaceable for the various analysed systems. In the TeleReality construct model, it is these constructs that are the minimum requirement, other constructs added would be supplemental.

Once chosen, the arrangement of the constructs can be open to question. A system that supports information transfer must have an informational construct in each of the participating actors. The most difficult placement is the environment model. However, once one starts to have a large multi-partition system (and consequently large data and processing requirements in the actor), concerns over security access and wishes to have the TeleReality actor as close in design to an AR actor as possible, make it more obvious that this construct should be located in the environment, where resources are more cheaply and readily available.

The technical complexity and lack of resources had two consequences, first the system was never implemented as fully as one would have wished. Second, there was not any full usability test done on the system. These are two significant deficiencies, as they prevent the conclusive result on the long term feasibility of the TeleReality concept. They do not, however, affect the core constructs. The test sorely lacked an ad-hoc radio element for the testing, as much could have been gained from that, although the author does not believe that would have changed the underlying constructs, but would have more of an affect on the detailed visual cell model and implementation process that was followed. The user interface that was implemented provided interesting ideas and concepts, as well as problems, but these did not change the underlying constructs either.

Considering the work done so far, the author anticipates that the following areas present the greatest difficulty at the moment, the scalability of the system, object clutter, object placement, light conditions, different HMD FOV’s and quality, health and safety issues. While the system is a multi-user system, the author has not attempted to calculate how many users would be able to participate synchronously. The capacity of the system is heavily dependent upon the numbers involved, how many TeleReality actors there are, what broadcast protocol is used, how many cameras are used and the resulting size of the encoded image and how much bandwidth does the system have. These parameters are highly variable and, hence, the author does not attempt to give rough and likely inaccurate, figures.

From the user interface experiments, it became obvious to the author that object clutter would most likely be a serious problem. The reason is that in the visual field, the objects must be easily visible and yet not obstruct gaze awareness. This means that the objects must inhabit a space much smaller than that available. Now in the case that there are lots of objects in the space, they either obscure each other, or overflow into areas that disrupt the users interaction with other people. Essentially, the objects must be designed and arranged so that they are usable and this places a limit on the number that can exist. What that number is was not explored by the author, though he believes it is a function of the object size and space available. This observation seems to be supported by Cockburn and McKenzie (2002) who noted that the subjects’ performance deteriorated in both physical and virtual systems as their freedom to locate items in the third dimension increased. They also noticed that subjective measures reinforced the performance measures, indicating that users found interfaces with higher dimensions more cluttered and less efficient.

Lighting has a detrimental affect on any vision system employed, for the simple reason that it is not a constant factor in most work environments. While recognising the impact it had in the two experiments, the author did not examine it closely. It can affect anything from cell size in outdoor areas to unwanted reflections. Lighting is one of those environmental effects which affect the cell, just as rain and snow affect radio cells. The last area of interest, which the author did not examine in great detail, is the issue of HMD’s. It is, however, perhaps the singular most important part of the system. HMD’s can have health and safety issues, causing cybersickness. In TeleReality systems, it causes an additional danger, as the TeleReality actor cannot see their local area while ‘visiting’ another real world location. Therefore they must be stationary. Whatever extra health and safety problems that may arise are not known by the author.

10.6 Summary

This chapter provides a discussion on the constructs in terms of the research question, analysis of the construct, restrictions on the technology and a framing of the subject against telepresence theory. The research question set out to use a virtual proxy to support interaction, a multi-user system and navigation. How these characteristics are supported by this research work is described here. The evolution of the basic constructs, from traditional telepresence systems to the final TeleReality model is also described, as is the reason for that evolution. Figure 55 visibly shows the location of each construct, as well as each larger actor in the system. The TeleReality system as described here faces a number of technical difficulties if it is to be accepted as a working technology. Problems associated with the heavy demands of video quality and the affect video codec’s have on bandwidth are described, with the primary problem being the need for good stereoscopic coding algorithms at a minimum. Problems associated with network lag are also described. Additional problems in relation to the search for a practical registration system and the deficiencies of ad-hoc security protocols are also covered. Lastly, the TeleReality system is compared against the telepresence theories of Sheridan, Steur, Zeltzer, Robinnet and Milgram. The result of this is to show that the concept of TeleReality fits into these theories, and as a result, shows that it is on a solid theoretical basis. The other conclusion is that it does not constitute a new theory on telepresence, but rather an alternative means to solve existing conceptual and implementation problems.

11 Conclusions

This thesis studied the feasibility of using virtual proxies to enhance telepresence systems. To support a multi-user collaborative work environment, the key areas of *navigation*, *interaction* and *multi-user support* are addressed. The thesis set out to study and answers the research question '*What form should an enhanced telepresence system relying on virtual proxies take so that it is multi-user, interactive and navigable'*

The virtual proxy itself is the user's presence within the telepresence environment and this can be best described as a free floating camera, capable of offering any image perspective in the remote environment. The three requirements that were placed on the virtual proxy are that it should be multi-user, support a rich set of interactions and support user navigation of the environment. Also, as the solution involved the use of video cameras, a question arose as to the impact such a system would have on the security and privacy of the participants. The author coined the term *TeleReality* to distinguish this form of enhanced telepresence system using virtual proxies from other solutions that rely on physical proxies. To address the research question, modification to the existing constructs of telepresence systems have been proposed and shown to fit logically and cohesively together by concept and experiment.

11.1 The TeleReality Constructs

The constructs for a TeleReality system are organised into four main components, the *virtual proxy*, the *TeleReality actor*, the *AR actor* and the *Cyber actor*. The TeleReality actor is the person who is accessing the remote space. The AR actor is the person who physically exists in the remote. The cyber actor is an independent computer controlled entity that exists in the remote environment. The constructs for both the TeleReality and AR actor are similar. What differs is the context in which they receive the images. The TeleReality user gets images from a remote space whereas the AR user gets them from their local space. In either case, they both receive the same information data. The TeleReality user must actively request the data based on exercising their control devices and whatever additional relevant information they get from their own environment sensors. The cyber actor can be more complex, but at a minimum it needs an information source (data), intelligence to react to stimuli (behaviour) and a communication element to connect with other actors. At its most complex, it can support physical interaction, including the capabilities of a full robot.

The largest changes come from the introduction of the virtual proxy, which creates a new environment model construct. This construct is responsible for receiving information from the environment sensors and from the various actors, and organising that information by constructing a model that can be queried by all users of the system. The environment model is identical in concept to that of VE models, but in this case represents the real world, not a computer generated model.

