CLRC

Technical Report
RAL-TR-95-077

Research in to Various Aspects of Visual Perception and Virtual Reality

M Prime

February 1996

COUNCIL FOR THE CENTRAL LABORATORY OF THE RESEARCH COUNCILS

The Central Laboratory of the Research Councils
Library and Information Services
Rutherford Appleton Laboratory
Chilton
Didcot
Oxfordshire
OX11 0QX
Tel: 01235 445384 Fax: 01235 446403
E-mail library@rl.ac.uk

ISSN 1358-6254

Research in to various aspects of Visual Perception and Virtual Reality

Martin Prime
Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, England
martin@inf.rl.ac.uk

1. Introduction

This report describes a six month European Research Consortium for Informatics and Mathematics (ERCIM) internal research fellowship which was completed between October 93 and April 94. The fellowship was fully funded by the Rutherford Appleton Laboratory. The work was carried out at both the Swedish Institute of Computer Science (SICS) and Kunggl Tekniska Hogskolan (KTH) which is the Royal Institute of Technology in Stockholm.

The objective of the fellowship program is to widen the contacts between the individual ERCIM partners allowing staff to contribute to research at another institute, gain insight into the workings of a different research organisation, and help to foster closer links in the future between ERCIM member organisations.

ERCIM is a grouping of the thirteen leading research organisations in Europe which are involved in basic and strategic research and developments with industry. It also acts as a major link between academia and industry.

SICS it the ERCIM member for Sweden but KTH and SICS have very close research links. High speed networks have been set up between the two institutes to facilitate these collaborations. Sections of the research period were spent researching with both institutes.

1.1. The Swedish Institute of Computer Science (SICS)

SICS was founded in 1988 and conducts research in computer science and artificial intelligence. SICS is a non-profit foundation sponsored by the Swedish National Board for Technical and Industrial Development (a government agency commonly known as NUTEK) and by an industrial association of five major companies, Asea Brown Boveri AB, Telefonaktiebolaget LM Ericsson, IBM Svenska AB, Telia AB and Fvrsvarets Materielverk, FMV (Defence Material Administration).

The main role of SICS is bridging the gap between academic research in computer science in Sweden and related industrial research and development, thus contributing to the competitive strength of Swedish industry. SICS carries out advanced research in selected areas of Computer Science on one hand and promotes use of new ideas and results for industry on the other hand. SICS' personnel are also engaged in education, both at undergraduate and PhD program level at several Swedish universities. In this way new insights and research results are made available to computer science and computer engineering students. Several senior researchers hold university positions part-time.

During 92/93 the Institute's total income was approximately 49 MSEK (there are around 11 Swedish Krona to the pound). The major part of the Institute's research has been conducted within the boundaries of SICS's Basic Research Program. Contracted research amounted to approximately 49 percent of income.

The Institute holds strong positions in areas such as programming methodology, distributed and parallel computer architectures, natural language processing, knowledge-based systems, formal design methods, advanced user interfaces, high-speed networks, and communication protocols.

SICS had sixty-one permanent employees at the end of the '93 fiscal year. An additional forty-six people worked on SICS projects as temporary employees during that year.

Among the many software prototypes that have been produced at SICS the most interesting for my potential research were the following:

Collaborative Desktop (CoDesk) is an extension of the desktop metaphor for a distributed desktop environment that integrates mechanisms for group communication (audiovisual) with distributed applications and underlying support for cooperation. CoDesk is also an interface to TheKnowledgeNet, which is a system in which members have access to joint information. The prototype is being developed in cooperation with KTH, The Department of Numerical Analysis and Computer Science (NADA) and Interaction and Presentation Laboratory (IPLab) in the MultiG research program and is accessible in a UNIX environment.

DIVE 2.0 (Distributed Interactive Virtual Environment) is a software system for the design of interactive, distributed virtual worlds in a heterogeneous network environment. DIVE 2.0 has been licensed for noncommercial use to over twenty institutions throughout the world, such as NASA and the US defence industry. DIVE is used to construct and experiment with distributed and shared applications based on the metaphor of virtual, three-dimensional space. Tools for distributed, computer-supported cooperative work, network monitoring, 3-D presentation for control-tower personnel, the visualization of file systems, and control of an industrial robot are some of the examples of applications demonstrations SICS has produced with DIVE. DIVE is developed in ANSI-C and currently runs on IBM, SUN, and Silicon Graphics workstations. A Swedish small business has entered into negotiations with SICS regarding the exploitation rights for DIVE 2.0.

A prototype of an audiovisual teleconferencing system has been developed. The system takes advantage of ATM technology and operates within the Stockholm Gigabit Network (SGN). International audiovisual conferences have been carried out via Internet with the participation of researchers in the United States, Australia, and Europe.

SICS is housed in the Electrum building at Kista. The Electum is a forum for cooperation between higher education, research institutes and industry, its purpose being to strengthen Sweden's position in electronics and computer technology.

1.2. Kunggl Tekniska Hogskolan (KTH)

KTH, The Royal Institute of Technology, is the largest of Sweden's six colleges of technology. The college provides one-third of Sweden's capacity for engineering studies and technical research at post-secondary level.

KTH trains engineers in the fields of materials technology, architecture, computer science, electrical engineering, engineering physics, chemical engineering, surveying, mechanical engineering, industrial economics, vehicle and civil engineering. KTH is also committed to the subjects of "Technology, Man and Society" in its basic studies, and also to making environment protection an integral part of all technical subjects.

The college is housed in listed buildings on Valhallavagen which were built in 1917. KHT has around 7,500 students, with 2000 being admitted every year, and 25% of these are women. There is a staff of 2,500 and a total expenditure of around MSEK 1,300 two thirds of which goes on research.

The motto of KTH is "Vetenskap och Konst" which means "Science and Art". The motto was used in the 19th century to indicate a holistic view of existence and social progress. "Science" in this context referees not only to natural science but also the knowledge compiled into systems. "Art" meant, not only the fine arts but also the professionalism and craftsmanship belonging to the "art" of engineering. This motto means that the college was designed with many examples of fine art through out the building. The structure is going through a five year renovation with many of the sculptures and frescos being repaired and replaced.

The IPLab was established in 1985 at NADA it is based on an interdisciplinary group consisting of researchers and research students in computer science, linguistics, psychology and sociology with interests in human computer interaction with graphics workstations. There are two main areas of study; computer supported writing; and distributed multimedia applications and programming in high speed networks. It was with the multimedia group that I chose to carry out my research for the length of the fellowship.

2. Areas of study

Initially time was spent getting to know both KTH and SICS, their work and the people who are researching there. From this staring point and reading around the diverse research of the groups and initial study plan was created. I tried to fit the work in with my prior research experience and interests. More reading was carried out on this reduced set of topics. I wanted my research to fit in with existing projects at the labs so I could possibly expand their scope. Through discussions with researchers at both KTH and SICS the plan was expanded with greater detail, and three research projects were identified. Colleague at the IPlab suggested papers to read to provided an understanding of the work that has already been carried out at both of the labs in similar areas.

The three areas of study were:

a) Distance perception

b) 3D input device evaluation

c) Virtual Reality (VR) issues

The report now describes these three research projects in more detail, for full information on the work the reader is advised to read the published papers referenced. Some of the papers that were worked on have not been published when this RAL report was printed. These papers are included as appendices to this document. They are a snap shot in time and comments are welcomed on the papers contents.

3. Distance perception

This research was aimed at studying depth perception in static images. It was hoped that it would be possible to extrapolate from this study in to the area of VR.

Two versions of this paper were written. One was published in Computers and Graphics and the other version was a more extended description of the work and published as a KTH report.

The papers describe the experimental results of viewing 3D objects on a 2D display. The effect of object relationship; presentation quality (wireframe and shaded); and illumination on depth perception were studied. After the depth estimation phase, subjects were then asked to make a subjective judgement: whether they thought better estimates had been made with the shaded rather than wireframe form. Subsequent analysis, showed that the experimental figures did not support this guess. Lighting influenced accuracy, which illustrates the importance of careful choices of light sources. Relative object placement was also established as an important factor.

Prof Kjellahl and I designed the experiment together, Prof Kjelldahl created the software while I developed the experimental design. We then spilt the experimental sessions between us. I carried out the statistics and interpretation of the results. Finally i wrote the papers with comments and input from Prof Kjelldahl.

Work was also started on a similar experiment but this time using stereo equipment to study accuracy of depth judgement, this work is continuing at KTH with input form the author.

3.1. References

L Kjelldahl and M Prime, A study of the effects on depth perception for different presentation methods of 3D objects on a 2D display, Computers and Graphics, Vol 9, No. 2. March/April 1995.

L Kjelldahl and M Prime, A study of the effects on depth perception for different presentation methods of 3D objects on a 2D display, report number - NADA, KTH, TRITA-NA-P30, 1994, published at KTH Sweden.

4. 3D input evaluation

This work extended research carried out by the author in 1990 evaluating 3D input devices in a 3D world. The research fellowship carried out similar experiments but this time in a VR world.

The experiment was designed to evaluate the usability of different input devices in a Virtual Reality (VR) environment. The evaluation involved both speed and accuracy to achieve a set of experimental tasks. Two input devices were tested in the experiment: the DataGlove <tm> a hand tracking device and the Bird <tm> a flying mouse. The tasks were developed under the DIVE VR environment.

