February 1992

The Reality of Virtual Environments: WPE II Paper

Rebecca T. Mercuri
University of Pennsylvania

Recommended Citation

Rebecca T. Mercuri, "The Reality of Virtual Environments: WPE II Paper", . February 1992.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-92-10.

The Reality of Virtual Environments: WPE II Paper

Abstract

Recent advances in computer technology have made it now possible to create and display three-dimensional virtual environments for real-time exploration and interaction by a user. This paper surveys some of the research done in this field at such places as: NASA's Ames Research Center, MIT's Media Laboratory, The University of North Carolina at Chapel Hill, and the University of New Brunswick. Limitations to the "reality" of these simulations will be examined, focusing on input and output devices, computational complexity, as well as tactile and visual feedback.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-92-10.

The Reality of Virtual Environments
WPE II Paper

MS-CIS-92-10
GRAPHICS LAB 49

Rebecca T. Mercuri

University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

February 1992

The Reality of Virtual Environments

WPE II Paper

Rebecca T. Mercuri
University of Pennsylvania

Recent advances in computer technology have made it now possible to create and display 3-dimensional virtual environments for real-time exploration and interaction by a user. This presentation will survey some of the research done in this field at such places as: NASA's Ames Research Center, MIT's Media Laboratory, The University of North Carolina at Chapel Hill, and The University of New Brunswick. Limitations to the "reality" of these simulations will be examined, focusing on input and output devices, computational complexity, as well as tactile and visual feedback.

Preface

Virtual reality, virtual environments, cyberspace --- these phrases invoke an imaginary world where users can transcend physical barriers and become travelers in the minutia of human blood vessels or the vastness of outer space. William Gibson, in *Neuromancer*, described cyberspace as "a consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts... A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged in the nonspace of the mind, clusters and constellations of data. Like city lights, receding...."[Gib84] In its ultimate form, individuals would interact with humans and computers in an experiential illusion, appearing to share the same space, although they may actually be hundreds or thousands of miles apart.

These ideas are hardly new. Ivan Sutherland, in 1965, described the ultimate display as "a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked."[Suth65] Researchers are hard-pressed to separate facts from the colorful hyperbole expressed by scientists who have joined camps with Timothy Leary [Mart91] and other futurists, encouraging us to believe that new-age technology will be able to provide us with drug-free illusions, safe virtual sex, and no-death wars.

The question upon us is to determine which portion of this is feasible using today's engineering capabilities (and that which could be achieved some number of years hence, say within the next decade). Are the limitations imposed by bandwidth, input/output devices, and computational complexity so insurmountable as to preclude the possibility of maintaining the realism necessary to produce desired psychosomatic responses? This paper explores these issues as propounded by some of the major researchers in the field of virtual environments.

Specifically, I will begin by establishing some basic definitions of terms and discuss the potential applications of this technology. Then, I will provide an overview of the basic construction of a virtual user interface, drawing from the classic paper by NASA scientists Fisher, McGreevy, Humphries, and Robinett. Next, I will review the general problem of how a user manipulates a camera in virtual space, through the work of Ware and Osborne at the University of New Brunswick. Force feedback devices and algorithmic methods of presenting surface textures will be examined through the work of Minsky, Ouh-young, Steele, Brooks and Behensky at the University of North Carolina. The proposals by MIT Media Lab's researcher, Alex Pentland, regarding the possible simplification to linear time of well-known computationally intensive tasks such as object dynamics, collision detection and constraint satisfaction, will be considered. Throughout, I will compare the assertions of the authors with expected and achieved theoretical and experimental results. Finally, I will conclude with a general assessment of virtual environment technology, with special consideration made to points emphasized by the key papers reviewed herein.

Introduction

A panoply of terms have been used to describe the virtual environment medium, many attached by trademark claims (which, more often than not, have been denied or revoked). I will attempt to clarify the jargon here for purposes of this paper, although it should be recognized that the distinctions are far from being well-defined, even among articles appearing in the same journal or conference proceedings.

Much debate has ensued over what actually constitutes a virtual *reality*, the only common ground appearing to be that the user must somehow experience the sensation of manipulating objects (typically visual, auditory, or tactile) in real-time, in a three-dimensional space. The use of the term virtual *environment* implies that additional restrictions are imposed. First, inputs from the real world to the user for at least one sense must be occluded and replaced by the displayed (virtual) material. Second, consistency must be maintained between objects that are being examined from different angles by the user, and all other objects that are currently in "view". The difference between a virtual environment and cyberspace is largely one of degree. *Cyberspace* is reserved to describe a complete sensory experience where the user believes that he or she is physically a part of the virtual space. A virtual environment need not require this feeling of total immersion, and only one sense may be involved (although a variety of sensory inputs and outputs are often combined to enhance the perception of realism and reduce undesirable side-effects such as nausea).

By these definitions, the Jack™ software developed at the University of Pennsylvania [Phil90] would qualify as being capable of producing virtual realities, since objects can be formed and moved within a three-dimensional matrix. The Jack™ software alone, as viewed on a terminal, can not presently simulate a virtual environment, since user input manipulates only the image on the CRT, and all other objects in the surrounding area (including those in visible but inactive windows on the screen) remain stationary. On the other hand, a room outfitted with equipment that could display a life-size, computer-generated holographic image for viewing at any vantage point relative to its projection in real space [Emm91] might be considered to form a virtual environment, as it would be possible to walk around (or possibly even into) the hologram and examine it as if it were an actual item within the domain. Note that in order to conform with the requirement for occlusion, the hologram must be opaque rather than transparent or translucent. Multiple holographic displays would have to deal with hidden-surface removal problems.

Applications

A wide range of applications have been proposed and explored using virtual environment technology. These generally fall into various categories, with teleoperation, simulation, and imaginary worlds encompassing the bulk of current research. I will elaborate on these areas below.

Teleoperation involves the use of machines that are controlled by human operators, working at a remote site location. Such equipment permits physical tasks to be carried out safely in a hostile environment. A remotely operated machine may be located as close as the next room, or as far away as a distant planet. As early as the 1940's, mechanical teleoperators were employed to remotely manipulate radioactive materials. More recently, the exploration of the Titanic wreck was performed by teleoperators, and the Hawaii laboratory of the Naval Ocean Systems Center is currently working to expand the uses of these devices. The addition of human perceptual, cognitive and motor abilities to a constantly varying task has been found to provide an advantage over totally autonomous robotics.[Utt89] Humans are also better at operating under degraded information conditions, such as those experienced while navigating an aircraft, than are fully-automated systems.[Cham85]

Virtual environments may be used in combination with computer-generated simulations of real-world conditions, many of which would normally be inaccessible to humans in traditional surroundings. The radiation planning simulation produced at the University of North Carolina allows a physician to view computerized scans of a patient's body through a 3-dimensional display system in order to optimally position gamma ray beams in the treatment of cancerous tumors within the lungs. A virtual racquetball simulation at Autodesk's Cyberspace Project permits handicapped persons to compete with human or computer opponents. Tom Furness, at the University of Washington's Human Interface Technology Lab, has experimented with what he calls *televirtuality*, to provide access to shopping facilities for elderly and physically challenged individuals. Michael McGreevy and others at NASA's Ames Research Center used data collected by Viking orbiters to produce a simulated "flight" through the Valles Marineris of Mars. This sort of simulation is often referred to as *telepresence*, the use of virtual reality to place the viewer within remote scenes. CAD/CAM and other design systems are enhanced by the ability to examine a constructed object from numerous vantage points.[Stew91]