The last set of constructs of interest concern the importance of security and privacy. A system that is designed with the concept of ubiquitous computing as a base must ensure that those communication elements generate confidence in the secure transfer of information. Privacy issues arise from the use of multiple camera systems, giving rise to a feeling of a big brother system. These constructs are therefore essential to address potential problems in the system. These basic constructs then, address the issues of navigation, interaction, security/privacy and multi-user support.

11.1.1 Navigation

Navigation means that each user can travel through the remote environment. This means that in a TeleReality system they can receive multiple camera viewpoints based on their position and orientation. To accomplish this, static cameras are placed through the world (environment sensors) and image processing software calculates the correct perspective images (environment model). This large camera system is sub/divided into *partitions*. To maintain a continuous image perspective for navigation, users can move from one partition to another in a process called *handover*. Handover can be accomplished based on the user’s position, their movement and most importantly the perspective (field of view) of the participants. Partitions are connected by *portals*, which provide the address to the next partition. A combination of partitions can be combined into a world system. Each partition can be addressed by the use of a symbolic address combined with geographical data.

The TeleReality world can be constructed with varying degrees of sophistication. A 3D or 2D video system can be utilised. A 3D system would obviously provide the greatest sense of flexibility and presence, but 2D systems can also be quite flexible, if somewhat more constrained. The world type can also be static or dynamic. A dynamic system updates the world in real-time, whereas a static world is generally stored in an audio-visual database and updated more rarely. While a dynamic system is the preferred option, static systems can work well in areas that do not involve working with people local to the area.

The TeleReality system requires a heavy use of good quality video transmission that may be too excessive. To counteract this, the video images should be divided into two parts, the visual scenery which represents the basic non changing nature of the environment, and the visual object which represent the interactive objects. Visual scenery objects are rarely updated, as generally very little happens. Visual objects are sent once, and afterwards, only their current state is changed.

11.1.2 Interaction

The environment is also defined by the degree in which it is interactive. This thesis proposes the *TeleReality Interactive Principle*. A measure is taken of the number of possible interactions in the environment by all users, TeleReality and AR, and it determines how many can be used by both TeleReality and AR users. The higher the measure, the richer the interactive environment and the more immersive the system is for the TeleReality user.

Interaction is achieved by the flow of sensory data to the users and the sending of command data. This circular flow of information means that interaction has to be supported to some degree by all of the constructs. The major parts of the interaction scheme are also described. These are the *actors*, who initiate actions/events in the systems, *artefacts*, which are representations of the services and applications within the system and *transactions*, which define how communication is carried out between actors and artefacts and between artefacts. The actors are the TeleReality actor, the AR actor and the cyber actor, usually a system entity. Each artefact belongs to at least one actor. The actors need to be able to copy, modify, create and destroy artefacts. Each artefact occupies up to two locations, the location of the code in a device and the location of the artefact in the visual field.

The relationship between the actors, artefacts and transactions are described by the use of a modified multi-sphere model, which is a geographically based system. Actors exist in a geographical space and each is given a personal user sphere in which resides all of their personal artefacts. Common artefacts are found in the local area sphere, and access is also provided for access to artefacts in the global cyber world.

11.1.3 Security/privacy

Providing security and privacy for the system is imperative given the possibility for big brother scenarios. To support privacy it is proposed that an organisational and technological solution is required. From an organisational point of view, *trust relationships* are established between the users of the system and the owners, indicating the rights and conditions in which the system is used. The users of the system should have *awareness* in the system, i.e., know who is sharing the space with them. They should have *control* over who can communicate with them, and *access* control to limit the amount of services and freedoms of the visitors. They should expect *integrity* in the system, i.e., knowing that the person they are dealing with is who they claim to be.

Architecturally, to support control, a point to point phone based model is used to allow users to have control over who is contacting them. Additionally, to prevent possible exploitation of the ad-hoc networks, the users may not allow other ad-hoc elements on their network while they are connected to other users. To prevent stolen keys, the users have a *network key* for accessing the fixed network and an *ad-hoc* key for dealing with other ad-hoc user networks. Also, ad-hoc users in the system should use the fixed network as an access point to 3rd party certification bodies to authenticate and certify other users and information.

11.1.4 Multi-user support

The advantage of the use of a fixed camera position and the image processing software is that the same images are sent to each person using the system, and it is then that their correct perspective is calculated. The use of radio elements such as ad-hoc networks also means that more than one person can use the system. Therefore the solution proposed is a multi-user system.

11.1.5 The Visual Cell Model

The TeleReality constructs were evaluated by determining their importance as basic building blocks of a TeleReality system. The evaluation method used was to identify the potential constructs, then develop a model and instantiate this model. By executing the model, any weakness or unanticipated absent constructs were identified and the original constructs were revised based on this evaluation, with a new revised model created instantiated and evaluated. The final outcome of this stage was the proposed visual cell model, which was developed using the TeleReality constructs. A partial implementation of this model was carried out within the PAULA project.

11.2 Future work

The work described in this thesis offers considerable opportunities for further research projects. While the basic constructs have been identified, the detailed and efficient implementation of those constructs have not being addressed, indeed they need not be as part of the chosen research method. Currently, the author is addressing the user interface aspect of collaborative telepresence and augmented reality systems. This is important, and while this thesis identifies interaction as being critical to presence and TeleReality, the mechanics of how the users interact has not been studied. This topic offers the potential for considerable research effort.

The interaction capability between multiple TeleReality actors has not been heavily studied in this thesis topic, as neither is the form of their avatar in the remote real world. These are optional features and are not part of the core constructs. The core constructs know where the remote user is supposed to exist in the environment and informs others, but what representation acts for them is a question of implementation. The form of this implementation is useful to study, as the richness of the solution can enhance non-verbal interaction.

Navigation has also been an important element in this work. However, the technology to implement it to the extent that the author wishes is not available at the moment. Future work on navigation must address three factors, a lightweight implementation with good support for handover, the means by which the remote user can initiate movement when they themselves must be stationary for safety reasons, and the degree to which navigation enhances presence and users' performance.

The Cyber actor offers significant sources of research. What specific tasks should these autonomous agents perform in the environment and how do they interact with other? Examples of cyber actors could be robots, memory assistance agents, automatic bus payment systems. Indeed the number of potential applications is probably larger than the author can think of. How will these agents be implemented, who will own them and under what conditions will they play a part? The author is looking at many of these research questions and has provided a number of future scenarios under the EU SSA MOSAIC project on future mobile work, which can be found at http://www.mosaic-network.org (Hickey 2004).