The Dive environment can be used either in a window or as a total immersion VR experience. Initially the system was run in the window mode as it was easier for the subjects to use. The second set of subjects used the full immersion VR Head Mounted Display (HMD) systems.

It was possible for subjects to move around the objects, in both immersion and window modes. The keyboard was used to achieve different views in the window mode. Four different view of the scene could be used by pressing four different keys. It was decided not to use the input devices to control movement (as would have been normal) as it was felt this would interfere with their use as pure manipulation devices.

I developed the experimental design and this allowed it to follow on from the work I had already carried out. Kai-Mikael Jaa-Aro, a PhD student, then designed the software that would be needed and Stephane Royer, an MSc student carried out the coding. Mr Jaa-Aro and I then jointly carried out the experimentation. Mr Jaa-Aro then carried out a statistical analysis of the dates. I then analysed the results and wrote up the work. The paper has not yet been published and is included in appendix 1 for information. The posters were also written to describe the work at the 6th ERCIM Workshop.

4.1. Reference
Posters published in 6th ERCIM Workshops, Stockholm, June 1-3, 1994. ERCIM-94-W001 SICS - M. Prime, K. Jaa-Aro and S Royer

5. Virtual reality issues

This work started by investigation issues around the"spacial Interaction model" developed at SICS. This continued with a study of Embodiment within VR and then into Sign Language in the VR environment.

5.1. Embodiment
When users enter a Virtual Reality (VR) environment they usually need some sort of body to inhabit there, this VR entity is the embodiment of the user. It is the body image which other users of the system will see when they look in the direction of the user. If the user is alone in the system they will still need an embodiment, when they look down at their own body they will need some sort of image to view. The VR

environment and thus the embodiment are purely abstractions of the real world and as such it should be remembered that they need not represent real life. The representation of a human need not look humanoid, the embodiment may be that of a lobster in an undersea world for example.The development, design and paper for this research was all my own work. This paper was not finished to the point of publication during the fellowship period. The version provided in appendix 2 acts as a snapshot in time of the paper as it existed at the end of the fellowship period. It is a working paper so any comments are welcome. It is presented here so that its initial ideas in the area can be presented more widely.

5.2. Sign Language
Three versions of the Sign Language paper were written, one published in ERCIM Report ECRIM-94-W001 The second version is and updated and extended version that has not yet been published, it is included in appendix 3 for information. Finally I was invited guest speaker at the IEE Colloquium on Visualisation of Three-Dimensional fields, a shortened version of the paper was produced for this event. The development, design and papers for this research was all my own work.

The papers describe the possible implementation of sign language within a Virtual Reality (VR) environment. The Spatial Interaction Model is defined with the concepts of focus, awareness, nimbus and aura to explain how sign language could be used in VR applications. An implementation of this system would provide communication as well as visual access to the deaf in a VR environment. The implementation trade-offs for such a system are addressed. The conclusion is that existing systems are capable of supporting an initial implementation.

5.3. References
A description of the spacial interaction model using sign language as an example. M. Prime, 6th ERCIM Workshops, Stockholm, June 1-3, 1994. ERCIM-94-W001 SICS

Integrating sign language into a virtual reality environment. M. Prime, IEE Colloquium on Visualisation of Three-Dimensional fields, ISSN 0963-3308, ref 95/205, 1995.

6. Other work

SICS and KTH have extensive research and lecturing programs, I carried out work in both these areas.

6.1. Lecturing
The lecture "The role of Visual Perception in user interface design" was presented as part of the undergraduate course on User interface design. This was presented to around 130 students.

I presented two of the IPlab's monthly technical discussions, one on visual perception of information and the other on hand tracking devices. These lectures were attended by around 20 researchers.

Many of the experimental sessions for my research were carried out by the students as assessed assignments. This gave the students an understanding of using experimental methods on a research program.

6.2. COMIC
I provided technical input to the COMIC project during the regular project meetings. The COMIC project is an ESPRIT Basic Research Action investigating configurable mechanisms of interaction for Computer Supported Cooperative Work. COMIC is a broad ranging multi-disciplinary project spanning Computer science, cognitive science and social science. Overall, COMIC aims to address a range of issues including organisational support for cooperative work, CSCW design methodologies, notations of group interaction and developing novel models of group interaction. COMIC meeting were every fortnight, bring together around 8 to 10 people involved with the project to exchange ideas, read papers and present results.

7. Conclusions

The objective of these fellowship programs is to widen the contacts between the individual ERCIM partners allowing staff to contribute to research at another institute, gaining insight into the workings of a different research organisation, and helping to foster closer links in the future between ERCIM member organisations. I very much feel this has been the case in this fellowship research trip and hope it has allowed RAL, SICS and KTH to gain a closer understanding of each others research. This will help create future collaborations between groups in all three institutes through my knowledge of all the relevant staff.

Quite a wide range of topics were covered by the research but they were all in the area of perception, object manipulation and how this could relate to a Virtual Reality environment. Both KTH and SICS are doing research in the VR area and the fellowship research was an attempt to fit in with that work.

8. Acknowledgments

I would like to thank the Rutherford Appleton Laboratory for its support and funding of this period in Sweden.
I would also like to thank the students at KTH who acted as subjects for the experiments.
Thanks go to KTH, SICS and ERCIM for making the fellowship possible.

Appendix 1

Evaluation of the usability of input devices in a Virtual Reality Environment, M. Prime, K. Jaa-Aro and S Royer, 1995.

Plus posters published in 6th ERCIM Workshops, Stockholm, June 1-3, 1994. ERCIM-94-W001 SICS - M. Prime, K. Jaa-Aro and S Royer

Evaluation of the usability of input devices in a Virtual Reality Environment

Martin Prime*, Kai-Mikael Jaa-Aro#, Stephane Royer#
*Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, England
martin@inf.rl.ac.uk
#IPLab, KTH, Lindstedtvagen 5
100 44 Stockholm, Sweden

Abstract

1. Introduction

The Interaction and Presentation Laboratory (IPLab) of the Department of Numerical Analysis and Computing Science (NADA) at the Royal Institute of Technology (KTH) is concerned with the field of Human-Computer Interaction. One of the projects at IPLAb is the development of new methods of interaction in computer-generated three-dimensional spaces.

This experiment was designed to evaluate the usability of different input devices in a Virtual Reality (VR) environment. The evaluation involved both speed and accuracy to achieve a set of experimental tasks. Two input devices were tested in the experiment: the DataGlove <tm> a hand tracking device and the Bird <tm> a flying mouse. The tasks were developed under the DIVE VR environment.

Dive can be used either in a window or as a total immersion VR experience. Initially the system was run in the window mode as it was easier for the subjects to use. The second set of subjects used the full immersion VR Head Mounted Display (HMD) systems.

It was possible for subjects to move around the objects, in both immersion and window modes. The keyboard was used to achieve different views in the window mode. Four different view of the scene could be used by pressing four different keys. It was decided not to use the input devices to control movement (as would have been normal) as it was felt this would interfere with their use as pure manipulation devices.

2. The DIVE environment

The device evaluations were run within the DIVE Virtual Reality environment developed at SICS (Swedish Institute of Computer Science). DIVE (Distributed Interactive Virtual Environment) is an experimental system used to test out interfaces and applications. The architecture is distributed and allows multi-user applications, this enables geographically dispersed users to share the same work space.

DIVE is a loosely coupled heterogeneous distributed system based on Unix and internet networking protocols within local and wide-area networks. Consistency and concurrency control of common data is achieved by active replication, reliable multicast protocols and distributed locking methods.

Users of the DIVE system are provided with a simple body-icon, when they navigate around the 3D environment they will encounter other users (with body-icons) and application processes, a white board for example. The user can interact with these other objects in their environment, any changes they make will be indicated to the other users of the system.

DIVE has a rendering application called the visualizer, this presents information about the state of the environment to the users. It also reads the user's input devices and maps the physical actions taken by the user to logical actions in the DIVE system. This includes navigation in 3D space, clicking on objects to select them and grabbing objects.

3. The input device interface

DIVE has much of its device driver code hard coded into it. Part of this experiment was to try and devise a more generic interface between the hardware devices and the DIVE software package.

3.1. Defining signals for generic event handling from the devices

Originally DIVE supported picking objects up with the mouse and with the wand, but this behaviour was hard-wired into the code. What was needed was to generalise this so that pressing a button is a signal to be interpreted by the system as a pickup or as some other kind of command appropriate under the circumstances. A standard "data structure" needed to be defined that could take in the raw data from the devices and then pass it in a standard form to DIVE to carry out the generic tasks.

It was hoped to develop drives for many input units (DataGlove, Magic Wand, GeoBall, Bird and Immersion Probe). The main experiment was to examine the difference between using a 6D mouse type input device and a hand tracking input device. Also a desk based device (probe) and a free standing one (Glove). Obviously as many different devices as possible should be able to use the newly created standard input system to DIVE.

3.2. The event handler

3.2.1. A layer model

The input handler is modular to make updating and maintenance simpler. Since this system will carry out communication, a natural way to represent this idea is to have a layer model. The general design of input handling is presented in the fig. 1.