The computer-generated video or audio display may be presented in such a manner that it only partially masks, or is superimposed upon, the natural environment. Certain forms of goggles project images to an otherwise transparent surface, and headphones can be made of materials that permit sounds to enter from the outside. These devices (which are not the focus of this paper) provide a means of connecting virtual and real environments, which can be useful for activities, such as air traffic control, that involve the manipulation and orientation of objects using traditional as well as virtual display equipment.[Stew91]

Objects displayed by a computer system need not be constrained to obey the rules of physical reality. For example, opaque items could be made transparent, and substances could be made to have negative mass.[Suth65] In an imaginary world, one could travel to prehistoric times, or inhabit the body of some other creature. Carnegie Mellon University's Oz Project explored the possibilities of interactive fiction, using computer-generated characters and content that could adapt to the actions of real-life participants in a story's plot.[Stew91] Alan Kay, at MIT's Media Lab, developed the Vivarium -- a virtual aquarium with a life-like ecology environment where people could interact with cartoon images.[Bran87] The Battletech Center is a commercial application of virtual environments, where (for a modest $7.00 fee) participants can compete against each other, alone or in teams, in a high-tech war game.[Mac90]

User Interface Devices

The majority of virtual environment presentation systems now follow the design used by Fisher, et al in their research at NASA Ames in the mid-1980's. Their goal was "to develop a multipurpose, multimodal operator interface to facilitate natural interaction with complex operational tasks and to augment operator situational awareness of large-scale autonomous and semi-autonomous integrated systems." They hoped to devise a uniform interface which would allow multiple task supervision, while offering human-matched displays and controls for ease of use and training, and reconfigurability to suit varying levels of operator skill and preference. Intended applications included: cockpit automation, space station automation and robotics, workstations for telerobotics and telepresence control, supervision and management of large scale integrated information systems, and human factors research.[Fish86]

Essentially, the system consisted of a central computer with various output and input devices. Output (to the human) was produced through a wide-angle stereoscopic display unit, 3D sound cueing and speech synthesis, and computer graphics and video image generation equipment. Input (to the computer) was provided using a multiple degree of freedom glove, connected speech recognition, and gesture and position tracking devices. I will now detail these components and discuss some of their limitations.

The display system took the form of a helmet, inside of which was mounted two medium-resolution, monochromatic, LCD screens. Images were presented to each eye through wide-angle optics, providing an effective field of view of 120 degrees (horizontal and vertical) with a common binocular field of up to 90 degrees. Binocularity was produced using parallax cues derived from different horizontal viewpoints of a 3D computer image, or by two remote video cameras transmitting separate views of a real-world scene. Calibration could be made to accommodate the interocular spacing of different users through an electronic shift of the computer-generated display, or through repositioning of the stereo cameras. The image was presented in NTSC standard video format. An alternate form of the display system was available in a workstation unit mounted on a movable arm.

Auditory feedback was given to the user to indicate task, system status and navigational information. It was felt that enhanced situational awareness would be produced through the use of auditory cues. Speech synthesis with unlimited vocabulary capability was used to generate voice reports and to acknowledge system input. Additional sound cues, localized in the virtual 3-space surrounding the user, were provided through stereo headphones. These cues were designed to maintain their spatial positions as the user moved about in the virtual environment.

The user could direct the system verbally, through connected speech recognition. Glove-like devices could be worn that transmitted gestures through flex-sensing devices at each finger joint, between the fingers, and across the palm of the hand. Additional motion-tracking sensors were placed on the hands and arms to yield position data. The computer maintained a 3D database of an articulated hand which corresponded with the viewer's hand, and was directly controlled by sensor transmissions from the glove, enabling the user to "pick up" and manipulate objects in the virtual environment. The head motion of the user was tracked in real time using a helmet-mounted device which provided information regarding 6 degrees-of-freedom. This position and orientation data was used to coordinate the displayed stereo images with the head activity.

A virtual environment system used for teleoperation, as implemented by the Naval Ocean Systems Center, is depicted in Figure 1.[Utt89] The NASA system is similar in style. The head unit is unwieldy and forward-heavy (although more so in the NOSC version), and data gloves and sensors transmit digital signals instead of the hydraulic control signals used by NOSC. I included these photographs because the NASA illustrations in the Fisher paper failed to exhibit the cumbersome appearance of the user which results when entrapped in a system containing multiple sensors and wires.

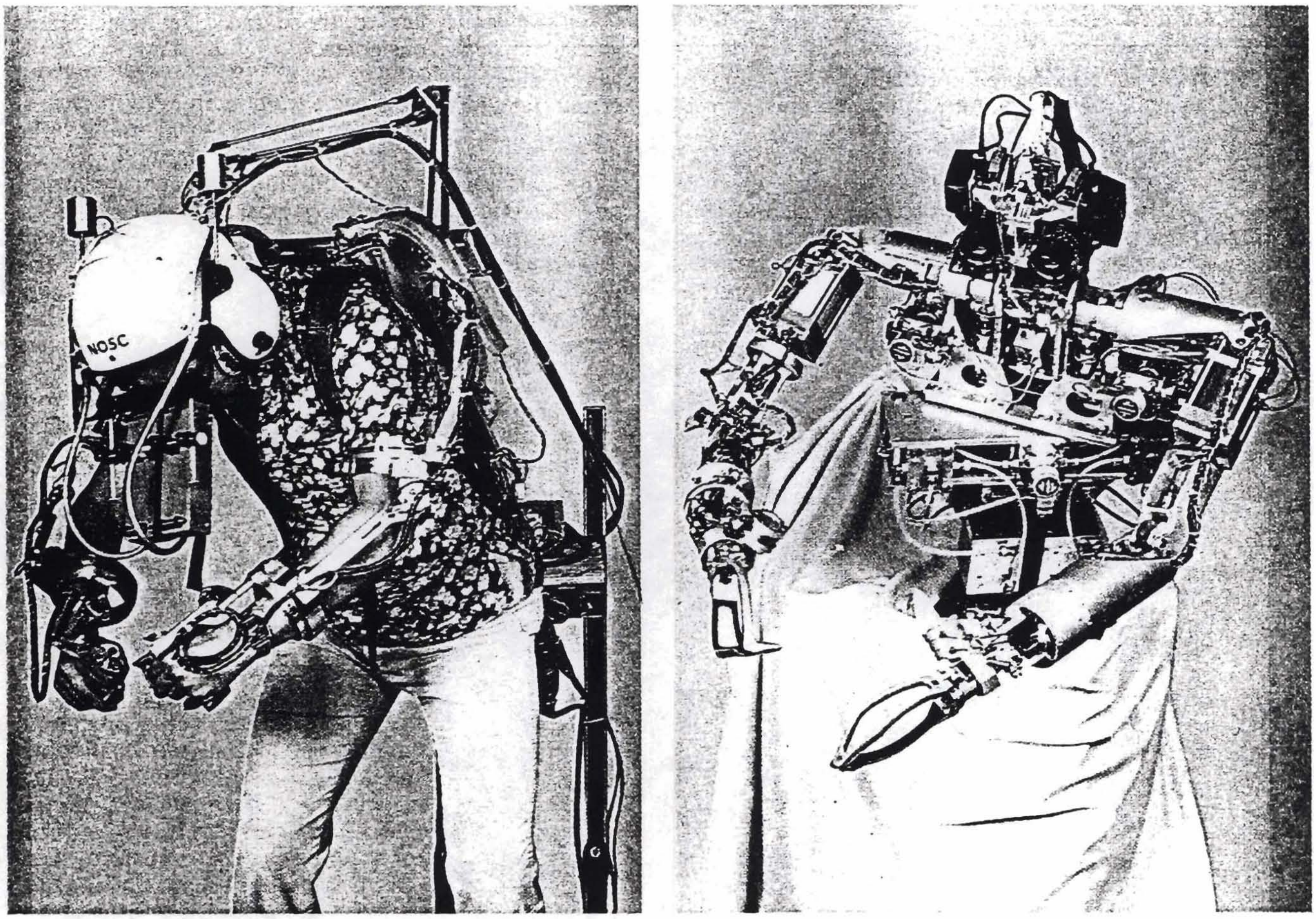


Figure 1

VPL Research offers a commercial version of this system, although it is hardly consumer-priced. The goggles (EyePhones) are priced at $9,400, gloves (DataGloves) are $8,800, a full length DataSuit is ~$100,000. A complete, single-user system including computers and multidirectional sound is priced at $250,000.[Stew91]