Many difficult technical issues remain, among them more efficient video codec's, communication broadcast protocols, the capacity of a system, acceptable HMDs of sufficient quality and improved registration.

11.3 Applications

TeleReality is unlikely to be widely implemented or used until satisfactory cost effective equipment is developed. However, there are a number of application areas where it might find extensive use. Other applications may suffice by utilising part of the system. Some of the potential applications are

Videoconferencing: Videoconferencing describes the situation where remotely located people can establish a video connection together, communicate, exchange information easily and use available objects irrespective of location. As an example, a remote person may want to give a presentation to a group of people in a meeting room. By using a TeleReality system, he can pass around a copy of the slides to each person by giving them a synthetic object representing the media file, then he can select the rooms projector and place his presentation file upon it, giving his normal presentation. They can move around the room, changing their perspective to talk with different people. This use of a TeleReality system can enrich the communication between people in remote areas.

Tele-assistance: Increasingly, world-wide specialists in a field are consulted by audio/visual telecommunication connections for their opinions on unusual and complex problems, especially in medical diagnosis. In tele-surgery, information can be exchanged by passing around an electronic version of the information. TeleReality would also give the remote specialists the ability to look around, as well as follow the experts on their rounds, or as they go between different sites. With TeleReality, as many specialist and students can take part as possible. Other examples include maintenance, distance learning, construction and disaster emergency services.

Tele-tourism: Tele-tourism allows a person to visit remote locations at a cheaper price than going there. It would enable people to visit art museums in foreign countries, remote inaccessible locations and archaeological digs. From these sites they can acquire information by interacting with local information points, i.e., synthetic objects that act as a front end for audio/visual files on the subject or contain links to more information.

11.4 Authors contribution

The work described in this thesis has been entirely the work of the author unless otherwise stated. The author received some help in the setting up of the mobile VR experiment, however, the concept, execution and analysis were done by the author. The author also received some help in the construction of the visual cell implementation by students working under my direction. The critical analysis on the limitations of telepresence, the analysis and derivation of TeleReality constructs, the detailed look of the issues of navigation, interaction, privacy, security and the final critical analysis were all performed by the author.

11.5 Final Thoughts

The significance and contribution of this thesis is the introduction of an enhanced telepresence system that supports multiple remote users while providing a rich interactive environment and a degree of presence. TeleReality is a system that was envisioned as a means of enhancing telepresence by removing the restrictions that limit its acceptability to a wider audience. Enhancing Telepresence requires a multi-disciplinary approach. This thesis work looks at a total solution, by defining all of the necessary parts of a successful technology, it aims to focus the research task by defining the basic elements and areas that the research field must acknowledge. TeleReality will be a technology of the future and this research aims to identify all of the relevant technologies and constructs that will be needed for a mature system.

References

Aggarwal S, Banavar H, Khandelwal A, Mukherjee S & Rangarajan S (2004) Accuracy in dead-reckoning based distributed multi-player games, Proceedings of ACM SIGCOMM 2004 workshops on NetGames '04: Network and system support for games, Portland, Oregon, USA.

Antoniac P, Pulli P, Kuroda T, Bendas D, Hickey S & Sasaki K (2002) Wireless User Perspectives in Europe: Handsmart Mediaphone Interface, Wireless Personal Communications 22(2): 161-174.

Arbanowski S, Pabst R, Mössner K, Pulli P, Zheng X, Raatikainen K, Uusitalo M, Lipka M, Ott K & Schieder A (2002) IST-999-12300 Wireless Strategic Initiative (WSI) Project, Deliverable D9: reference model for the Wireless World [online], [Cited 11 February 2005] Available from http://www.ist-wsi.org/D9_final.pdf .

Arbanowski S, Van der Meer S, Steglich S & Popsecu-Zeltin R (2001) The Human Communication Space: Towards I-centric Communications, Personal and Ubiquitous Computing 5(1): 34-37.

Avancha S, Joshi A & Finin T (2002) Enhanced Service Discovery in Bluetooth, Computer 35(6): 96-99.

Azuma R T (1997) A Survey of Augmented Reality, Presence 6(4): 355-385.

Bajura M & Neumann U (1995) Dynamic registration correction in video-based augmented reality systems, Computer Graphics and Applications 15(5): 52-60.

Ballantyne J, Greenspan M & Lipsett M (1997) Virtual Environments for remote operations, Proceedings of the ANS 7th Topical Meeting on Robotics and Remote Systems, Augusta, GE.

Banatre M, Coudrec P & Weis F (2000) Spontaneous Communication, Proceedings of the IST Mobile Communications Summit, Galway, Ireland.

Begault D R (1994) 3D Sound for Virtual Reality and Multimedia, Academic Press, Cambridge, MA.

Bellotti V & Bly S (1996) Walking Away from the Desktop Computer: Distributed Collaboration and Mobility in a Product Design Team. Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work, Cambridge MA, USA.

Benford S, Brown C, Reynard G & Greenhalgh C (1996) Shared Spaces: Transportation, Artificiality and Spatiality. Proceedings of ACM Conference on CSCW, Boston, Massachusetts.

Bergmann N & Mudge J (1994) Automated Assistance for the Telemeeting Lifecycle. Proceedings of the 1994 ACM conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, USA.

Bergqvist J, Dahlberg P, Ljungberg F & Kristoffersen S (1999) Moving Out of the Meeting Room: Exploring support for mobile meetings. European Computer Supported Cooperative Work '99, Copenhagen, Denmark.

Berruto E, Gudmundson M, Menolascino R, Mohr W & Pizarroso M (1998) Research Activities on UMTS Radio Interface, Network Architectures and Planning. IEEE Communications Magazine 36(2): 82-95.

Bertelsen O & Nielsen C (2000) Augmented Reality as a design tool for mobile interfaces. Proceedings of the ACM Designing Interactive Systems, New York, USA.

Bhatia P & Uchiyama M (1999) A VR-Human Interface for Assisting Human Input in Path Planning for Telerobots. Presence 8(3): 332-354.

Bobick A, Intille S, Davis J, Baird F, Pinhanez C, Campbell L, Schutte A & Wilson A (1999) The KidsRoom: A perceptually-Based Interactive and Immersive Story Environment. Presence 8(4): 369-393.

Bowers J, O'Brien J & Pycock J (1996) Practically Accomplishing Immersion: Cooperation in and for Virtual Environments. Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work, 380-389, Cambridge MA USA.