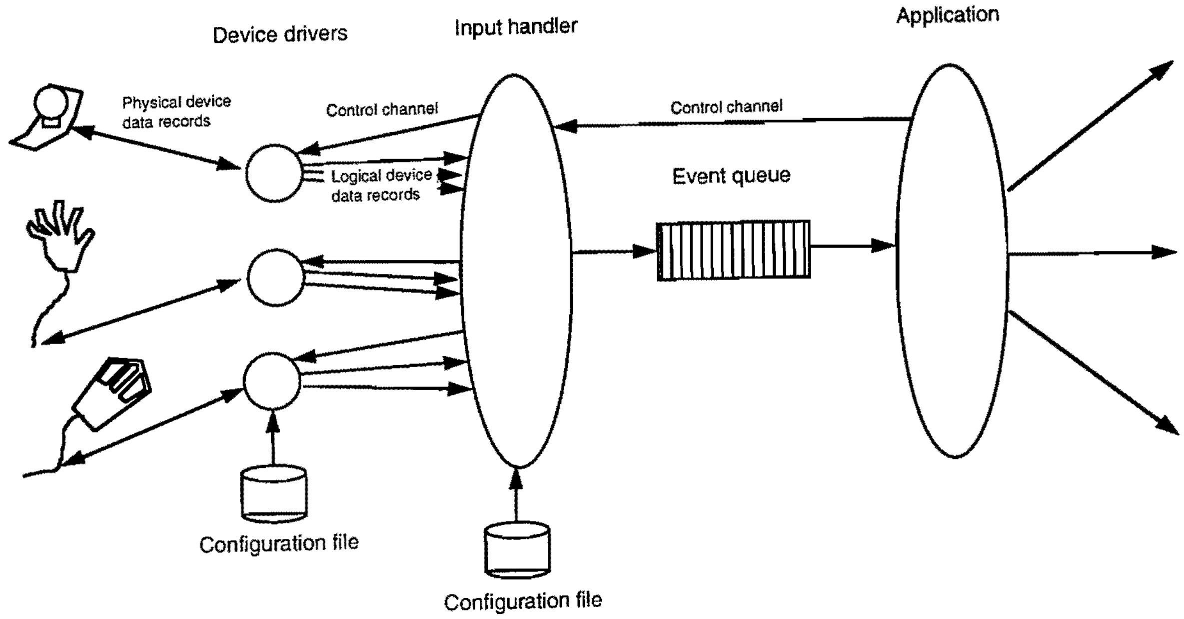


Figure. 1 General design of input handling

The PHYSICAL LAYER :
is under the control of the server of a given physical device, this layer deals with the signals sent from the physical device and converts the signals to explicit events (eg : For the mouse "Button-2 pressed", the glove :"Position of the index").

The LOGICAL DEVICES :
This layer converts the signals of the physical layer to configurable commands (eg : For the mouse "Button-2 pressed" can be translated into "Grasp an object of the world" or into "Open a pop-up menu").

The LOW-LEVEL EVENTS :
At this layer a set of logical-input values can be compiled in order to generate a more complicated command (eg : 2 positions "M1" and "M2" can be translated in "MOVE from M1 to M2").

The HIGH-LEVEL EVENTS :
Any command is translated as an action on the world (eg : "MOVE an object in the world").

The PHYSICAL-OUTPUT :
Aims at rendering the world, this can be via screen or 3D glasses.

3.2.2. The model for logical input devices

As in GKS a list of logical input device classes can be defined, these logical classes depend on the physical devices. But some problems may occur because of the large number and complexity of these physical devices, and the possible new creations. This list is not complete, and is as general as possible and easily extendable.

According to the existing physical devices a possible list may be the following:

POSITION : a position in 3 dimensions
ORIENTATION : a vector in 3 dimensions.
SPEED : a speed with direction, and absolute value of the speed.
ROTATION : a rotation around a given axe, and the absolute rotational speed.

SELECTION : command of selection of an object.
DESELECTION : command of deselection of an object.
GRASPING : command for grasping an object.
UNGRASPING : command to stop grasping an object.

3.2.3. The implementation of the logical devices

Each physical device (mouse, joystick, geoball, dataglove ...) is connected to one and only one device driver, this device driver receives raw data from the physical device.

The driver, according to the content of a file defining the mapping between physical and logical device(s), set the corresponding logical device(s). The existence of this file allows the reconfiguration of the mapping between the physical and logical device. These logical devices are implemented as sockets to which the driver writes data, and from which the visualizer reads the logical output values.

3.2.4. The mapping between the physical and logical device

This mapping is defined by a data file. The file will be different for each class of driver, but a standard file is useful if the user wants to create a tool to help in generation of such files.

Examples for the mouse and the glove could have the following structure :

MOUSE :

STATES:
button-1 (pressed or released)
button-2 (pressed or released)

SELECTION : if button-1 (pressed)
DESELECTION : if button-1 (released)
GRASPING : if button-2 (pressed)
UNGRASPING : if button-2 (released)
POSITION : xmouse, ymouse, 0
ANGLE :
SPEED :
ROTATION :

GLOVE :

STATES :
Status (SELECTED or GRASPED)
other-fingers (FLEXED or NONFLEXED)
index (FLEXED or NONFLEXED)
thumb (FLEXED or NONFLEXED)

SELECTION :
if status ^(SELECTION) and
if other-fingers (FLEXED) and index (NONFLEXED)
DESELECTION :
if status (SELECTION) and
if other-fingers (NONFLEXED)
GRASPING :
if status ^(GRASPING) and
if other-fingers (FLEXED) and index (FLEXED)
UNGRASPING :
if status (GRASPING) and
if other-fingers (NONFLEXED) and index (NONFLEXED)
POSITION :
function1 = (TR-Matrix)
ANGLE :
function2 = (TR-Matrix)
SPEED :
if other-fingers (FLEXED) and thumb (NONFLEXED) and index (NONFLEXED)
then direction = direction of the index, speed value=10 units.
ROTATION :
if other-fingers (FLEXED) and thumb (NONFLEXED) and index (FLEXED)
then axe=direction of the thumb, rotational value=5 units.

4. The Tasks

This is a listing of the 3D input tasks the subjects had to carry out. The test set took around one hour to complete. The tests were a study of manual dexterity with the main interest in speed and accuracy to complete the tasks. All these tasks are based on work carried out in previous experiments carried out by one of the authors (Prime 91).

First set of tasks analysed the accuracy of positioning objects. Three tests were completed in this set, each with three blocks.
The subjects had to:

a) make a horizontal line of the set of blocks

b) make a virtual pile of the set of blocks, in both these tests the central block had a fixed location so it was not possible for the subject to move it.

c) make a stack of a large, medium and small block. The largest block had a fixed position in this test.

A variable of drop shadows was used in some of these tests as an additional factor.

There were two light sources to illuminate the scene and provide contrast for the objects.

The next three tests analysed orientation tasks, position was not important.

In each of these tests the subjects were presented with two identical objects, one in a fixed orientation and the other a test object. The subject had to twist the test object until the thought it was the same orientation as the fixed object.

The next three tests were more complex tasks involving both position and orientation.

a) Hole test, the subject has to place a cube block in a similarly shaped hole within a wall of other blocks. The wall was fixed.

b) Inside test, the subject has to place one object inside another, then place both those objects inside a third. The largest box was fixed.

c) The third test in this set involved pushing 3 rod objects through the axis of a cube. The main test block was fixed.

Collision detection was either turned on or off for these tests to examine what effect it had on the test completion.

Finally the subjects were asked to carry out a complex user task. This test was carried out either with or without total immersion VR equipment. The subjects were not given much information about the task, they had to try work it out from within the experiment.

They were placed in a VR world with 5 gateways to other rooms:

1) this room had the instructions for the task they are trying to achieve,

2) had a world globe in it which could be rotated by the subject,

3) contained a list of place names they had to pin on the map,

4) contained the typing room where they could attach the names to the pins which they then had to pin onto the world,

5) the robot room, this has nothing to do with the test.

They had a set time limit to carry out the task, this added some stress to the task for the subject.

5. Experimental method

The experiment was carried out on a Silicon Graphics Indigo with 32Mbytes memory.

When the subjects started the experimental session they got a chance to use the input device for five minutes in a play world to get a feel for them.

The first six tests were carried out using both glove and bird and were presented at random to the subject to remove any learning effects that might build up in the experiment. The next 3 slightly more complex tasks were then completed, again with glove and bird. The tenth complex task was always presented last as it took more time and needed some competence at the use of the input device. Though it was random in which order the subject used the glove or bird for this test.

In each of the sets of tests an accuracy score was calculated by DIVE.

The aim of the final complex test was to see if a real task could be completed. It was also interesting to see whether total immersion equipment was a problem, and what the effects of fatigue could be.

Half of the subjects carried out the experiment using DIVE in the window mode, the other half used the system in full immersion VR mode.

At the end of the experiment subjects were asked if they felt any ill effects of using the equipment, for example bad back, or neck.

6. Subjects

The subjects were 25 engineering students (aged between 18 - 22) at the Royal Technical Institute in Stockholm, Sweden. The experiment was run as part of a Third year course on User interface design, all the subjects were students taking this course. The experiment was used as a teaching aid to present the concepts of object manipulation in a VR space. The students also used the experiment as a way to study experimental technique in a research setting and to understand presentation of graphics objects.