The NASA head tracking system was based on earlier research by Ivan Sutherland at Harvard and MIT. Sensors mounted on a helmet worn by the user transmitted signals that were picked up by a receiver mounted on the ceiling. Translation and rotation information was measured within the resolution of 1/100 of an inch, and one part in 10,000 of rotation. Distortion was about 3%. Although these tolerances seem to be satisfactory, they resulted in a display error of as much as 3/10 of an inch which, depending upon the application, could be critical. Furthermore, motions were restricted to an area six feet in diameter and three feet high.[Suth68]

Wang, et al recently devised an 3D tracking system which promises slightly improved performance over a larger working volume (1,000 cubic feet). Instead of mounting the emitters on the helmet and the receptor in the ceiling (a scheme which they call outside-in), they placed a grid of flashing infrared LEDs on the ceiling, and three photodiode receptors on the user's helmet (inside-out). Achieved resolution was approximately 0.1 degree rotational, and 2mm translational. What was most impressive was the fact that the lag time in computing the user's position was reduced to about 5ms, or 200 updates in a second using a micro-Vax-II (outside-in systems traditionally have poorer performance, the Polhemus 3D position tracker, for example, only providing 60 updates per second). Helmet size and weight continue to be a problem, this unit weighing in at 1kg.[Wan90]

If television-style transmissions are to be used, bandwidth and long-distance delay become problems which must be taken into consideration. Without data compression, a standard color TV frame contains nearly 3/4 of a million bytes of data. Researchers at Intel have achieved compression to 4500 bytes/frame (1.2 Mbits/second), using an asymmetric, lossy process. Difficulties with this method include the fact that although decompression can be performed in real time, the returned resolution is reduced by a factor of two in each of the horizontal and vertical directions. Additionally, the delta process used for the compression takes about two seconds per frame. They have a different compression algorithm which can reduce a frame in real time but the results are even more lossy.[Rip89, Tink89, Keit91] Ongoing standardization efforts by CCITT, ISO and other organizations are expected to spur on further algorithmic and technical advances in video compression.[LeGa91]

Let us assume that it were possible to achieve compression to 4500 bytes/frame, with reasonable image quality, in real time (the reader should note that this number is useful as it is related to the speed at which a CD player can read data from a disk). This is well within the data rate of 1.544 Mbits/second that has been defined for the North American DS-1 fiber optic interface.[Hac89] Neither is this a problem for satellite transmissions, which can encode a single 50 Mbps data stream per transponder. The difficulty encountered with both methods is propagation delay. For the satellite, end-to-end transmission time is between 250 and 300 msec.[Tan88] For a 1 Gbit/second fiber optic link, the propagation delay in spanning the United States is estimated at 15 msec. under typical response time.[Kle85] For time-critical control operations, these delays may be unacceptable.

With regard to the other components of the system, it is certainly possible to generate displays that produce reasonable stereopsis, and audio for intelligible speech synthesis. On the other hand, rigorous algorithms to perform continuous speech recognition for multiple (or even single) users over a substantial vocabulary have eluded researchers for decades, and had surely not been resolved in 1986, when the NASA paper was written. The statement that "the system includes commercially available connected speech-recognition technology that allows the user to give system commands in a natural, conventional format in contrast to highly constrained discrete word recognition systems or keyboard input"[Fish86] is somewhat questionable, as it implies that a natural user interface was available for verbal input. Connected speech systems at that time were as highly constrained (if not more so) as the discrete systems were in terms of vocabulary as well as speed of processing. Years later, although advances had been made, this was still viewed as a computationally intensive task which was not yet fully understood.[You89]

The authors also indicated that 3D sound cueing was incorporated in the system. Suffice it to say that in 1990, a paper was published from the same research group within NASA Ames entitled "Challenges to the Successful Implementation of 3-D Sound", indicating problems with such things as front-back reversals, intracranially heard sound, localization blur, data reduction, and low frequency response characteristics.[Beg90] The use of headphones, rather than loudspeakers, for sound presentation introduces further complications due to the fact that the filtering effect of the pinnae (convolutions in the outer ear) is subverted by presenting the stimulus directly to the ear canal. Some success in overcoming this problem has been achieved through the use of Head-Related Transfer Functions, individually tuned to the listener, and applied to the signal during the generation process.[Beg90, Wenz90] Extremely small earpieces are available which can be placed in the ear canal, so weight and pressure is probably only of minimal concern.

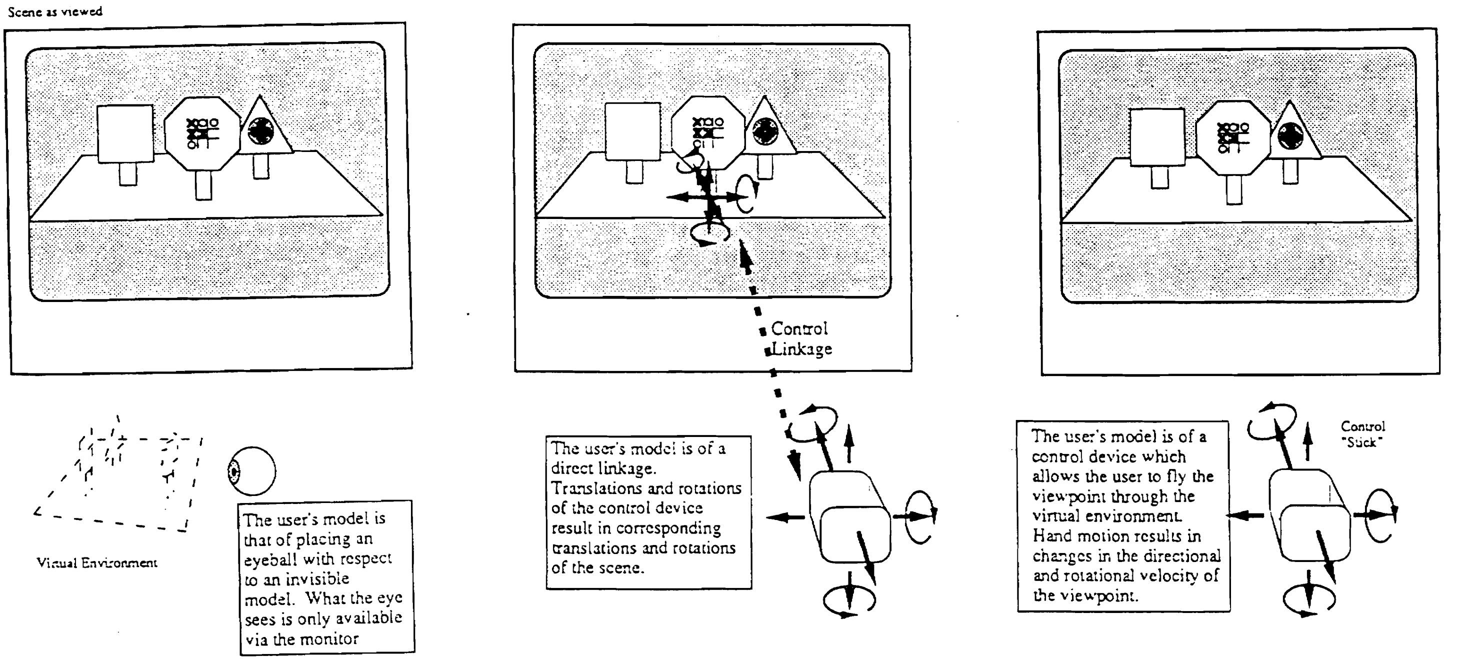
There is no doubt that Fisher, et al developed and demonstrated a functional virtual environment display system, and that their model was a viable one for some applications. I contend, though, that the system they implemented was hampered by fundamental limitations, none of which were addressed, or even alluded to, by the authors. Perhaps it would have been somewhat more honest if the researchers had described the system as a "work in progress".