Brasche G & Walke B (1997) Concepts, Services, and Protocols of the New GSM Phase 2+ General Packet Radio Service. IEEE Communications Magazine 35(8): 94-104.

Broll W (1995) Interacting in Distributed Collaborative Environments. Proceedings of IEEE Virtual Reality Annual International Symposium - VRAIS'95, Los Alamitos, USA.

Burbidge D & Murray P (1989) Hardware improvements to the helmet-mounted projector on the Visual Display Research Tool (VDRT) at the Naval Training Systems Centre. SPIE Proceedings Volume 1116: Head Mounted Displays.

Burkhardt J, Henn H, Hepper S, Rindhoff K & Schack T (2002) Pervasive Computing: Technology and Architecture of Mobile Internet Applications. Pearson Education, Addison-Wesley, USA.

Capin T K, Esmerado J & Thalmann D (1999) A dead-reckoning technique for streaming virtual human animation. IEEE Transactions on Circuits and Systems for Video Technology, 9(3): 411-414.

Carlsson C & Hagsand O (1993) DIVE - a Multi-user Virtual Reality System. Proc. IEEE Virtual Reality Annual International Symposium (VRAIS'93) Seattle, Washington, USA.

Coates G (1992) Program from Invisible Site- a virtual show, a multimedia performance work. Presented by George Coates Performance Works, San Francisco, CA, March 1992.

Cockburn A & McKenzie B (2002) Evaluating the Effectiveness of Spatial memory in 2D and 3D Physical and Virtual Environments. The CHI 2002 Conference on Human Factors in Computing Systems, Minneapolis, USA.

Cooke E, Karl M, Kauff P & Schreer O (2000) Realtime View adaptation of Video Objects in 3-Dimensional Virtual Environments. 3rd International workshop on Presence, Delft, Netherlands.

Darken R, Allard T & Achille L (1998) Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces: An introduction. Presence 7(2): 101-107.

Darken R, Cockayne W & Carmein D (1997) The Omni-Directional Treadmill: A Locomotion Device for Virtual Worlds. ACM Symposium on User Interface Software and Technology, Alberta, Canada.

Davey B, Comeau R, Munger P, Pisani P, Lacerte D, Olivier A & Peters T (1994) Multimodal interactive stereoscopic image-guided neurosurgery. Proceedings of Visualisation in Biomedical Computing (VBC) 94, Rochester, Minnesota, USA.

Dourish P & Bly S (1992) Portholes: Supporting Awareness in a Distributed Work Group. Computer Human Interaction '92, Monterey, California, United States.

Ebrahimi T & Horne C (2000) MPEG-4 Natural Video Coding - An overview. Signal Processing: Image Communication 15(4-5): 365-385.

Edigo C (1988) Video Conferencing as a technology to support group work: A review of its failure. CSCW 88: 2nd Conference on Computer Supported Cooperative Work, Portland, USA.

Edwards E, Rolland J & Keller K (1993) Video see-through design for merging real and virtual environments. Proceedings of the IEEE Virtual Reality Annual International Symposium '93, Seattle, USA.

Elrod S, Bruce R, Gold R, Goldberg D, Halasz F, Janssen W, Lee D, McCall K, Pedersen E, Pier K, Tang J C & Welch B (1992) LIVEBOARD: A Large Interactive Display Supporting Group Meetings, presentations and remote Collaboration. CHI 92: Conference on Human Factors in Computing Systems, Monterey, California, United States.

Emura S & Tachi S (1994) Compensation of time lag between actual and virtual spaces by multi-sensor integration. Proceedings of the 1994 IEEE International conference on multisensor Fusion and Integration for Intelligent Systems, Las Vegas, USA.

Endo T, Katayama A, Tamura H, Hirose M & Tanikawa T (1999) Cybercity Walker - Layered Morphing Method. Proceedings of HCI International 99, Munich, Germany.
Fahlen L, Stahl O, Brown C & Carlsson C (1993) A Space Based Model for User Interaction in Shared Synthetic Environments. Proceedings of the SIGCHI conference on Human factors in computing systems, Amsterdam, Netherlands.
Fairchild K, Lee B, Loo J, Ng H & Serra L (1993) The heaven and earth virtual reality: Designing applications for novice users. IEEE Virtual Reality Annual International Symposium(VRAIS), Seattle, Washington, USA.
Finkelstein A, Jacobs C & Salesin D (1996) Multiresolution Video. SIGHGRAPH '96, New Orleans, USA.
Fish R, Kraut R, Root R & Rice R (1992) Evaluating Video as a technology for Informal Communication. Proceedings of the SIGCHI conference on Human factors in computing systems, Monterey, California, United States.
Fitzpatrick G, Kaplan S & Mansfield T (1996) Physical spaces, virtual places and social worlds: a study of work in the virtual. Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work, Boston, Massachusetts, United States.
Fluhrer S, Mantin I & Shamir A (2001) Weakness in the key scheduling algorithm of RC4. Selected Areas in Cryptography: 8th Annual International Workshop, SAC 2001, Toronto, Ontario, Canada.
Fuchs H, Bishop G, Arthur K, McMillan L, Bajcsy R, Lee S, Farid H & Kanade T (1994) Virtual Space Teleconferencing Using a Sea of Cameras. Proceedings of the First International Symposium on Medical Robotics and Computer Assisted Surgery, Pittsburgh, USA.
Gagnon E, Greenspan M & Ballantyne J (1997) Real-time surface following for teleoperation in virtual environments. Proceedings of SPIE Conference on Telemanipulator and Telepresence Technologies IV, Pittsburgh, USA.
GARMIN Corporation (2000) GPS Guide for Beginners web [online]. [Cited 3 January 2005], Available from http://www.garmin.com/manuals/GPSGuideforBeginners_Manual.pdf.
Gaspar J, Decco C, Okamoto J & Santos-Victor J (2002) Constant Resolution Omnidirectional Cameras. OMNIVIS'02 Workshop on Omni-directional Vision, Copenhagen, Denmark.
Gausemeier J, Fruend J & Matysczok C (2002) AR-Planning Tool -Designing Flexible manufacturing Systems with Augmented Reality. Eight Eurographics Workshop on Virtual Environments, Saarbrucken, Germany.
Gaver W, Moran T, MacLean A, Lovstrand L, Dourish P, Carter K & Buxton W (1992) Realizing a Video Environment: EUROPARC's RAVE system. Computer human Interaction (CHI) '92, Monterey, California, United States.
Gaver W, Sellen A, Heath C & Luff P (1993) One is not enough: Multiple Views in a media space. Proceedings of the SIGCHI conference on Human factors in computing systems, Amsterdam, Netherlands.
Gaver W, Smets G & Overbeeke K (1995) A Virtual Window On Media Space. Proceedings of the SIGCHI conference on Human factors in computing systems, Colorado, United States.
Geist R, Stinson T, Schalkoff R & Gurbuz S (1997) Autonomous Virtualization of Real Environments for Telepresence Applications. Presence 6(6): 645-657.
Gibbs S, Arapis C & Breiteneder C (1999) TELEPORT - Towards immersive copresence. Multimedia Systems 7(1): 214-221.
Goldberg K (2000) The Robot in the Garden: Telereobotics and Telepsitemology in the Age of the Internet. MIT Press, Cambridge, Massachusetts, USA.
Greaves A (1994) State of the art in nuclear telerobotic: focus on the man/machine connection. Transactions of the American Nuclear Society 71(1): 520-521.
Greenbaum P (1992) The lawnmower man. Film and video, 9: 58-62.
Greenhalgh C & Benford S (1995) MASSIVE: A Collaborative Virtual Environment for Teleconferencing. ACM Transactions on Computer-Human Interaction 2(3): 239-261.
Greenhalgh C & Benford S (1999) Supporting Rich And Dynamic Communication in Large-Scale Collaborative Environments. Presence 8(1): 14-35.
Hansmann U, Merk L, Nicklous M & Stober T (2003) Pervasive Computing. Springer Professional Computing, Heidelberg, Germany.