7. Results

Analyses on distance error and angle error were calculated separately. The results showed that the distance error is not significantly different between the two devices, but the rotation ability is. The DataGlove is significantly better at the 1% level than the probe at rotation tasks.

It was felt that the subjects should have had a much longer practice period as they kept improving (= decreasing) their manipulation times throughout the experiment. They did not however increase their accuracy at all.

The tasks of rotating the ellipsoid and sticking thin objects through a box were apparently overwhelming, there were huge errors for these tasks.

Clearly, the amount of error seems to depend on the task in particular the rotation of an ellipsoid seems to suffer from large errors. On the other hand, the order in which the tasks were performed does not seem to affect the error. It would seem that the subjects tended to perform faster, but not with greater accuracy. Possibly this means that there is a limit to the accuracy they could achieve, but that they reached this limit faster.

8. Conclusions

It was not surprising that there was marked learning happening as the subjects got used to handling the devices. The subjects soon learnt the problems and limitations of each of the devices. They then tried to work around these problems to achieve their task. This was hopefully factored out by the randomizing of the experiments. The final complex task need the experience of the other tests to carry it out, this meant it was the best to present last.

Accuracy while using the devices was very poor, the stack of blocks created by both glove and bird was very inaccurate.

There is a problem when releasing the grasp on a block while using the glove, the action of making the gesture to release the object moves it to an unwanted position. A great deal of care is taken to manoeuvre the object and this work is destroyed when it is released. This is very annoying for the user. It maybe better to take the position of the object a second before the release gesture so any displacement by the gesture is minimised.

The movement of the cursor and the blocks was not very smooth this sometimes caused problems for the subjects.

When the cursor was behind a block it could not be seen by the subject and this meant the subjects could accidentally loose where the cursor was. This was sometimes very annoying for the subjects.

Depth perception was a problem when trying to place the objects in a stack. From one direction they look fine, from other directions they were very poorly balanced. It was much easier to move around the objects in the total immersion environment this allowed for faster more accurate adjustment.

The smooth squashed sphere object was harder to align than the more angular blocks, this was probably because the subject could use the angles to align to the axies of the objects.

Physical effort was needed to carry out the tests with these devices, the subjects got fairly tired after an hour's testing.

When the subject used the window based DIVE system they had four keys to move them around the scene so they were limited to only four views in which to carry out the task. When they used the HMD they could see the blocks from any angle they wished.

In the positioning tasks two had sets of cubes of the same size, this meant the subject could use their size as a factor in deciding if they were in the correct position. The other test had a set of three different sized cubes, this made it much harder to build the stack due to the size differences. The objects could have been the same size but just further away from the subject to make them appear smaller, there is no way for the subject to relate size to distance.

9. References

Carlsson, C. and Hagsand, O.
The MultiG Distributed Interactive Virtual Environment.
Proceedings of the 5th MultiG Workshop, Stockholm, Dec 92.

Carlsson, C. and Hagsand, O.
DIVE - A Platform For Multi-user Virtual Environments.
Comput. & Graphics Vol. 17, No. 6. 1993.

Prime, M. J. 1
Hand tracking devices in complex multimedia environments.
Proceedings of the 6th MultiG Workshop, Stockholm, May 93.

Prime, M. J. 2
Human Factors Assessment of Input Devices for EWS.
Rutherford Appleton Laboratory Report 91-033, 1991.

3D Input Device Usability Evaluation in a Virtual Reality Environment

Kai-Mikael Jää-Aro, Martin Prime and Stephane Royer

• Introduction

– This experiment is designed to evaluate the usability of different input devices in a Virtual Environment. The evaluation involves both speed and accuracy to achive a set of experimental tasks. Two input devices are currently being tested in this experiment: the DataGlove™ — a hand tracking device — and the Bird™ — a flying mouse. The tasks were developed in the DIVE VR environment.

• The input devices

– Work is in progress to develop device drivers for many input units (DataGlove, Magic Wand, GeoBall, Bird and Immersion Probe). The main experiment is to examine the difference between using a 6D mouse type input device and a hand tracking input device. In addition the difference between a desk based device (Immersion Probe) and a free standing one (DataGlove) will be explored.

• The tasks

– The tests are a study of manual dexterity, with the main interest in speed and accuracy to complete the tasks. The test set takes around one hour to complete.

– The first set of tasks analyse the accuracy of positioning objects. Three tests are in this set, each with three blocks. In this set, the subjects have to:

  » make a horizontal line of the set of blocks

  » make a vertical line of the set of blocks

  » make a stack of a large, medium and small block

– The next three tests analyse orientation. In each of these tasks the subjects are presented with two identical objects, one in a fixed orientation and the other a test object. The subjects have to twist the test object until they think it has the same orientation as the fixed object.

– The final three tests are more complex tasks involving both position and orientation.

  » Hole test, the subject has to place a cube in a similarly shaped hole within a wall of other blocks.

  » Inside test, the subject has to place one object inside another, then place both these objects inside a third.

  » The third test in this set involves pushing three rods through the axes of a cube.

– The intention is to use a variable of drop shadows as an additional factor in future tests.

• Experimental method

– The subjects initially get a chance to use the input devices in a play world for a few minutes to get a feel for them.

– In each of the sets of tasks an accuracy score is derived.

– The first nine tests are carried out using both DataGlove and Bird and are presented at random to the subject to remove any learning effects that might build up in the experiment.

– It will be possible for subjects to move their viewpoint and thus improve their accuracy.

• Future experiments

– It is the intention to carry out experiments with more complex tasks, in order to see if a real task can be completed, if total immersion equipment will be a problem, and the effects of fatigue.

– The subjects will be placed in a VR world with gateways to other rooms:

  » A room with the instructions for the task they are trying to achieve.

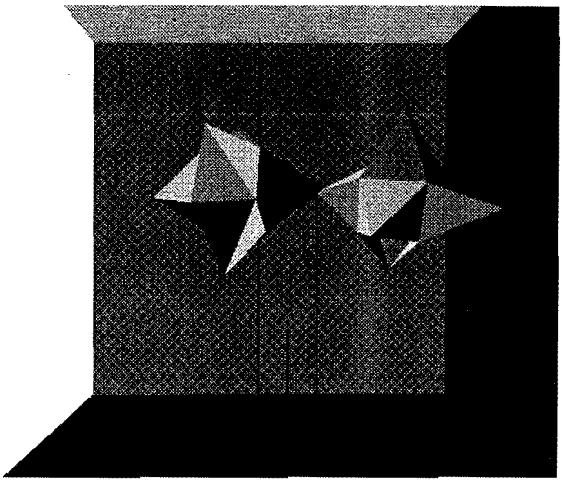
  » A room with a world globe which can be rotated by the subject.

  » A room containing a list of place names they will have to pin on the map.

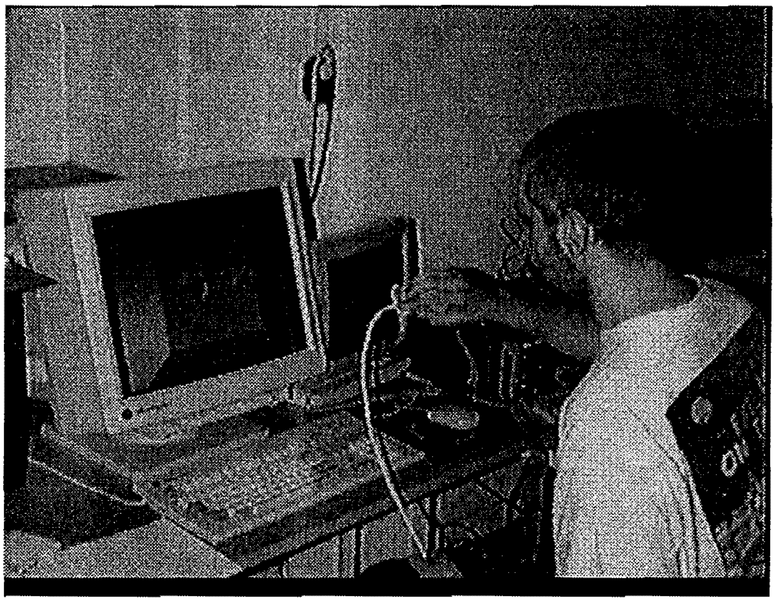
  » A room containing the typing room where they can attach the names to the pins which they then have to pin onto the world.

  » A foil room which has nothing to do with the test.

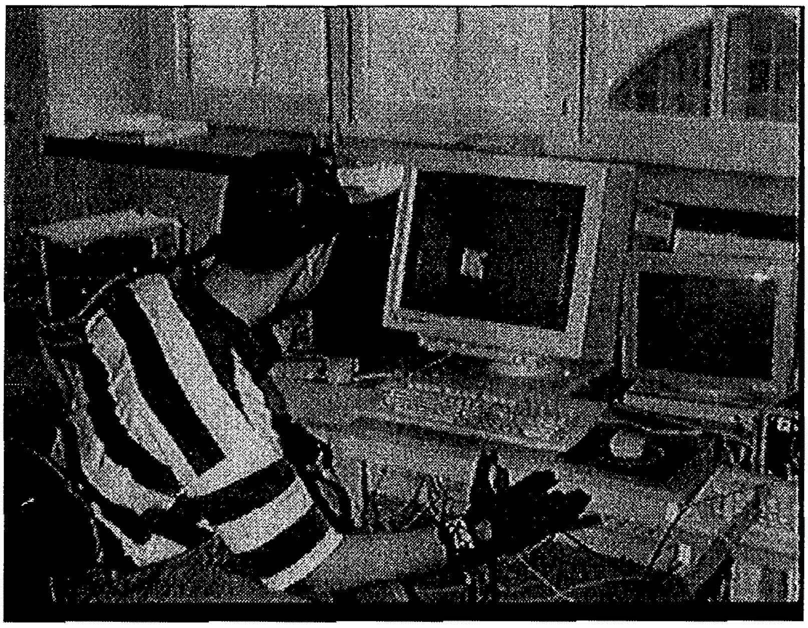
– The subjects will have a set time limit to carry out the task which will add some stress to the experiment.