Camera Control

It has been observed that users of virtual environment systems tend to rapidly become accustomed to the display devices and experience the presented world as if it were, to use Myron Krueger's oxymoron, an "artificial reality".[Stew91] As part of the adaptation process, the user develops a viewpoint from which the virtual scene is observed. In real life we are not conscious of the images of objects moving on our retinas as we pivot our heads in space, rather we recognize that the objects remain stationary as our heads move. In fact, a complex mechanism called the vestibulo-ocular system coordinates the motion-activated stimuli transmitted from fluid-filled canals in the ear with rapid (saccadic) and slow (tracking) eye movements in order to assist in image stabilization.[How82] This reflex, as well as other, more complex ones involving coordination of visual, somatic and auditory stimuli to the brain, are common to various species including mammals.[Spa87] The incapacitating nausea or "simulator sickness" experienced by some teleoperators and users of virtual display devices has been attributed, by some, to the disparity between the operator's motion perception and the lack of inner-ear stimulation.[Utt89] Further research in this area is indicated.

One step in the direction of understanding human perception of virtual environments involves the examination of metaphors developed for exploration and virtual camera control. Colin Ware and Steven Osborne identified three such metaphors and implemented them in "toy" graphical environments using an IRIS workstation and the Polhemus 3Space™ Isotrak™ input device. They observed the users' reactions to the application of various metaphors in conjunction with different environments, and categorized them as to their suitability in the performance of specific tasks.[Ware90]

The 3Space™ input device consists of a 6 degree of freedom spatial sensor that encodes 3 forms of position placement and 3 forms of angular placement. This is sufficient to place a viewpoint within a 3D environment if you omit the additional degree of freedom required to allow for a view scale factor (zoom). The researchers mounted the 3Space™ unit in a hand-sized, rectangular case that included a button, permitting it to be used in "mouse mode" (changes are registered while the button is depressed). A ratcheting feature allows the user to perform a sequence of hand moves, releasing the button in between, so that the controller can be kept within a comfortable arm position when large motions across the scene are required. The 3Space™ is a low-frequency, electromagnetic device, with cables connecting the source and sensor components.[Iso85] The authors did not address whether the cabling posed any maneuverability problems, or if the equipment had been modified to eliminate the wires.



The user's model is that of placing an eyeball with respect to an invisible model. What the eye sees is only available via the monitor

The user's model is of a direct linkage. Translations and rotations of the control device result in corresponding translations and rotations of the scene.

The user's model is of a control device which allows the user to fly the viewpoint through the virtual environment. Hand motion results in changes in the directional and rotational velocity of the viewpoint.

Eyeball in hand metaphor

Scene in hand metaphor

Flying vehicle control metaphor

Figure 2

The three metaphors are identified as "eyeball in hand", "scene in hand", and "flying vehicle control" (see Figure 2). The "eyeball in hand" metaphor refers to the use of the 3Space™ device as a virtual video camera or eyeball which can be moved about the virtual scene. Similar to the Jack™ system[Phil90], the graphical image is viewed from the vantage point of a hand-held eye which may change its position in the 3D space. (Indeed, the authors reference the 1986 paper by Badler, Manoochehri and Baraff in this context.) In the "scene in hand" metaphor, the entire scene is linked to a specific point within the viewed scene. All motions of the 3Space™ become translations and rotations of the scene about the link point. The "flying vehicle control" uses the 3Space™ as a control device for a virtual vehicle. Actual flight, with acceleration, is not truly modelled, only velocity may be manipulated as the cube of the displacement of the controller.

The three "toy" environments were designed to permit the users to experiment with different task domains. Within each environment, three areas of detail were contained, which the user was to examine. Users were asked to explore the three environments using each of the three metaphors, and they were also asked to make a "movie", recording an exploratory sequence in each environment with each metaphor. The first toy scene consisted of three objects that were roughly similar in appearance to road signs, placed on a rectangular grid. There was an area of detail on one side of each sign. The second toy scene was a maze within a T-shaped hallway. Areas of detail were placed at different locations on the inside halls. The third scene was a single cube with details on three of its faces.

The experimenters elected to use a semi-structured interview technique to evaluate the experiences of the subjects. Instead of asking specific questions, a general dialog was carried on with the examinees. Only seven, nominally paid, subjects were used. Ware and Osborne state that they tried to select subjects "with a variety of previous experience[s]". They assert that they selected subjects with/without computer, mouse, movie-making, and aircraft or simulator flying experiences. It should be obvious that with four binary variables, the selection of only seven subjects would not enable a complete survey of all experience combinations, but the authors do not comment on this apparent problem in their study. Even more disconcerting is the fact that their results have dubious statistical significance, with the exception of a high correlation between responses to questions regarding ease of control, ease of making the movie, and ease of exploration for the each of the three metaphors in each of the three environments. Among the users, it was generally agreed that flight was the best for the maze and worst for the cube, and scene in hand was best for the cube and worst for the maze, so some polarization may be at play here. All subjects were able to adapt to the use of each metaphor within 20 minutes of training. Some attempt was made by the experimenters to randomize the presentation of the metaphors and environments through the course of the 3-hour test sessions, but here again, the small number of subjects prohibited full exploration of all permutations.

Allowing for the rather poor quality of the experimental design, some interesting observations were put forth:

With the eyeball in hand metaphor, users noticed limitations in fine control of hand positions, resulting in a jerky appearance for the movie sequence.

Contrived physical maneuvers had to be applied in order to manipulate the 3Space™ device to obtain the desired viewpoint. Some users experienced the virtual scene as literally being in front of the monitor, and would avoid trying to "touch" it or stand within it. Looking under or behind the scene required a reversal of the movement of the scene relative to the hand motions, some individuals found this to be disorienting. For users who imagined the controller to be a camera, some registered confusion over the actual scale of the scene and the perceived size of the virtual camera. The eyeball in hand metaphor was indicated as the one easiest to learn. Certain subjects felt that it was most appropriate for the maze and least appropriate for the cube, and others had the exact opposite opinion.