Harter A & Hopper A (1994) A Distributed Location System for the Active Office. IEEE Network Magazine 8(1): 62-70.

Held R & Durlach N (1992) Telepresence. Presence 1(1): 102-112.

Hickey S (1999) Distributed Meeting Service to Support Mobile Users. Fifth International Conference on Intelligence in Networks (SMARTNET'99), Bangkok, Thailand.

Hickey S (2002) Security issues for evolving ad-hoc networks. European Intensive programme on Information and Communication Technologies security, Oulu, Finland.

Hickey S (2004) Software Audit Scenario [online].http://www.mosaic-network.org/bscw/bscw.cgi/0/64583, [Cited 2005,7.1.2005]

Hickey S, Manninen T & Pulli P (2000) TeleReality - The next step for Telepresence. Proceedings of the World Multiconference on Systemics, Cybernetics & Informatics (SCI 2000), Florida, USA.

Hickey S & Pulli P (2001) Developing a User Centric Multisphere Service Model for Mobile users using a Visual Cell Platform. Proceedings of the WWRF kick-off Meeting, Munich, Germany.

Hindmarsh J, Fraser M, Heath C, Benford S & Greenhalgh C (1998) Fragmented Interaction: Establishing mutual orientation in Virtual Environments. Computer Supported Collaborative Work 98, Seattle Washington USA.

Hindus D, Ackerman M, Mainwaring S & Starr B(1996) Thunderwire: A field Study of an Audio-Only Media Space. Computer Supported Cooperative Work '96, Cambridge, MA, USA.

Hirose M, Tanikawa T & Endo T (1999) Building a Virtual World from the Real World. Mixed Reality, Merging Real and Virtual Worlds, Ohmasha Books, Tokyo, Japan.

Holloway R (1995) Registration errors in augmented reality. Department of Computer Science, North Carolina at Chapel Hill, USA.

Hudson S & Smith I (1996) Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. Computer Supported Cooperative Work '96, Cambridge, MA, USA.

Iltanen M, Kosola H, Palovuori K, Vanhala J & Tuokko R (1998) Positioning and localization system for an indoors mobile robot platform based on spread spectrum technology. Proceedings of the Eighth International Symposium on Measurement and Control in Robotics (ISMCR'98), Prague, Czech Republic.

Isaacs E, Morris T & Rodriguez T (1994) A Forum for Supporting Interactive Presentations to Distributed Audiences. Proceedings of the Conference on Computer Supported Cooperative Work CSCW'94, Chapel Hill, NC, USA.

Isaacs E, Morris T, Rodriguez T & Tang J C (1995) A Comparison of Face-To-Face and Distributed Presentations. Computer-Human Interaction Mosaic of Creativity (CHI '95), Denver, USA.

Ishii H & Kobayashi M (1992) ClearBoard: A seamless Medium for Sharing Drawing and Conversation with eye contact. Computer Human Interaction '92, Monterey, California, United States.

IST (2000) Book of Visions 2000 [online]. [Cited 11.2.2005] Available from : http://www.wireless-world-research.org/general_info/Bookofvisions/BookofVisions2000.pdf.

Jia H, Gao W Lu Y (2003) Stereoscopic video Coding based on Global Displacement Compensated prediction. Fourth International Conference on Information, Communications & Signal Processing, Singapore.

Julier S, Baillot Y, Brown D & Lanzagorta M (2002) Information Filtering for Mobile Augmented Reality. IEEE Computer Graphics and Applications 22(5): 12-15.

Kac E (1993) Telepresence Art. In Kriesche R and Hoffman P (ed) Teleskulptur 3, Kulturdata and Division of Cultural Affairs of the City of Graz, Graz, Austria : 48-72.

Kalawsky R S (1993) The Science of Virtual Reality and Virtual Environments. Addison-Wesley Publishing Company, Cambridge, UK.

Kanade T, Rander P, Vedula S & Saito H (1999) Virtualized Reality: Digitizing a 3D Time-Varying Event As Is and in Real-time. Mixed Reality: Merging Real and Virtual Worlds, Ohmasha Books, Tokyo, Japan.

Kaplan A, Keshav S, Schryer N & Venutolo J (1997) An Internet accessible telepresence. Multimedia Systems 5(1): 140-144.