Two star-shaped objects that have to their orientation aligned. Part of the experiment is too see whether objects with many angles and protrusions are more easily brought into alignment than simpler objects.



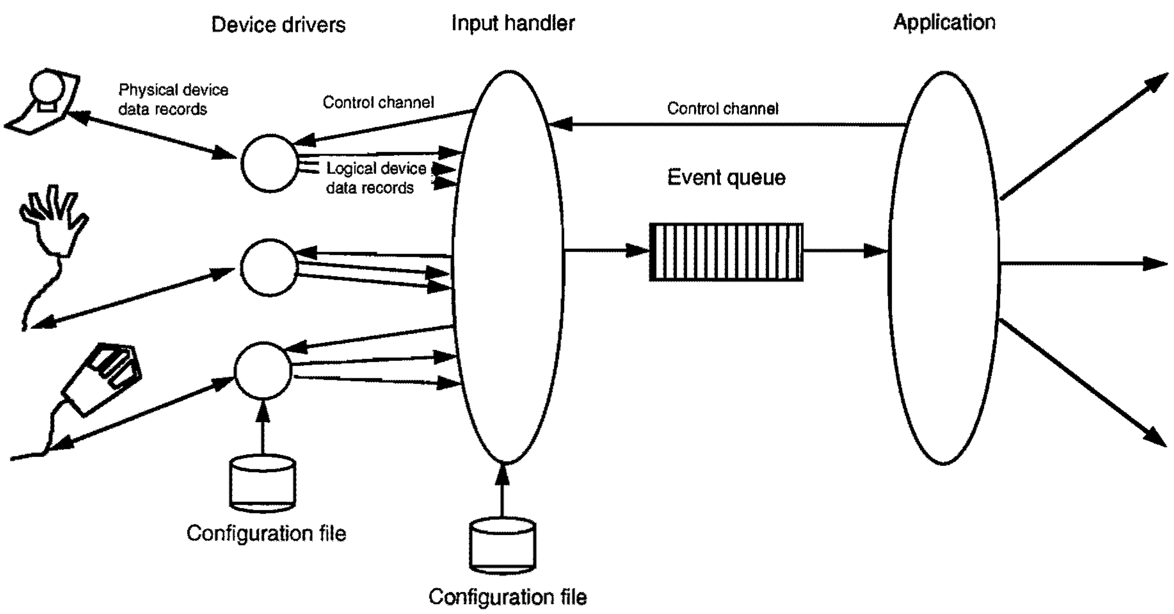
Martin Prime using the Ascension Flying Mouse to rotate objects on the screen.



Vlad Ionesco using the VPL DataGlove to insert differently sized boxes in to each other.

A new input model for DIVE

Kai-Mikael Jää-Aro & Stephane Royer, NADA, KTH



- **Physical level**
  - Corresponds to the physical input devices (DataGlove, 6D Mouse, ...) which continuously transmit data on movement, button presses, gestures, etc.
- **Driver level**
  - For each physical device there is an associated device driver which listens to the device, and then transforms the data into a uniform format (common coordinate system, scale, ...).
  - The mapping of button presses, gestures, ... to COMMAND tokens is done according to a configuration file for each device and particular circumstances.
- **Input handler**
  - The input handler is told by a vehicle what kind of data it requires and starts the various device drivers and instructs them to deliver data to different ports corresponding to the relevant logical devices (POSITION, ORIENTATION, COMMAND, ...).
  - The data from the different devices are synchronised by the input handler and entered into an event queue which is sent on to the vehicles.
- **Vehicles**
  - Vehicles are mappings between events and actions in the world.
- **An Example**
  - A vehicle requires two POSITION devices, two ORIENTATION devices and one COMMAND device. These are then mapped to the position and orientation of the viewpoint in the world and the position and orientation of a 3D cursor which moves relative to the viewpoint. The possible command events are SELECT, GRASP and MENU, which result in messages being sent to objects touched by the cursor.
  - For a particular experiment, we may set up a HMD (Head Mounted Display) to send one set of POSITION and ORIENTATION data, and an Ascension Bird to send the other set, as well as the COMMAND data, defined as button presses.
  - For a different experiment, we may decide to substitute a DataGlove for the Bird and just write a new configuration file without any changes to the vehicle.

Appendix 2

Issues involved with embodiments in virtual reality environments, M. Prime, 1994.

Issues Involved with Embodiments in Virtual Reality Environments

Martin Prime
Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, England
martin@inf.rl.ac.uk

1. Introduction

When a user enters a Virtual Reality (VR) environment they usually need some sort of body to inhabit there, this VR entity is the embodiment of the user. It is the body image which other users of the system will see when they look in the direction of the user. If the user is alone in the system they will still need an embodiment, when they look down at their own body they will need some sort of image to view. The VR environment and thus the embodiment are purely abstractions of the real world and as such it should be remembered that they need not represent real life. The representation of a human need not look humanoid, the embodiment may be that of a lobster in an undersea world for example.

The Virtual Reality environment is a simulation that includes human users as a necessary element, this means that some of the objects will be inhabited by human users. The users actions are monitored by a set of sensors and these motions are transferred to the virtual entity. The embodiment will normally react in direct relation to these sensory movements in the physical world. The embodiment will monitor changes in the virtual world and send this information to the human user. Thus a tight feedback linkage is formed between virtual and physical reality. It is interesting to note that the input sensors of the embodiment will act as output to the human user and an output from the human will provide input to the embodiment.

It is important to remember that non user objects within the environment will also need a representation of some sort to allow the users of the system to interact with it. Similar issues will be important for these non-user entities.

This paper tries to present some important issues that should be taken into consideration when VR embodiments are constructed. The paper does not always answer the questions raised by the problems but poses the questions that should be addressed.

The paper is split into two main sectioned, these describe user issues and environment issues.

2. Self Awareness

This section of the paper address embodiment concerns involved with a user of the VR environment.

When a user enters into the VR world they will be represented there by an embodiment. The question being addressed here is whether the user feels they are actually inhabiting the body or just using it like a robot waldo. The user may feel they are actually involved in the environment, a "sense of presence" (Fontain, 1992), or simply viewing it as if it were on a screen in front of them. It is important the feedback loop between input to the embodiment and output from the embodiment to the user is fast and accurate, this will allow the user to feel more a part of the virtual environment. The more intimate this loop is the more likely it is that the user will feel they inhabit the virtual embodiment.

The embodiment may do more than just create a believable body image it may also give a feeling that the user actually inhabits the creation which it feels is its new body. It is the disassociation of self from the real physical body and into an association with the virtual one that is needed to get a realistic feel for the VR environment.

The embodiment may just provide a level of indirection for the user. If they feel they are simply using a robot it would just be an extension of themselves into the environment and not a feeling of inhabiting an embodiment.

There is a middle ground between fully inhabiting the embodiment and using the embodiment as a robot. This would be a migration into the embodiment but still just using it as a tool to navigate around the VR space.

Personality types may help to predict the way users may employ the embodiment. There is a theory of two types of people Internal and external. Internal - believe they have a effect on the world, everything that happens is under their control, they get a sense of what is up from inside their body. External - believe it is all the worlds fault, they are not in control, fate is what is happening, they get a feeling for what is up from external clues. These two groups of people may need different types of cues to use the a virtual body effectively.

Projected emotion - What emotion do you want to present to the Virtual world? Is this a facial expression or some other coding - maybe colour - red for anger green for passive? These conventions are culturally based though, so they may not be a good idea. Facial expression would be much easier to recognise - but then we do make errors of prediction in the real world as well. How do you predict that someone or some animal (a dog for example) is about to attack? It is the way they move as well as the way they look. It could also be through other perceptions, maybe their smell changes. Again some people are much more perceptive of the world around them and can pick up these small details of danger or other states.

Masks for emotion/information - These have been used in the theatre for centuries to present information about the character who wears it. One change of a mask and the character changes emotion or even changes into another character in the play. For the Orestia in London the actors wore the masks for hours and looked in a mirror to get a feel for what the behaviour of that mask on a person should be like, what the masks personality was - not what an actor wearing the mask would be like. There are also issues here about what are the possible consequence of allowing a user to consciously manipulate the expression that is presented to the other users. How would the user change the mask - with a gesture? The problem with this is that the emotions will not change automatically as they do in real life, the user would have to actively change the facial image or mask. It would take a conscious effort, when it should really be automatic.