For the environment in hand metaphor, users experienced difficulty in manipulating the scene when the viewpoint was far from the center of rotation. Rotating through a large angle required a full twist of the 3Space™ device, or ratcheting, either of which were perceived as undesirable maneuvers. Subjects felt that this metaphor was best for hand-sized objects like the cube, and had difficulty when the believed size of the scene was large relative to the size of the hand. Users had difficulty performing simultaneous rotations and translations within this metaphor, and movie making was experienced as the most difficult.

With flying vehicle control, the maze was agreed to be the easiest scene to traverse. Continuous movement of the 3Space™ was not required. It appeared to be least suited for observing the cube, as users found it hard to fly around an object which was not in their line of sight, and some also felt that it was unnatural to fly around what they imagined to be a small object. Users felt uncomfortable flying through objects or making angular motions. Generally, though, this metaphor was perceived as "less restrictive" than the others. Movies produced through flying were smooth and had the best quality.

The order in which the metaphors were learned seemed to have some effect on the manner in which users would explore the scenes, as they would attempt to transfer some of the manipulation methods, with which they had prior success, to other metaphors that they were experiencing for the first time. Subjects with previous flight experience seemed to be least flexible in their use of the flying vehicle control metaphor, tending to only make the types of movements that would be possible in an actual aircraft.

It is difficult to draw any concrete conclusions from the results of this paper, owing largely to the vague manner in which the observations were reported by the authors. Citing some of the problems associated with rapid motions through large spaces using the aforementioned metaphors, researchers at Xerox Palo Alto have recently identified another form of movement. They allow the user to identify a Point of Interest on a target item in the scene, and then the viewpoint is altered logarithmically, in the animated sequence, according to distance.[Mack90] It can be seen that there is much validity in examining the effects of metaphors in producing viewpoints for manipulating virtual realities, and the concepts presented here will likely continue to provide excellent launching points for further study.

Feeling and Seeing

A number of virtual environments have included tactile feedback as part of the system. Michael McGreevy at NASA has provided this output in the form of an integral pressure differential mechanism built into the handset of a experimental unit intended to be used for remote construction and repair of vehicles, such as on the space station.[Mac90] The University of North Carolina currently maintains the most active research group dealing with the subject of human haptic response (sense of touch) in virtual environments. A graphics system that they developed allows users to construct custom drugs to kill dangerous cells, by complementing the molecular structure and electric charges of the target enzyme molecule's receptor cites. In working with this system, the designers realized that items could be more easily positioned if the user was provided with tactile information indicating the strength and direction of molecular attractions and repulsions. Force feedback on a pistol grip was determined to be most effective and enhanced the sense of realism in the simulation.[Stew91]

Minsky, et al at the University of North Carolina reported on a real-time Macintosh-based system called Sandpaper, that they devised in order to experiment with force display technology. Their intention was to simulate the physics of the user's virtual world and allow tactile exploration of a variety of textures. Control theory analysis was used to determine the stability of the simulation, taking into consideration the inherent properties of the human arm, so that the effects of changing the sampling period, mass, stiffness and viscosity variables could be reliably predicted.[Min90]

The experiment conducted was based on earlier psychophysical studies by Lederman and Taylor[Led72], and was adapted to the computer/ simulation environment. Lederman and Taylor's psychophysical research is of particular importance as it led to the development of significant theories regarding haptic perception. The notable result of the cited work was the discovery that groove width is the most important factor in the sensation of roughness, although contact force and the resultant skin deformation also play significant roles.[Boff86]

In the simulation, patches of texture were created and presented to users in combination with no visual feedback or with a graphic display that was either intended to match the simulated texture, or to differ from it. Subjects were asked to rank different textures according to their perceived roughness. Although only a small number of subjects were used, patches were ordered with a moderate degree of consistency. Users were also asked to adjust the force amplitude parameter for Perlin's noise and Pentland's fractal depth maps, and a groove spacing parameter in grooved patches, and were requested to identify the minimum and maximum values where each patch appeared to have a "surface texture".

Textures were created as a series of small bumps, ranging from a few microns to a few millimeters in virtual height and width. A 2-degree-of-freedom joystick allowed the user to explore the texture, and motors driven by the computer system provided sensory feedback regarding the surface under examination. The illusion of a bump was created by simulating spring forces using the motors. As the user attempted to move the joystick in a direction that would go "up" a bump, the motion was opposed by a spring force proportional to the height of the bump. In this manner, it would be difficult to move to the top of a bump (similar to the difficulty in pulling a spring outward), and easy to fall off of a bump to a lower region of the simulated surface (ease in letting a spring relax). This is shown in Figure 3. Textures are computed in real time from a stored texture depth map, or from a procedural representation of the texture. An analog-to-digital converter was used between the joystick sensors and the computer for input, and a digital-to-analog converter was used between the computer and the motors for output.

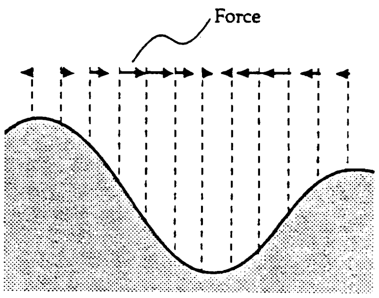


Figure 3

The essential formulas used in creating the surface system were as follows:

spring force = stiffness * position
damper force = viscosity * velocity
mass force = mass * acceleration

with force, position, velocity and acceleration all being vectors, and stiffness, viscosity and mass all being scalars. For the simulation, position, velocity and acceleration were measured using the joystick. Adjustable parameters (accessed by screen-based sliders) controlled the height and spacing of grooves or bumps and the viscosity of the surface, as well as the apparent mass of the joystick.

It was observed that certain body motions are associated with the feature exploration of objects. Lateral (rubbing) motions of the hand are typical for investigating texture. Patches needed to be of a certain virtual "size" before the users found the experience acceptable. The simulation was further enhanced, for about half of the users, by reworking the joystick so that it had the physical appearance of a ball being pushed across a black surface.

The model was analyzed thoroughly using control theory in order to determine the effect of the sampling rate on the stability of the system. Unstable systems result in deterioration of the tactile illusion. Quantization, thermal, transmission, and electrical noise all produce jitter problems in input, and this also has a destabilizing effect on the simulation. In particular, instability problems result from 1) sampling delay due to the use of a digital device, 2) noise and delay from the use of a velocity derivative for measuring the external force applied to the system, and 3) the variance in the locations of the input sensor and the output actuator (the mass is therefore a distributed model instead of a point).

In a continuous (analog) system, the simulation equation is:

mass*acceleration + viscosity*velocity + stiffness*position =
force generated by motor - force measured by sensor

The entire model is considered to be a feedback loop with the following structure:

input position -> spring constant -> delay -> output force
| |
<----------$1/(Ms^2 + Bs)$ <---------

where M = mass, B = viscosity, s = natural frequency of the system. If the product of the delay (T) and the natural frequency of the system is small, then the delay in the feedback loop can be approximated using a second-order Taylor series expansion. The transfer function can be derived as the ratio of output force/input position, using the feedback model and substitution of formulas. The authors then used the Nyquist stability criterion to determine where the system will be unstable. The poles occur where the denominator (input position) is zero. By this determination, instability occurs when (stiffness * delay) - viscosity > 0. A constant C, which was experimentally determined to be ~2, in this relation, yields the instability equation:

delay > 2 * viscosity / stiffness

The authors indicate that the addition of the human arm to this equation does not affect the stability formula. Note here that the delay value is the same as the sampling period used for the system.