Kapralos B, Jenkin M & Milios E (2003) Auditory Perception and Spatial (3D) Auditory Systems, York University, USA.
Kim M, Miligram P & Drake J (1997) Virtual tape measure for 3D measurements in micro-surgery. Proceedings 41st Annual Meeting of Human Factors and Ergonomics Society, Alburquerque, NM, USA.
Kim S & Kuc T (1998) Goal-directed Motion Control of 3D Human Character. IEEE International Conference on Systems, Man and Cybernetics, San Diego, USA.
Klinker G, Ahlers K, Breen D, Chevalier P, Crampton C, Greer D, Koller D, Kramer A, Rose E, Tuceryan M & Whitaker R T (1997) Confluence of Computer Vision and Interactive Graphics for Augmented Reality. Presence 6(4): 433-451.
Kraut R, Fish R, Root R & Chalfonte B (1993) Informal communication in organisations: Forms, functions and Technology. Readings in Groupware and Computer Supported Cooperative Work, Morgan Kaufmann, CA, USA.
Krueger W M (1991) Artificial Reality II. Addison-Wesley Publishing Company, USA.
Kuutti K (1999) Small Interfaces -- a blind spot of the academical HCI community? Proceedings of HCI International '99, Munich, Germany.
Kuutti K, Pulli P, Pyssysalo T, Hickey S & Antoniac P (1999) CyPhone Mediaphone Project - Taxi Trip Scenario. 9th International Conference on Artificial Reality and Telexistence ICAT '99, Tokyo, Japan.
Kuzuoka H (1992) Spatial Workspace Collaboration: A Sharedview Video Support System for Remote Collaboration Capability. Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '92), Monterey, California, United States.
Kuzuoka H, Kosuge T & Tanaka M (1994) GestureCam: A Video Communication System for Sympathetic Remote Collaboration. CSCW '94 Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, NC, USA.
Lanier J (2001) Virtually There. Scientific America, April issue.
Lee A, Girgensohn A & Schlueter K (1997) NYNEX Portholes: Initial User Reactions and Redesign implications. International Conference on Supporting Group Work (GROUP '97), Phoenix, Arizona, United States.
Leonhardt U (1998) Supporting Location-Awareness in Open Distributed Systems. Department of Computing, London, Imperial College.
Lin Q & Kuo C (1999) Assisting the teleoperation of an Unmanned Underwater Vehicle Using a Synthetic Subsea Scenario. Presence 8(5): 520-530.
Luff P& Heath C (1998) Mobility in Collaboration. Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work (CSCW 98), Seattle, Washington USA.
Macedonia M, Zyda M, Pratt D R, Barham P & Zeswitz M R (1995) NPSNET: A Network Software Architecture for Large Scale Virtual Environments. Presence 3(4):265-287.
MacIntyre B & Feiner S (1996) Future Multimedia User Interfaces. Multimedia Systems 4(5): 250-268.
Manninen T (2001) Rich Interaction in the Context of Networked Virtual Environments - Experiences Gained from the Multi-player Games Domain. In Joint Proceedings of HCI 2001 and IHM 2001 Conference, Lille, France.
Manninen T (2002) Contextual Virtual Interaction as part of Ubiquitous Game Design and Development. Personal and Ubiquitous Computing 6(5-6); 390-406.
Mantei M, Baecker R, Sellen A, Buxton W, Milligan T & Wellman B (1991) Experiences in the Use of a Media Space. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'91), New Orleans, Louisiana, United States.
March S & Smith G (1995) Design and natural science research on information technology. Decision Support Systems 15(4): 251-266.
Maxfield J, Fernando T & Dew P (1995) A Distributed Virtual Environment for Concurrent Engineering. Virtual Reality - Annual International Symposium '95, Research Triangle Park, NC, USA.
Maxfield J, Fernando T Dew P (1998) A Distributed Virtual Environment for Collaborative Engineering. Presence 7(3): 241-261.

Mellor J (1995) Realtime camera calibration for enhanced reality visualisation. Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95, Nice, France.
Milgram P & Ballantyne J (1997) Real World teleoperation via virtual environment modelling. Proceedings International conference on Artificial Reality and Tele-existence ICAT'97, Tokyo, Japan.
Milgram P & Colquhoun H (1999) A Taxonomy of Real and Virtual World Display Integration. Mixed Reality: Merging Real and Virtual Worlds, Tokyo, Japan, Ohmsha / Springer-Verlag.
Minsky M (1980) Telepresence. Omni 2: 45-51.
Mohr R, Buschmann R, Falkenhagen L, Van Gool L & Koch R (1998) Cumuli, Panorama, and Vangaurd Project Overview. European Workshop on 3D Structure from Multiple Images of Large-scale Environments (SMILE), Freiburg, Germany.
Munro A, Geulen E, Deguine M, Melpignano G & Martinez A (1998) Services and Applications: Requirements and Realizations in the UMTS ERA. IEEE Communications Magazine, February issue : 118-126.
Myers B, Stiel H & Gargiulo R (1998) Collaboration using Multiple PDAs Connected to a PC. Compute Supported Cooperative Work '98, Seattle, Washington, USA.
Nardi B, Schwarz H, Kuchinsky A & Leichner R (1993) Turning away from talking Heads: the Use of Video-as-Data in Neurosurgery. INTERCHI'93 Proceedings of the conference on Human factors in computing systems, Amsterdam, The Netherlands.
Neumann U & Fuchs H (1993) A Vision of Telepresence for Medical Consultation and Other Applications. Proceedings of the 6th International Symposium on Robotics Research, Hidden Valley, PA, USA.
Nigay L, Salembier P, Marchand T, Renevier P & Pasqualetti L (2002) Mobile and Collaborative Augmented Reality: A scenario Based Design Approach. Fourth International Symposium on Human Computer Interaction with Mobile Devices, Pisa, Italy.
NRC, U. S. N. R. C. (1997) More than Screen Deep - Toward Every-Citizen Interface's to the Nation's infrastructure, Washington, DC, National Academy Press.
Obata A & Sasaki K (1998) OfficeWalker: A Virtual Visiting System based on Proxemics. Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, Seattle, USA.
Okada K, Maeda F, Ichikawaa Y & Matsushita Y (1994) Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. CSCW '94 Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, USA.
Olives J, Mottonen R, Kulju J & Sams M (1999) Audio-Visual Speech Synthesis for Finnish. AVSP'99, Audio-Visual Speech Processing, Santa Cruz, USA.
Olives J, Sams M, Kulju J, Seppäla O, Karjalainen M, Altosaar T, Lemmetty S, Töyrä K & Vainio M (1999) Towards a High Quality Finnish Talking Head. IEEE Signal Processing Society 1999 Workshop on Multimedia Signal Processing, Copenhagen, Denmark.
Orwell G (1949) Nineteen Eighty-Four. Secker and Warburg, London, UK
PAULA (2003) www.paula.com, Last accessed 2003,15.7.2003
Paulos E & Canny J (1997) Ubiquitous tele-embodiment: applications and implications. International Journal Human-Computer Studies 46(6): 861-877.
Paulos E & Canny J (1998) PRoP: Personal Roving presence. Computer Human Interaction 1998, Los Angelus, USA.
Pearmain A, Cosmas J, Carvalho A & Typpi V (1999) The MoMuSys MPEG-4 Mobile Multimedia terminal and Field trials. ACTS Mobile Summit 1999, Sorrento, Italy.
Pearsall J (2001) New Oxford Dictionary of English. Oxford University Press, UK.
Pereira F (2000) MPEG-4: Why, What, How and When? Signal Processing: Image Communication 15(4-5): 271-279.
Piekasski W & Thomas B (2002) ARQuake: The Outdoor Augmented Reality Gaming System. Communications of the ACM 45(1): 36-38.
Popsecu-Zeletin R (2001) Towards I-Centric Communications. Proceedings of the WWRF Kick-Off Meeting, Munich, Germany.