Hiding behind a mask - If you wear a mask it can have dramatic effects on your behaviour, people do not know who you are. Others can not identify you. It is possible to be someone else. You could even pretend to be some other VR actor, a charlatan. Some users of MUD systems (Multi User Dungeons, a form of text based virtual reality) enjoy pretending to be a different gender to their own (Curtis, 1992), or even swap between genders to experience living as a different person.

Flaming - When a user pretends to be some one else or just gains anonymity from being embodied in a VR system can provide an environment in which flaming can occur. This is the use of language or behaviour the person would never use in a real life setting. Flaming can be quite aggressive and hostile The person does or say things they would never do if they were themselves. This behaviour may be due to a feeling of disembodiment.

False Front - A person is showing one emotion but actually feeling another, are they telling a joke or are they really angry?

Social groups - People behave differently in groups than on their own. An Awareness of others is what moderates behaviour. Fitting in with social norms, the problem is that these social norms may not have been identified in a VR setting, people may push to see how far they can go. Misunderstanding of other users behaviour may occur. Social gravity - people move to the more populated areas.

Why have a body in the first place - If the world the user enters is just a data base, do not really need a real humanoid body. If it is a meeting room then maybe a fully rendered body is essential for the social interaction to take place.

Decentralise - There is a point in our development as children at which we can "leave" our body and see some thing, an idea or a view of something, from someone elses point of view (need to look up development text book for age) (some people never manage to do this) To get into a virtual body it would seem essential to be able to image moving out of your own position in space in to the VR environment.

Phobias - It is possible that a VR system could be used to treat a neurotic patients phobia. If a patient has agoraphobia, which is a fear of open spaces, they could use the VR system to gradually get used to the experience of being in the open from the safety of a computer system. The patient would be confronted with their worst fears and feel safe enough to face them. This treatment would only work if they actually felt they had migrated in to the embodiment and were really in an open environment. If they just felt they were viewing a picture the experience and thus the cure would not cross over into their real life experience. Another example would be if the patient has arachnophobia, the fear of spiders, the patient could gradually be desensitized to the images of spiders, then to moving images of live spiders. There is another therapy of total immersion of the fear where the user would be dropped in a room filled with, virtual spiders (this is not to be used with users with a weak heart!)

3. The Virtual Environment

This section will cover issues that are concerned with the environment that has been created to be inhabited by the embodiments of the users.

Size, Does the users projected size matter? People who are tall are looked up to, in both a physical and psychological way. Psych tests have shown that people trust and respect tall people more than short. So maybe the size of the embodiment is important.

It is difficult to understand the dynamics of the sun because it is so large, It is difficult to understand the dynamics of crystals because they are so small, It is difficult to understand the dynamics of quantum chemistry because it is so fast, It is difficult to understand the dynamics of geology because they are so slow, In the VR it is possible to change the size of the embodiment or the speed of the action to reduce these problems to a more manageable size for the user. When a scientist can study these phenomena at the right size or speed they can deduce much more about the mechanisms taking place.

Non 3D worlds, If the VR world is not a normal 3D rendering of a real world view could users cope with it. An example of this would be a multidimensional chemistry experiment that the VR system is visualising for the user.

Time, in a cooperative VR environment people will be entering and leaving at different physical times. This means they will be in different social moods. On the IRC it is sometimes strange to have some one from america join the system and say "good morning every one how are we today" when for users in Stockholm it is early evening and time to go home. There will be a distributed user population through time. The sun never sets in a VR world.

Social conventions, A new sets of social conventions will always be set up between humans on any communication media, it is impossible for it not to happen. Some will be good some bad. Users will decide what is acceptable behaviour and make it clear when others overstep that line. We are inherently social animals. Awareness of others as real or virtual actors may also effect the way we respond to them. You could make a mistake and be very rude to a body and then find out they are inhabited by a real person. The social environment will create "softly enforced" social conventions - applied by users not rigorous controls by machine.

Resource allocation, How should the resources of the system be allocated to users, for example if there is a tool of some sort who decides which user should have access at any particular time. It is probable that this will be carried out by social convention. Another issue is how should one interrupt a user with out being rude?

Theatre Magic, It is possible sometimes to get a feeling of escape in the theatre or cinema for example. The viewer can get so absorbed by the scene that they feel they are part of it, this causes heightened emotion and a greater belief in the realism of the virtual environment.

Isolation, In the pelenque multimedia system (reference) children were asked if they wanted to visit an Aztec palace, they said yes but only with a "friend" along as well. The system had a virtual friend to be with the child. They were like an naive guide, they knew nothing about the place they were visiting, just like the user child. The system had other knowledgeable guides (historians, anthropologists etc.) that could be asked questions. But this virtual actor was a complete innocent that removed the feeling of isolation that the child felt.

A Friend, Is someone we call a "friend" someone about whom we can predict their behaviour, and so will feel comfortable in there presence. We can relax with friends.

Safety/fear, What is it that makes us feel "safe" or "fearful in a space? The image of self in a comfortable place surrounded by friends, or a fear of the dark syndrome. Past learning of when we are or are not safe.

Disembodiment, To understand embodiment there is a need to understand what it feels like to be disembodied - a feeling of not being connected to a tangible space or time.

Puppets, is that other actor a virtual person (a puppet of the system) or is it an actual human entity - should you be able to tell the difference? May be there should be some equivalent of the "Turing test" to discover if an embodiment is a real user or just a puppet of the VR system.

Non human form, There is no reason why the virtual body needs to be human, it may not even be living in the real world. The user could inhabit a lobster, a table, the ground, or even someone elses body What would it be like to place a human into a non human form. Could the user cope. Would they act differently or would it be difficult for them to even tell what they inhabited. They may feel they are human, it is only when they look at themselves, maybe in a mirror, that they would realise the body was not human.

Social Chameleon, This is a characteristic of being able to fit into any social situation with ease. Imagine picking someone up and dropping then into a) A croquet party at the big house b) A soup kitchen under the arches of Kings cross in London c) Going to a Thai Brothel with your mother d) A Swedish Anachist Rave Party. It would need a special facility to cope and feel at home in all these situations - you would need a social chameleon. Would you need this facility feel at ease in any thing a virtual world could throw at you.

Degree of actual presence, While in the VR environment the user may encounter the embodiment of another user and try to interact with it. The problem is that there maybe no way of telling if the embodiment is active or not, the entity may appear active but the real life user may have popped out for a coffee leaving the "shell" representation in the environment. They look as if they are there are inhabited but they are not.

Privacy, privacy is easy to provide for users, but what should be made public and what private. Some people will care about some things but not others. There is a need for an embodiment of some sort for all users. If a user does not have a body then it would be possible to spy on others without them realising another user is present. Should all users be forces to have an embodiment of some kind? Voyeur, Is it necessary to have a body if you are just viewing the virtual scene. How would other users feel if they knew that there were voyeur users. Just watching but not taking part. This is related to the issue of users entering the VR system just appearing out of thin air and into the existing users field of view. This may disrupt the feeling of reality the users have already set up. Sentient viewer, This is slightly different from the voyeur, a sentient can feel and get sensations from the VR environment but cannot take part or give out feeling, they are responsive to or conscious of sense impressions, they are aware of the feeling of others but can not effect them.

Temperature, Should the real body get a sensation of the outside virtual world?

Sound, It is possible to identify people by the sounds they make (and smell?) Work out who is coming down the corridor and into your office, just be the noise they make as they walk.

Power seat, In any formal social setting there are physical positions of power. In any meeting table there will be one seat that can command the meeting. Its position is the power not the person in it, a good leader will always look round the table and find the power position and sit in it. (maybe a bad leader as well!) People will find that the body they inhabit in VR will have similar properties and use them to extract power from social interactions. There can also be the inverse power position, it can sometimes be to a persons advantage not to use the power position, it will make the opponent uneasy if they do not realise what you are up to.

Immersion vrs. fish tank VR Are we in the world or just looking in from outside. It will effect the feelings of linking of self to the VR body created for us by the system. Does the user feel like the audience at a play or the real actors.

Navigation Will it be easier to navigate round the world in total immersion or in Fish tank. Navigation through VR worlds to VR sub-spaces. Use cognitive mapping techniques like those used in Cognitive Psych for a long time. Sub space can be created by user for their own work, carry out in privacy, area under their own control.

Futility, In the theatre, when we watch or take part in an activity we expect something to happen, this allows us to get hold of a feeling or image. In the VR environment if users see an object they will expect to be able to interact with it, it is very frustrating if this is not possible The play Waiting for Godot is the most comprehensive account of futility, the two tramps have a dogged desire for something to happen.

Legal status in VR, is a VR signature binding?, held to what you say in VR?
Common objects, is it a trap to think that just because I can see "it" so can you? are we looking at the same thing anyway. This is very important in a CSCW environment, in such a shared environment there must be some way of keeping all objects up to date. There must also be some form of object locking so that two people can not alter an object at the same time and cause potential confusing to the system. It is also possible that a user embodiment may appear different to different users

4. Conclusions

This paper has presented issues that should be address when embodiments are used in a VR environment. Some of the potential problems can be addressed within the system itself, but many are created as a result of human social interactions. It is anticipated that these sorts of problems will be resolved through social conventions by the actual users of the system.