Unfortunately, the system under investigation was discrete, not continuous, so numerical simulation was performed in order to obtain insights about the model. Three observations were made, as follows:

1. T* is linearly related to 1/spring constant.
2. T* is linearly related to viscosity over a wide range, and then becomes nonlinear.
3. T* is not related to spring mass when the mass is over a certain threshold.

T* = maximum sampling period that produces stability

These relationships were clarified through examples. If viscosity is small, and the spring constant is large, then instability can be easily produced, as the system delay may exceed 2*viscosity/spring constant. A slow system could therefore only simulate a restricted family of these types of force fields. On the other hand, if one considers the metaphor of "stirring a rod in a tank of viscous oil", the high viscosity will produce great stability, therefore the sampling rate can be low, and the simulation will still be effective. The hypotheses implied by these computations were confirmed through the use of simulations where sampling rates and viscosities were varied.

After assessing the results, the authors were puzzled as to why users observed differences in the feeling of the system when sampling time was increased from 500 to 1000Hz, considering the fact that the neural-muscular response time is ~200ms. They conjectured that perhaps the human system is not digital, rather a form of digitally supervised analog control, although anomalies in the construction of the joystick could not be excluded.

Further research is planned, including an enhancement of the simulation with auditory feedback and visual cues. The authors would like to improve the surface models by defining them in terms of degrees of roughness, softness and stickiness instead of the density and placement of bumps. They would also like to study the motions made by users performing surface exploration tasks. Additional avenues of investigation include the mapping of surfaces onto three-dimensional objects, the creation of soft surfaces and volume textures (fluids), and applications in virtual environment scenarios.

The use of a pre-existing study from the field of perception clearly enhanced the viability of this experiment. I was able to confirm the correctness of the formulas and derivations used, and should note that they are consistent with current practices in control theory. My general assessment of this paper is that the ideas presented are sound, and that this detailed style of work will go a long way in providing an understanding of the complex mechanisms in the human perceptual system, which can be applied toward the design of better user interfaces in virtual environments.

Computational Complexity

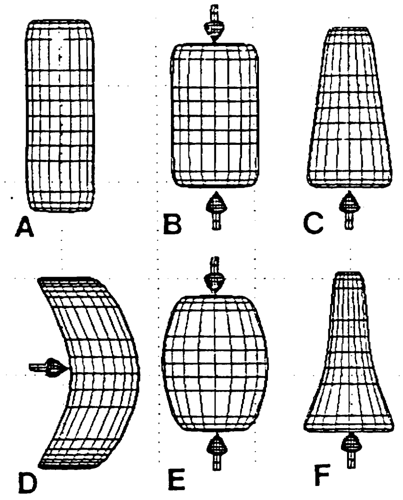
The subject of computational complexity has traditionally been a concern of computer scientists, and has its place in virtual environments research as well. If, for example, the New York City skyline can be modeled to within a meter using 270,000 polygons[Bran87], we need to assess the impact this will have on the difficulty of maneuvering objects in a similarly detailed construction. Alex Pentland examined the problems of rendering, dynamic simulation, collision detection and constraint satisfaction in this regard. He asserts that algorithms for these processes should be ones which scale linearly with increasing problem size. He first discusses the traditional approaches, and then proposes some solutions which may offer enhanced performance while still maintaining the necessary realism within the simulation.[Pent90]

The author developed a prototype virtual world, ThingWorld, which he used in investigating the scaling properties of the aforementioned algorithms. His hypothesis was that if small, controlled errors in dynamics calculations were permitted, if the geometry was represented using implicit functions, and if constraints were restricted to quadratic energy functions (springs) and holonomics (degrees of freedom), the savings in computational complexity could be sizable. He discusses the salient formulas used in his simulation, but admits to having omitted considerable details of analysis in this paper.

In the area of rendering, Z-buffer algorithms have been established which provide linear scaling of computational complexity with polygons and pixel displays. For those unfamiliar with this process, a buffer is used to maintain the z-coordinate (depth) of every visible pixel in the image space. If a new pixel is to be written to the screen, it is first compared to the Z-buffer information at that location to see if it is in front of or behind the pixel already being displayed. If it is behind it, it is discarded. Otherwise, the Z-buffer information is updated and the new pixel is written to the display. Essentially, this only requires a probe to retrieve the value of z(x, y). The computation time, therefore, is related to that used by polygon scan conversion.[Rog85] Pentland considers this matter to have been resolved, and therefore does not further elaborate on this topic in his paper.

With reference to the dynamics of moving objects and deformation, the author discusses the use of the finite element method. Here, an object is broken down into components by identifying a finite number of nodal points, instead of using a continuous displacement function. The energy equation for the finite elements is: $M\ddot{u} + D\dot{u} + Ku = f$, where u is a 3n x 1 vector of the displacements of the n node points, M, D and K are 3n x 3n matrices for the mass, damping, and material stiffness of each point, and f is a 3n x 1 vector of the forces acting on each node. (Note the similarity of this formula to the one used in the texture simulation.) Matrix multiplication is well known to have $O(n^3)$ calculation complexity and $O(n^2)$ storage locations are required. The rigid-body approach, though computationally simpler, is discarded as it is inadequate in predicting the effects of contact and friction, and non-rigid behavior must be modelled using waves and compressions when describing the resilience which occurs during collisions.

Pentland's approach to dynamics simultaneously diagonalizes M, D and K in order to characterize an object as a set of natural strain or vibration modes, each with a separate resonant frequency. The *whitening transform* (the solution to the eigenvalue problem: $\lambda\phi = M^{-1}K\phi$ where λ represents the eigenvalues and ϕ represents the eigenvectors of $M^{-1}K$) is then used to convert the energy equation (above) into 3n independent differential equations, describing the time course and deformation of each of i separate vibration modes. Linear superimposition of these modes will determine how an object responds to a given force. Non-linear materials can be modeled by summing the modes at each time step to form a stress state which can be input to a nonlinear material transformation function. Some examples are shown in Figure 4.



(a) A cylinder, (b) a linear deformation mode in response to compression, (c) a linear deformation mode in response to acceleration, (d) a quadratic mode in response to a bending force, (e) superposition of both linear and quadratic modes in response to compression, (f) superposition of both linear and quadratic modes in response to acceleration.

Figure 4

The modal method decouples the degrees of freedom in the system, but does not, by itself, reduce their number. Further analysis yielded the observation that object shape is not affected much by modes with high resonance frequencies because those have small amplitudes, dissipate quickly, and generally receive less energy. By eliminating all modes but linear and quadratic, an adequate simulation can be produced (see Figure 5). This has the effect of reducing the computation and storage requirements to linear scaling (O(n)) in the number of vertices. If low frequencies are used, further savings result through the use of a larger time step in the calculations. Some additional time may be gained by precomputing modes for similar shapes.