Poupyrev I, Tan D, Billinghurst M, Kato H, Regenbrecht H & Tetsutani N (2001) TILES: A Mixed Reality Authoring Interface. INTERACT2001 Eight IFIP TC.13 Conference on Human-Computer Interaction, Tokyo, Japan.
Poupyrev I, Tan D, Billinghurst M, Kato H, Regenbrecht H & Tetsutani N (2002) Developing a Generic Augmented Reality Interface. Computer, IEEE Computer Society Press 35(3): 44-50.
Pratt D, Pratt S,Barham P, Barker R, Waldrop M, Ehlert J & Chrislip C (1997) Humans in Large-scale, Networked Virtual Environments. Presence 6(5): 547-564.
Prince S, Cheok A, Farbiz F, Williamson T, Johnson N, Billinghurst M & Kato H (2002) 3-D Live: Real Tine Interaction for Mixed Reality. CSCW 2002 Conference on Computer Supported Cooperative Work, New Orleans, USA.
Pulli P, Pyssysalo T, Kuutti K, Simila J, Metsavainio J & Komulainen O (1998) CyPhone - Future Personal Telecooperation Device. Telecooperation, Proceedings of the XV IFIP World Computer Congress, Vienna/Austria & Budapest/Hungary.
Pyssysalo T (1998) The Application of Augmented reality in Future Mobile Communication Services. Working Conference; Personal Wireless Communications, Tokyo, Japan.
Raskar R, Welch G, Cutts M, Lake A, Stesin L & Fuchs H (1998) The Office of the Future: A Unified Approach to Image-Based Modelling and Spatially Immersive Displays. SIGGRAPH '98 Computer Graphics Proceedings, Annual Conference Series, Orlando, Florida, USA.
Rauthenberg S, Kauff P & Graffunder A (2000) A Realtime Implementation of a Shared Virtual Environment System Using Today's Consumer Technology in Connection with the MPEG-4 Standard. 3rd International Workshop on Presence, Delft, Netherlands.
Reddy M, Watson B, Walker N & Hodges L (1997) Managing level of Detail in Virtual Environments: A perceptual Framework. Presence 6(6): 658-666.
Regan C (1995) An investigation into nausea and other side-effects of head-coupled immersive virtual reality. Virtual Reality: Research, Development, Applications 1(1): 17-32.
Reijers N, Cunningham R, Meier R, Hughes B, Gaertner G & Cahill V (2002) Using Group Communication to support Mobile Augmented Reality Applications. Proceedings of the Fifth IEEE International Symposium on Object-Orientated Real-Time Distributed Computing (ISORC'02), Washington, USA.
Reitmayr G & Schmalstieg R (2001) A Wearable 3D Augmented Reality Workspace. Proceedings of 5th International Symposium on Wearable Computers (ISCW 2001), Zurich, Switzerland.
Rekimoto J, Ayatsuka Y & Hayashi K (1998) Augmentable Reality: Situated Communications through Physical and Digital Spaces. In Proceedings of the 2nd International Symposium on Wearable Computers, Pittsburgh, USA.
Renevier P & Nigay L (2001) Mobile Collaborative Augmented Reality: The Augmented Stroll. Engineering for Human-Computer Interaction: 8th IFIP International Conference, EHCI 2001, Toronto, Canada.
Robinett W (1992a) Synthetic Experience: A taxonomy, Survey of Earlier Thought, and speculations on the Future. University of North Carolina, Chapel Hill: 30.
Robinett W (1992b) Synthetic experience: a proposed taxonomy. Presence 1(2): 229-247.
Rolland J, Holloway R & Fuchs H (1994) A comparison of optical and see-through head-mounted displays. SPIE Proceedings Volume 2351: Telemanipulator and Telepresence Technologies, Boston, USA.
Rubin P & Vatikiostis-Bateson E (1998) Talking Heads. Proceedings of Auditory-Visual Speech Processing, Sydney, Australia.
Ruddle R, Payne S & Jones D (1998) Navigating Large-Scale Desk-top Virtual Buildings: Effects of Orientation Aids and Familiarity. Presence 7(2): 179-192.
Ruddle R, Payne S & Jones D (1999) Navigating Large-Scale Virtual Environments: What differences occur Between Helmet-Mounted and Desk-Top Displays? Presence 8(2): 157-168.
Satzinger J & Olfman L (1992) A Research Program to assess user perceptions of Group work support. CHI '92, Human factors in computing systems, Monterey, CA, USA.
Schloerb D W (1995) A Quantitative Measure of Telepresence. Presence 4(1): 64-80.
Schmalstieg D & Hesina G (2002) Distributed Applications for Collaborative Augmented Reality. Proceedings of the IEEE Virtual Reality (VR'02), Orlando, Florida.

Sellen A (1995) Remote conversations: The effects of mediating talk with technology. Human-Computer interaction 10(4): 401-444.

Sheridan T (1991) Telepresence and Artificial Reality: Some Hard Questions. International Conference on Artificial Reality and teleexistence, Tokyo, Japan,

Sheridan T (1992a) Musings on Telepresence and virtual presence, Presence 1(1): 120-126,

Sheridan T (1992b) Telerobotics, Automation and Human Supervisory Control, MIT Press, Massachusetts, USA.

Sheridan T (1999) Descartes, Heidegger, Gibson and God: Towards an Eclectic Ontology of Presence. Presence 8(5): 551-559.

Singh G, Serra L, Png W, Wong A & Ng H (1995) BrickNet: Sharing Object Behaviours on the Net. Virtual Reality Annual International Symposium (VRAIS'95), Research Triangle Park, North Carolina.

Singhal S & Zyda M (1999) Networked Virtual Environments Design and Implementation. ACM Press, NY, USA.

Slater M, Usoh M & Steed A (1995) Taking Steps: the influence of a Walking technique on Presence in Virtual Reality. ACM Transactions on Computer-Human Interaction 2(3): 201-219.