MUDs become true communities after a time. The Participants slowly come to consensus about a common (private) language, about appropriate standards of behaviour, and about the social roles of various public areas (e.g. where big discussions usually happen; where certain crowds can be found) There have even been marriages conducted through MUD systems. All these conventions will be created when multi user VR systems and communities become more common.

The DIVE system (Distributed Interactive Virtual Environment) developed at SICS (Swedish Institute of Computer Science) (Carlsson, 1992) does not support such advanced embodiments at this time. Improvements are being made to the type of embodiments possible and it is hoped that in the near future some psychology experimentation can take place in this area.

5. References

Carlsson, C. and Hagsand, O. The MultiG Distributed Interactive Virtual Environment. Proceedings of the 5th MultiG Workshop, Stockholm, Dec 92.

Curtis, P. (1992). Mudding: Social phenomena in text-based virtual realities. Proceedings of the 1992 conference on directions and implications of advanced computing, Berkely, May 1992.

Fontaine, G. (1992). The experience of a sense of presence in intercultural and international encounters. Presence, 1 (4).

Appendix 3

Integrating sign language into a virtual reality environment, M. Prime 1995.

Integrating Sign Language into a Virtual Reality Environment

Martin Prime
Rutherford Appleton Laboratory
Chilton, Didcot, Oxon, England
martin@inf.rl.ac.uk
Work carried out while at:
IPLab, KTH, Lindstedtvagen 5
100 44 Stockholm, Sweden
martin@nada.kth.se

Abstract

This paper describes the possible implementation of sign language within a Virtual Reality (VR) environment. The Spatial Interaction Model is defined with the concepts of focus, awareness, nimbus and aura to explain how sign language could be used in VR applications. An implementation of this system would provide communication as well as visual access to the deaf in a VR environment. The implementation trade-offs for such a system are addressed. The conclusion is that existing systems are capable of supporting an initial implementation.

Introduction

In a multi-user distributed Virtual Reality environment, such as DIVE (Distributed Interactive Virtual Environment) [2] where the various participants each have a 3-D representation, a model for interaction is necessary. The Spatial Interaction Model [4] describes how objects should interact with each other in the virtual environment. This model has since been extended to cover more interactions and is described in the next section.

Language and gesture can be coordinated to form a single communication system more powerful than either alone. Hand gesture in particular is critical for fast interactive human to human communication [8]. A VR system that allows a sign language channel and thus also a gestural channel will have an enriched communication medium [9].

The main concept of this paper is that in a Virtual Reality environment it is possible to support users with different capabilities for interaction. The factors considered here are whether the user can hear or not; their ability to use sign language; and the technology they have at their disposal on entry into the VR environment. A user is thus represented in the virtual environment as an entity with those capabilities available to it. The interactions that are possible between the entities naturally depend on these different abilities.

Not all people with hearing difficulties are profoundly deaf: most have some residual hearing. It could be possible to help communication for such people purely by increasing sound levels. However, this paper is more interested in the use of sign language for the profoundly deaf in a VR system. The assumption will be made that all deaf people within the system have a knowledge of sign language.

The Spatial Interaction Model

The basic spatial interaction model [5] can be defined in terms of Aura, Focus, Nimbus, and Awareness.

Aura
The aura of an entity or object is defined as the area surrounding it in which it is possible for interactions to take place. It is a predefined sub-space which is usually invisible to the user. When two objects auras overlap in the system it is then possible for some sort of interaction to occur. The aura concept allows objects that are far apart to ignore each other, their auras have not collided so they take no notice of each other.

Awareness
Once auras have collided the type of interactions that are available can be calculated. Awareness represents the amount of contact or affinity between objects in the environment.

Nimbus
The nimbus of an entity are the aspects of it that it makes available for others to use. It is a projector of the information that the entity has available that others might be interested in.

Focus
The focus of an entity is the area where a user directs their attention.

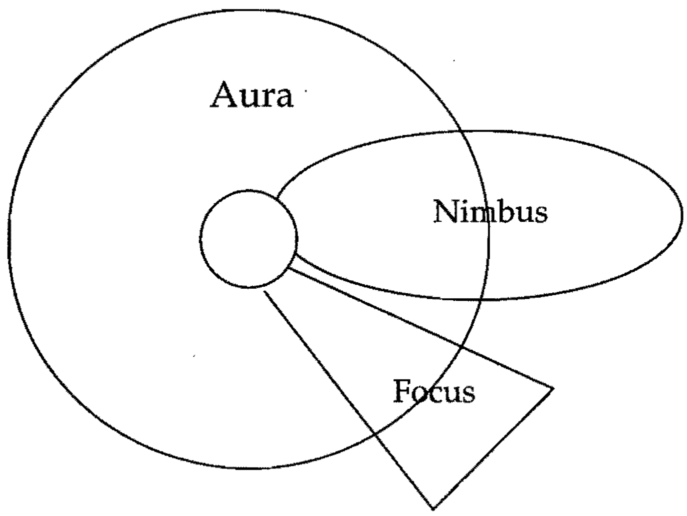


Diagram of the Spatial Interaction Model

The awareness calculations of one user for another only start when their auras overlap. This helps reduce unnecessary calculations in the system. It is only possible for object A to gain any information about object B when object A has its focus in object B's nimbus.

It is not necessary for the nimbus and focus to be within the aura of the user, they can extend outside this region. The overlap of a nimbus with a focus is not a sufficient condition for awareness to take place, the appropriate calculations will only be initiated when auras collide.

The types of user in the signing VR system

There are 4 types of user entity in the system depending on their capability and the technology they have available. These are:

a) HN - hearing but can not use sign language
b) HS - hearing and are able to use sign language
c) DS - deaf and can sign (they have the necessary technology)
d) DN - deaf but can not sign (they do not have the technology)

Note: *S is used to denote someone who can sign and is either hearing or not. i.e. both HS and DS.

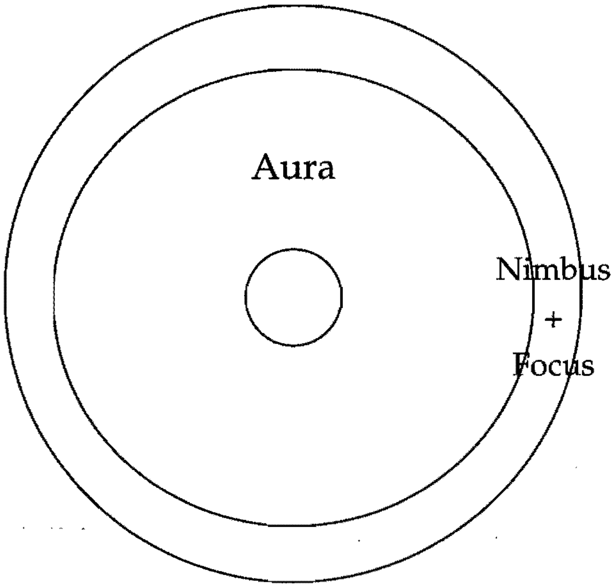
Let us assume for the metaphor that all hearing people can speak and all deaf people cannot produce sound. It is possible that a hearing user can appear deaf purely because they do not have the technology to perceive sound in the environment i.e. they do not have sound detectors. This is similar to the DN user, who can not produce sign language because they do not have the technology.

It is important that the embodiment used within the environment reflects the capability of the user. If the embodiment is known to be truthful then the users can gain all the information the need to know about a user from the body they inhabit. An example of this would be the use of ears, if the body has ears then a viewer of that entity can make the assumption that it can hear. It could be possible that the embodiment just has ears to make it look more life like and they are decoration. It is much more useful for a viewer to use ears as a clue that the entity can hear (this is not of course true in real life). The embodiment would need a pair of well rendered hands and arms to show that the user can produce sign language.

The combination of the spatial model and signing

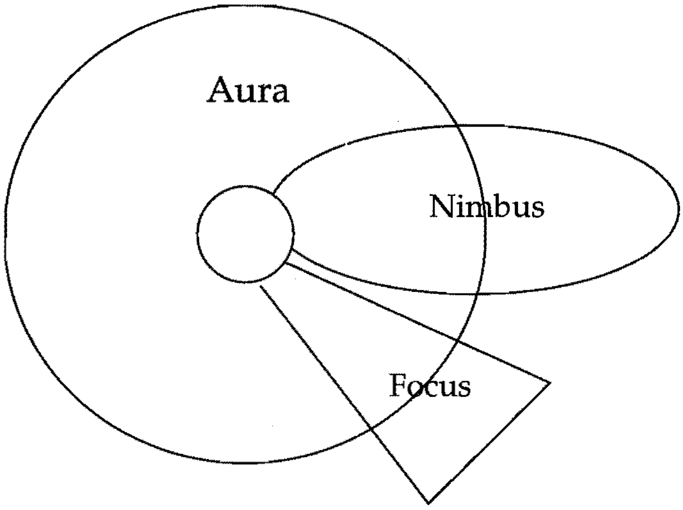
The main difference between using sign language and speech for communication is that sign language is unidirectional whereas sound is multidirectional. This will probably also be true in a VR environment. It could be possible to change this but it seems sensible to keep the real world intact in the VR environment. Overlaying a real world metaphor onto the VR world will aid the users understanding of its operation.

The nimbus shape for sound is different for that of signing. The aural nimbus surrounds the user as does the aural focus.