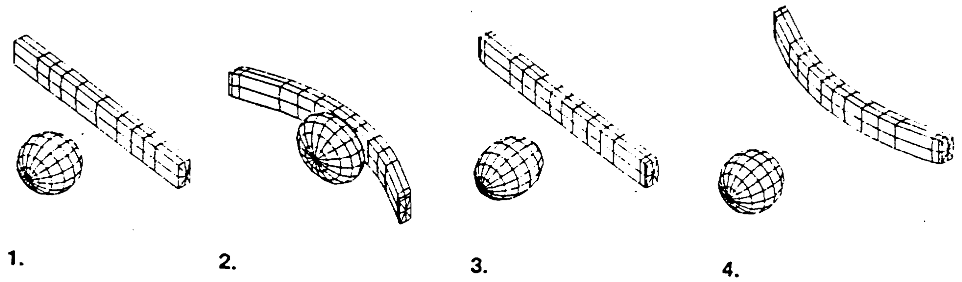


Figure 5

Collision detection using an octree scheme involves O(log m) operations and $O(m^2)$ storage locations, with 1/m being the allowed spatial resolution for detection of collisions. This method has the undesirable feature of requiring constant recomputation for moving or deforming objects. The bounding box method is more computationally expensive, but is only O(m) in storage. Pentland asserts that neither of these methods provides enough precision for physics simulation, and that in a polygon-based system, only a polygon-by-polygon comparison, costing O(Nn) where N is the number of points causing a collision hazard, and n is the number of nodes in the object's geometry, would be adequate. Collision detection for rough, irregular objects, or smooth ones that require a large number of node points to define them accurately, would therefore be costly.

If implicit inside-outside functions $f(x, y, z) <= d$ are used to define objects, rather than point-wise representations, collision detection can be reduced to O(N). For example, if a sphere is represented as:

$$f(x, y, z) = \mathrm{sqrt}(x^2 + y^2 + z^2) = d$$

and defined as all points (x, y, z) such that $f(x, y, z) <= 1.0$ then substitution with each of the N potential hazards in the equation and comparison to the threshold will determine which ones result in collisions. This imposes a limitation on the shapes to those that can be described in such a manner, but combinations of such volumetric primitives along with modal deformations can produce a reasonable range of objects. The combination technique requires an additional calculation of distance between the potential hazard point and the deformed object, possibly with some scaling, but this can be achieved in O(N) calculations. Objects need not be restricted to simple primitives: ThingWorld makes use of the family of superquadratics (a generalization of ellipsoids where the squared variables are replaced by arbitrary powers). The bump mapping technique for providing surface texture already employs inside-outside primitives, and could be incorporated naturally into a collision detection system of this variety.

Constraint systems (such as those used to control the range of motion of elements in a jointed figure) that are physically-based require $O(ck^2)$ complexity to compute the minimization of system energy (this being the number of operations needed to solve c linear equations in k variables), and $O(ck)$ storage complexity, where c is the number of constraints and k is the number of constrained parameters. Overconstrained and inconsistent systems can have as much as $O(ck^2 + k^3)$ computational complexity. Pentland asserts that general constraint satisfaction is isomorphic to general-purpose equation solving, thus no efficient method is evident. In ThingWorld, he elected to simplify the problem by keeping the system of equations linear, and by trying to make it well-conditioned and diagonal. The linearity requirement restricted the constraints to quadratic energy, for inexact models, or holonomic (which he defined as linear equations of the form $f(u) = 0$, where u is the vector describing the nodal displacements), when precise constraints must be enforced. These restrictions were not as limiting as they at first appeared to be. The addition of constraints can actually increase the speed of a simulation because degrees of freedom are reduced. The requirement for well-conditioning and diagonalization meant that natural parameters of the energy function had to be used. The variable x was defined as a natural parameter x of the energy function $E(x)$ if the derivative of $E(x)$ taken with respect to x is 1. The use of well-conditioned and diagonal functions reduces the cost of such things as matrix inversion and multiplication.

Pentland stresses that faster, larger and more parallel computers are not the solution to computational complexity problems within virtual environments. His work here suggests that research in efficient algorithm design, and a thorough understanding of the nature of the objects being modelled, can provide considerable reductions in computation time. For the most part, his research appears to be sound, indeed his comparison with existing techniques used the lowest complexity figures currently available (to the best of my knowledge). It is difficult, though, to assess the impact of some of his simplification methods, as an in-depth analysis over a wide range of object types, motions, and constraints was not included in the paper. The merit of his approach will only be revealed as other researchers attempt to implement his algorithms.

Concluding Remarks

It does appear to now be possible to produce viable applications which allow users to experience interactive, sensory contact in a limited context. It is my belief that currently the primary barrier to an effective implementation of this technology is the unwieldy nature of the input and output devices. Although small earpieces are available for audio output, a similar glasses-sized device is needed for video output. The heavy transmission or reception headpieces, and joint sensors used for motion detection are unacceptable. I have wondered if it would not be feasible to place a cloth body-suit and head cap, with an elaborate printed design, on the user, and employ pattern recognition techniques to determine position (via ceiling- and wall-mounted cameras). If a printed pattern is inadequate, perhaps a system containing embedded fiber-optics using variously colored lights, might provide the necessary resolution for motion detection, and still retain light weight and flexibility. Such a system would likely involve considerable computation power, but real-time outline recognition devices are available using custom hardware, as recently demonstrated at a Franklin Institute exhibition in Philadelphia by Myron Krueger (also depicted in [Stew91]), and this technology could be extended and refined. I feel that only when the speed and resolution of input and output devices are improved to the point that they compare with real-world expectations by users, and their size and weight reduced, will subjects find the experience to be natural.

Second in importance is the development of sophisticated tactile output devices that match the quality of auditory and visual presentations. I have heard reference to work by Minsky and Hillis in the development of a "skinlike material that can 'feel' and transmit small tactile surface features".[Hlab91] Constructions of this sort may be more appropriate than motor-driven hand-grips which are fatigue-inducing and provide an unnatural setting for haptic perception.

Third, experimentation with control and feedback units, and user metaphors, should enable virtual environments to become more closely matched to human perceptual processes. It has been observed by David Sparks and others that the superior colliculus in the mammalian brain contains individual neurons that respond to auditory, somatosensory and visual stimuli. It should be noted that these senses interact primarily to assist in identification of object placements in three-dimensional space.[Spa87] In my opinion, further investigation regarding the combination of sensory presentations is indicated. The senses of smell and taste have been mentioned [Suth65, Stew91] but more for comedic effect, than as serious research topics. Babies do place objects in their mouths, but it is thought that they do this to explore the shape of the object using tactile sensations in the mouth cavity, rather than to actually taste the objects themselves. I speculate whether the lack of investigation in virtual smell and taste is due to the fact that these senses are not used in the location of objects. Indeed, taste certainly does not involve this process, and smells are more often diffuse rather than directional. Some scientists believe that these senses are biologically more ancient, and convey strong emotional associations.[Pugh91] I feel that research involving these areas could be deferred until the more critical matters (listed above) have been resolved to a greater extent.

Finally, fine-tuning of algorithms will allow greater precision in simulations with lower computational costs. Certain fixed problems such as bounds on speed of video transmissions may become less of an issue as the level of sophistication of computer-generated graphic presentations increases.

The ultimate virtual environment system, therefore, should be one in which the sensory displays provide a total sense of realism, where presentation and response delays would be unobservable by the user, and where the tools used to manipulate the surroundings would be suitably matched to the task. In essence, the virtual setting would be indistinguishable from natural surroundings, if required to do so, and would also be able to provide an imaginary world that a user could accept as real. The extent to which all of this will be achieved, and its impact and applicability is, as yet, unknown.