Slater M & Wilbur S (1997) A Framework for Immersive Virtual Environments(FIVE): Speculations on the Role of Presence in Virtual Environments. Presence 6(6): 603-616.

Slay H, Thomas B & Vernik R (2002) Tangible User Interaction using Augmented Reality. Australian User Interface Conference, Melbourne, Australia.

Sohlenkamp M & Chwelos G (1994) Integrating Communication, Cooperation, and Awareness: the DIVA Virtual Office Environment. CSCW '94 Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill North Carolina, USA.

Stajano F (2002) Security for Ubiquitous Computing. John Wiley & Sons Limited, Chichester, UK.

Stajano F & Anderson R (1999) The Resurrecting Duckling: Security issues for Ad-Hoc Wireless Networks. Security Protocols, 7th International Workshop, Cambridge, UK.

Stanney K, Mourant R & Kennedy S (1998) Human factors in Virtual Environments: A Review of the Literature. Presence 7(4): 327-351.

Stansfield S, Miner N, Shawver D & Rogers D (1995) An application of Shared Virtual Reality to Situational Training. Virtual Reality Annual International Symposium '95, Research Triangle Park, NC, USA.

Starner T, Mann S, Rhodes B, Levine J, Healey J, Kirsch D, Picard R & Pentland A (1997) Augmented reality through Wearable Computing. Presence 6(4): 386-398.

State A, Livingston M, Garrett W, Hirota G, Whitton M, Pisano E & Fuchs H (1996a) Technologies for Augmented Reality Systems: Realizing ultrasound -guided Needle Biopsies. Computer Graphics Proceedings, SIGGRAPH '96, New Orleans, USA.

State A, Hirota G, Chen D, Garrett W & Livingston M (1996b) Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. Computer Graphics Proceedings, SIGGRAPH '96, New Orleans, USA.

Steur J (1992) Defining Virtual Reality: Dimensions Determining Telepresence. Journal of Communication 4(4): 73-93.

Stubblefield A, Ioannidis J & Rubin A (2001) Using the Fluher, Mantin and Shamir Attack to break WEP. AT&T Labs.

Stytz M (1996) Distributed Virtual Environments. IEEE Computer Graphics and Applications 16(3): 19-31.

Sutherland I (1968) A Head Mounted three - dimensional display. Proceedings of Fall Joint Computer Conference, AFIPS Conf Proceedings, Washington, USA.

Tachi S, Kawakami N, Inami M & Zaitsu Y (2004) Mutual Teleexistence System Using retro-Reflective Projection technology. International Journal of Humanoid Robotics 1(1): 45-64.

Tan D, Poupyrev I, Billinghurst M, Kato H, Regenbrecht H & Tetsutani N (2001) On-demand, In-place help for Mixed Reality Environments. Ubicomp 2001: Third International Conference Ubiquitous Computing, Atlanta, USA.

Tang J C, Isaacs E & Rua M (1994) Supporting Distributed Groups with a Montage of Lightweight Interactions. CSCW '94 Proceedings of the Conference on Computer Supported Cooperative Work, Chapel hill, North Carolina, USA.

Thomas B, Close B, Donoghue J, Squires J, De Bondi P & Piekasski W (2002) First person Indoor/outdoor Augmented Reality Application. Personal and Ubiquitous Computing 6(1): 75-86.

Thomas B, Krul N, Close B & Piekasski W (2002) Usability and playability issues for ARQuake. IFIP First International Workshop on Entertainment Computing (IWEC 2002), Makuhari, Japan.

Torkin O (2000) Architectures for Location-Based Services. MSc thesis, Department of Information Processing Science, University of Oulu.

Van Dam K, Pitchers S & Barnard M (2001) Body Area Networks: Towards a Wearable Future. Proceedings of the Wireless World Research Forum, Munich, Germany.

Van Diggelen F & Abraham C (2001) Indoor GPS Technology. Cellular Telecommunications and Internet Association CTIA Wireless-Agenda, Dallas, USA.

Vittorini D & Robinson B (2003) Optimizing Indoor GPS Performance. GPS World.

Vlahakis V, Ionnidis N, Karigiannis J, Tsotros M, Gounaris M ,Stricker D, Gleue T, Daehne P & Almeida L (2002) Archeoguide: An Augmented Reality Guide for Archaeological Sites. IEEE Computer Graphics and Applications 22 (5): 52-60.

Waggoner B (2001) The Many Faces of MPEG-4. DV Web Video: 44-51.

Wang Q, Green M & Shaw C (1995) EM - An Environmental Manager for Building Networked Virtual Environments. Virtual Reality Annual International Symposium (VRAIS'95), Research Triangle Park, North Carolina, USA.

Watson B & Hodges L (1995) Using texture maps to correct for optical distortion in head-mounted displays. Proceedings of the Virtual Reality Annual International Symposium '95, Research Triangle park, North Carolina, USA.

Weisbin C & Lavery D (1994) NASA rover and Telerobotics technology program. IEEE Robotics & Automation Magazine 1(4): 14-21.

Weiser M (1991) The Computer for the Twenty-First Century. Scientific American, 263(7): 94-104.

Weiser M (1993) Some Computer Science Problems in Ubiquitous Computing. Communications of the ACM 36(7): 74-84.

Winograd T & Flores F (1986) Understanding computers and cognition: A new foundation for design. Addison-Wesley, New Jersey, USA.

Witmer B & Kline P(1998) Judging Perceived and Traversed Distance in Virtual Environments. Presence 7(2): 144-167.

Woods E, Billinghurst M, Aldridge G & Garrie B (2004) Augmenting the science centre and museum experience. Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia, Singapore.

Yamaashi K, Cooperstock J, Narine T & Buxton W (1996) Beating the Limitations of Camera-Monitor Mediated Telepresence with Extra Eyes. Proceedings of CHI'96, ACM Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada.

Zahorik P & Jenison R (1998) Presence as Being-in-the-World. Presence 7(1): 78-89.

Zeltzer D (1992) Autonomy, Interaction, and Presence. Presence 1(1): 127-132.

Zhao Q & Stasko J (1998) Evaluating Image Filtering Based Techniques in Media Space Applications. CSCW '98 Proceedings of the ACM 1998 Conference on Computer Supported Cooperative Work, Seattle Washington USA.

Zimmerman T, Lanier J, Blanchard C, Bryson S & Harvill Y (1987) Hand Gesture Interface Device. CHI'87 Conference Proceedings: Human Factors in Computing Systems and Graphics Interface, Toronto, Canada.