Nimbus for a sound field

The signing nimbus is only available from the front of the user. The focus in this example is the direction the user is looking.



Nimbus for signing

If someone is to receive sign language they must be looking in the direction of the signer. On the other hand, sound is transmitted in all directions so the receiver does not need to be looking at the sender.

For a *S to use sign language they must have the attention of the recipient. If the recipient is not looking at the signer then no communication is possible and the *S must gain their attention somehow.

An embodiment that does not have arms can be assumed not to be able to sign. A DN can only watch a *S and gain information, they are not able to send communication nor understand a HN.

Some example interactions

It is important to know what communication facilities another user is capable of: are they a signer; can they hear or can they produce sound? This information is provided when auras collide and could be a simple visual symbol: arms and hands for a signer, mouth/ears for a speaker/hearer. There may be problems with this as the user would have to look at the front of the body for a mouth, but it is possible to hear someone from behind before a mouth at the front is seen. An alternative would be to use badges on the embodiment which would be symbolic icons of the capabilities of the user. There is a standard icon of an ear with a red cross over it to represent deafness, a similar icon of arms with a cross through it could represent a non signer. These icons may be more polite than the removal of ears or arms of the embodiment but could take longer for a user to spot and identify.

For an H to be heard by another H they have to simply move their nimbus into the focus of the other and then speak. If the recipient has their attention switched to sound then they will hear the speaker.

It is not this simple for a signer, as the recipient has to be looking in the appropriate direction. If the signer does not want to move round to the front of the recipient how does a signer get the attention of the recipient? Normally in the real world a signer has to touch the recipient, but this will not be possible in a VR environment until tactile feedback is available. A way round this problem is that when the nimbus of a signer over laps the focus of the recipient an arrow will appear in the visual field of the recipient indicating the position of the signer. The recipient can then turn to move the signer within their attention area. They will then be able to communicate with the *S, if they are themselves a *S.

Implementation issues

A device driver for a glove input device already exists within the DIVE environment. A simple rendered hand is controlled by the input device. Currently only one hand is in use as gloves are expensive, but there is no difficulty in providing a two-handed system.

An alternative method to allow sign language to be visible in the environment would be through the use of video. Device drivers are being written to permit video to be shown on screens in the virtual reality space within DIVE. Instead of the user having rendered hands in the environment they could have a video image of their real hands. The user of the system would have a video camera pointed at their torso to pick up the hand movements.Viewers of the embodiment would then see the video image on the chest of the user displaying the users hands carrying out the signing. This is a much simpler and possibly cheaper method than using fully rendered hands and arms combined with glove input devices.

In a VR environment it is not necessary to stick to real world concepts, for example there is no real need to have a video of the arms in the position of the real arms of the user. The video screen could be placed above the head of the signer this may make it easier to view the sign language from a distance. The area above the head is physically unoccupied and spatially associated with communication. It is also important to remember that it is possible for each user of the VR environment to be provided with a different representation of the world. This would mean that it would be possible to have all the video screens above the heads of the other users turned towards the viewer. Each user of the system would have the view that all the other users of the system were communicating with them alone. This also partly removes the problem of gaining a users attention and then signing to them. It takes the signing world into a similar state to the sound world, where voice is carried in all directions at the same time.

A base level for this system would be equivalent to video conference where each individual is represented as a pair of signing hands on a screen. All speakers have an equal status, which could be confusing and prevent private conversation. This does raise problems of privacy within sign and sound communication, the answer would be to allow only selective viewing or hearing of particular conversations. This may be provided by the user selecting the receiver or receivers of their communication. A VR system would also overlay a spatial dimension on communication, allowing the physical groupings that occur in normal social contexts for conversation.

As well as the VR system allowing sign language communication it must also be remembered that it provides a channel for the non-verbal gestures used in everyday communication between humans. These gestures may help to remove some of the confusion, misunderstanding and frustration that sometimes occurs in computer communications[9].

Lip reading is a fundamental part of communication for many deaf people. While this paper is fundamentally concerned with sign language it is useful to note that the implications for lip reading are similar to those for signing. Lip reading may be made more difficult due to low light levels, obscurement, obliqueness and distance. VR would allow an image of the speakers lips to be presented to the recipient always under conditions of perfect visibility. The video image of the lips could be placed on the face of the embodiment but again the video screen may be better placed above the head of the user. Two lip readers either side of a speaker would each get a square-on view of the lips, not physically possible, but possible in a VR environment. VR can remove restraints of this sort. Different virtual realities are presented for different observers in the same virtual space.

Translation

A VR environment can be a leveller within society. Just as a swimming pool can put a physically handicapped person and an able-bodied one on a par, VR could extend this egalitarian function to those with communication difficulties.

It should be technically possible to carry out some form of image analysis of the sign language input and translate this to audio output. It is possible that image analysis of the video image of hands may be superior to glove analysis. Similarly an audio to sign language analysis could be developed.

Work has already been carried out to use glove input to interpret the hand motion and drive a speech synthesizer [6]. The approach used a three-stage back propagation neural network trained to recognise gestural words. The paper also describes an analysis of hand-to-language mapping at various levels of granularity, from using hand motions for the control of parameters of an artificial vocal tract, to interpreting whole hand motions as words and concepts.

There has been less work carried out in the area of recognition of video images. What is really needed in this system is a method for all people to interact with computers without the need of encumbering equipment. Using custom hardware to process silhouette images [7] it is possible to overcome some of the image processing speed problems. The techniques are successful at recognising parts of the body, head, legs, arms, fingers, as long as they can be seen in the silhouette. One of the goals of the system is to develop an entire computer-based workspace that requires a minimum of mechanical devices, instead relying on vision techniques to interpret the user's hands and body. The main limitations of the system as it stands is the spatial and temporal video resolution, and separation of foreground from background in a cluttered environment.

Any distinction between those who could/couldn't sign, could/couldn't hear, or could/couldn't speak would be lost. The system would thus support communication between those with incompatible capabilities in the real world. There would be an unprecedented breaking down of barriers.

Conclusions

This paper has presented the major issues involved in using sign language and gesture in a VR environment. Although the system is not implemented it would seem to be relatively trivial to create an initial video linked sign language environment when the DIVE video channel has been written.

One of the main problems in any system is for a signer to get the attention of another user. Only when the system is in use will it be possible to determine the extent of this problem. It seems possible that, as in the real world, social conventions will be set up by the users of the system themselves to solve such problems.

Mechanisms to modify the range of the nimbus and focus have been created for special situations in some existing systems [5]. One example of a nimbus extension in this system would be for a user to broadcast their signing to more than one user. A large virtual video screen could be used to provide this function. When the user is close enough to the broadcast screen they will have the image of their video hands transferred to the larger screen. In this way many viewers will simultaneously be able to have their focus in the larger nimbus of the screen thus allowing a large group of users to see the signing at the same time. Having to be near the screen creates a transparent prioritisation mechanism to stop multiple users from broadcasting at once. Many other scenarios of this type are possible.

The DIVE environment is a distributed system and allows access to users from different geographic locations to interact together [3]. Long distance networking like this would seem to be highly important for the deaf where isolation can be all too common.

It would be a challenge to extend a virtual reality environment to allow access to blind users. This would need extensive additions in the area of sound field and tactile feedback. For an H 'spatial audio' is very important, spatial audio is sound that has been processed to give the listener a sense of the location of the virtual sound source and the characteristics of a virtual listening space. True binaural 3-D audio, when presented over headphones, appears to come from a particular point in the space outside of the listener's head [1]. Spacial and 3-D audio can be useful whenever a listener, blind or not, is presented with multiple auditory streams, requires information about the positions of events outside of the field of vision, or would benefit from increased immersion in an environment.

Acknowledgments

I would like to thank the ERCIM Internal Fellowship program for making this research possible.

References

[1] Burgess, D. and Verlinden, J.
An Architecture for Spatial Audio Servers.
VR Systems '95, New York.

[2] Carlsson, C. and Hagsand, O.
The MultiG Distributed Interactive Virtual Environment.
Proceedings of the 5th MultiG Workshop, Stockholm, Dec 92.

[3] Carlsson, C. and Hagsand, O.
DIVE - A Platform For Multi-user Virtual Environments.
Comput. & Graphics Vol. 17, No. 6. 1993.

[4] Fahlen, L. E. and Brown, C. G.
The use of a 3D Aura Metaphor for Computer Based Conferences and Teleworking.
Proceedings of the 4th MultiG Workshop, Stockholm, May 92.

[5] Fahlen, L. E., Brown, C. G., Stahl, O. and Carlsson, C.
A Space Based Model for User Interaction in Shared Synthetic Environments.
Proc. of INTERCHI'93: 43-48.

[6] Fels, S.
Building adaptive interfaces with neural networks: The glove-talk pilot study.
Dept. Computer Science, University of Toronto, Technical report CRG-TR-90-1, Toronto, Canada, Feb 1990.

[7]Krueger, M.
Artificial reality (2nd ed.).
Addison-Wesley, Reading, MD, 1990.

[8] Prime, M. J. 1
Hand tracking devices in complex multimedia environments.
Proceedings of the 6th MultiG Workshop, Stockholm, May 93.

[9] Prime, M. J. 2
Human Factors Assessment of Input Devices for EWS.
Rutherford Appleton Laboratory Report 91-033, 1991.