Acknowledgements

I would like to take this opportunity to sincerely thank Dr. Norman Badler for providing both direction and guidance in the preparation of this report. I also wish to acknowledge Drs. Jonathan Smith and Xiaoping Yun for their time in forming the examination committee. My colleagues in the graphics lab are appreciated for their lively repartee and suggestions throughout this process. Thanks are due to many others, most notably Mae Churchill for timely forwarding of pertinent articles and unflagging confidence in my abilities, Mike Keith for statistics on data compression, and Patrick Mercuri for support and the extended loan of his Macintosh during text preparation. I also thank the Audio Engineering Society for having provided me with an educational grant toward my Ph.D. studies, the Drexel University library for access to their engineering collection, and (last but far from least) the Computer and Information Science Department of the University of Pennsylvania's Moore School of Engineering for research facilities and the opportunity to further my educational pursuits.

Figures

1. NOSC Teleoperator, from Uttal, W. R., *Teleoperators*, Scientific American, December 1989, page 125.

2. Three Metaphors, from Ware, C., Osborne, S., *Exploration and Virtual Camera Control in Virtual Three Dimensional Environments*, Proceedings of SIGGRAPH, 1990, pages 177 and 178.

3. Surface Gradient Technique, from Minsky, M., Ouh-young, M., Steele, O., Brooks, F. P., Behensky, M., *Feeling and Seeing: Issues in Force Display*, Proceedings of SIGGRAPH, 1990, page 236.

4. Object Deformation, from Pentland, A. P., *Computational Complexity Versus Simulated Environments*, Proceedings of SIGGRAPH, 1990, page 189.

5. Collision Deformation Modelled Using Only 1st and 2nd Order Modes, from Pentland, A. P., *Computational Complexity Versus Simulated Environments*, Proceedings of SIGGRAPH, 1990, page 190.

Bibliography

[Beg90] Begault, D. R., *Challenges to the Successful Implementation of 3-D Sound*, 89th Convention of the Audio Engineering Society, September, 1990, Preprint #2948, Session-Paper #C-II-1.

[Boff86] Boff, K. R., Kaufman, L., Thomas, J. P., editors, *Handbook of Perception and Human Performance*, Chapter on Tactual Perception by Loomis, J. M. and Lederman, S. J., Wiley, NY, 1986.

[Bran87] Brand, S., *The Media Lab*, Viking Penguin, New York, NY, 1987.

[Cham85] Chambers, A. B., Nagel, D. C., *Pilots of the Future: Human or Computer?*, Communications of the ACM, November 1985, Volume 28, Number 11, pp. 1187-1199.

[Emm91] Emmett, A., *In Search of the Miracle Hologram*, Computer Graphics World, February 1991, pp. 44-52.

[Fish86] Fisher, S. S., McGreevy, M., Humphries, J., Robinett, W., *Virtual Environment Display System*, ACM 1986 Workshop on Interactive Computer Graphics, October 1986.

[Gib84] Gibson, W., *Neuromancer*, Ace Books, New York, NY, 1984, page 51.

[Hac89] Hac, A., Mutlu, H. B., *Synchronous Optical Network and Broadband ISDN Protocols*, IEEE Computer, November 1989, pp. 26-34.

[Hlab91] Human Interface Technology Lab, *Telepresence*, Posting on sci.virtual-worlds Newsgroup, April 17, 1991.

[How82] Howard, I. P., *Human Visual Orientation*, Wiley, NY, 1982.

[Iso85] *3Space™ Isotrak™ User's Manual*, Polhemus Navigation Sciences Division, McDonnell Douglas Electronics Company, August 1985.

[Keit91] Keith, M. J., *Personal Communication with R. Mercuri*, Intel Corp., Princeton, NJ, February 18, 1991.

[Kle85] Kleinrock, L., *Distributed Systems*, Communications of the ACM, November 1985, Volume 28, Number 11, pp. 1200-1213.

[Led72] Lederman, S. J., and Taylor, M. M., *Fingertip Force, Surface Geometry, and the Perception of Roughness by Active Touch*, Perception and Psychophysics, Volume 12, Number 5, 1972, pp. 401-408.

[LeGa91] LeGall, D., *MPEG: A Video Compression Standard for Multimedia Applications*, Communications of the ACM, April 1991, Volume 34, Number 4, pp. 46-58.

[Mac90] MacNicol, G., *What's Wrong with Reality?*, Computer Graphics World, November 1990, pp. 102-108.

[Mack90] Mackinlay, J. D., Card, S. C., Robertson, G. G., *Rapid Controlled Movement Through a Virtual 3D Workspace*, Proceedings of SIGGRAPH, 1990, pp. 171-176.

[Mart91] Martin, D., *Virtual Reality! Hallucination! Age of Aquarius! Leary's Back!*, The New York Times, March 2, 1991, Section Y, page 11.

[Min90] Minsky, M., Ouh-young, M., Steele, O., Brooks, F. P., Behensky, M., *Feeling and Seeing: Issues in Force Display*, Proceedings of SIGGRAPH, 1990, pp. 235-243.

[Pent90] Pentland, A. P., *Computational Complexity Versus Simulated Environments*, Proceedings of SIGGRAPH, 1990, pp. 185-192.

[Phil90] Phillips, C. B., Zhao, J., Badler, N. I., *Interactive Real-time Articulated Figure Manipulation Using Multiple Kinematic Constraints*, Proceedings of SIGGRAPH, 1990, pp. 245-250.

[Pugh91] Pugh, E., *Personal Communication with R. Mercuri*, University of Pennsylvania, Department of Psychology, Philadelphia, PA, April 16, 1991.

[Rip89] Ripley, G. D., *DVI - A Digital Multimedia Technology*, Communications of the ACM, July 1989, Volume 32, Number 7, pp. 811-822.

[Rog85] Rogers, D. F., *Procedural Elements for Computer Graphics*, McGraw-Hill, NY, 1985.

[Spa87] Sparks, D. L., Nelson, J. S., *Sensory and Motor Maps in the Mammalian Superior Colliculus*, Trends in Neuroscience, Volume 10, Number 8, 1987, pp. 312-317.

[Stew91] Stewart, D., *Through the Looking Glass into an Artificial World - via Computer*, Smithsonian, January 1991, pp. 36-45.

[Suth65] Sutherland, I. E., *The Ultimate Display*, Proceedings IFIP Congress, 1965, pp. 506-508.

[Suth68] Sutherland, I. E., *A Head-Mounted Three Dimensional Display*, Proceedings of the Fall Joint Computer Conference, Volume 33, 1968, pp. 757-764.

[Tan88] Tanenbaum, *Computer Networks*, 2nd Ed., Prentice Hall, NJ, 1988.

[Tink89] Tinker, M., *DVI Parallel Image Compression*, Communications of the ACM, July 1989, Volume 32, Number 7, pp. 844-851.

[Utt89] Uttal, W. R., *Teleoperators*, Scientific American, December 1989, pp. 124-129.

[Wan90] Wang, J., Chi, V., Fuchs, H., *A Real-time Optical 3D Tracker For Head-Mounted Display Systems*, Proceedings of SIGGRAPH, 1990, pp. 205-215.

[Ware90] Ware, C., Osborne, S., *Exploration and Virtual Camera Control in Virtual Three Dimensional Environments*, Proceedings of SIGGRAPH, 1990, pp. 175-183.

[Wenz90] Wenzel, E. M., Stone, P. K., Fisher, S. S., Foster, S. H., *A System for Three-Dimensional Acoustic "Visualization" in a Virtual Environment Workstation*, Visualization '90 Conference Proceedings, October 1990, pp. 329-337.

[You89] Young, S. R., Hauptmann, A. G., Ward, W. H., Smith, E. T., Werner, P., *High Level Knowledge Sources in Usable Speech Recognition Systems*, Communications of the ACM, February 1989, Volume 32, Number 2, pp. 183-194.