Virtual Reality Technology: A Tutorial

by Frank Biocca, University of North Carolina, Chapel Hill

"Virtual environments . . . are communication media," says a NASA scientist, Stephen Ellis. "Like other media, they have both physical and abstract components" (Ellis, 1991, p. 321). This article is part of a series that considers "virtual reality"[1] (VR) as a communication medium (Biocca, 1992a, b, c; Lanier & Biocca, 1992; Meyer, Applewhite, & Biocca, in press). This tutorial surveys the "physical components" of the medium, that is, the technology.[2] I will consider what the scientists at the multinational firm Fujitsu call the device and system levels of the technology. I will also concentrate on the basic rationale behind each component of the multicomponent technology and point to trends in its development. Just as this article focuses on the technology, one of the companion articles considers the emerging techniques of communication design within virtual reality (Biocca, 1992c). Ellis (1991) calls these the "abstract components" of the communication medium.

Any discussion of this medium inevitably must address the issue of cognition. Virtual reality is defined by its effect on human perception and cognition. For some researchers active in the development of this medium, the ultimate goal of the technology is nothing short of the amplification of human perception and cognition (e.g., Furness, 1988; Krueger, 1991; Lanier & Biocca, 1992; Rheingold, 1991). While similar goals can be found in the development of other media (e.g., Biocca, 1987, 1988; Czitrom, 1982), no medium in history has been so self-consciously designed as an extension of our senses. According to Warren Robinette, a key designer at NASA and at the University of North Carolina, "The electronic expansion of human per-

[1] See footnote 1 in preceding article.

[2] Readers who want to pursue more detailed discussions of specific aspects of the technology should consult the works cited in each section and the valuable discussions of VR technology that have helped shape this tutorial (Benedikt, 1991; Durlach, 1992; Ellis, 1991; Ellis, Kaiser, & Grunwald, 1991; Krueger, 1991; Rheingold, 1991).

Frank Biocca is an associate professor and director of the Center for Research in Journalism and Mass Communication at the University of North Carolina, Chapel Hill. The writing of this article was partially funded by a grant from the University of North Carolina. The author is grateful to Gary Bishop, Nat Durlach, Douglas Holmgren, Warren Robinette, and Mike Shimamoto for helping him get access to unpublished materials. The author would also like to acknowledge the helpful comments of Gary Bishop, Glen Bleske, Ben Delaney, and Kenny Meyer on earlier versions of this article. The author must bear, of course, the burden of any errors or omissions that remain.

ception has, as its manifest destiny, to cover the entire human sensorium" (Robinette, 1991, p. 19).

The creation of VR technology is an interesting communication experiment. Some immediately seize and expand on the philosophical implications of the idea of a virtual reality (e.g., Benedikt, 1991; Heim, 1991). But even when our attention turns to a discussion of the "nuts and bolts" of each component, intriguing psychological and epistemological questions inevitably emerge (e.g., Loomis, 1992). As we will see, such questions strike at the very heart of this medium.

New communication technologies are either a small refinement on existing technologies or a paradigmatic shift in the way information is presented and absorbed. High-definition television (HDTV) is an example of a refinement. As with virtual reality, a desire for a "greater sense of presence" is one of the goals that drove its development (Benson & Fink, 1990, p. xvii). HDTV brings a high-resolution video image with a wider aspect ratio into the home. The result is an image similar to what we experience with 35mm film in theaters (Fink, 1989). But impressive as it is, HDTV is a significant refinement, not a radical shift in the way we communicate.[3]

Virtual reality technology is another matter. Virtual reality promises to be a multisensory merger of the telephone, the graphic workstation, and the television set (see Figure 1). Virtual environments may provide the platform for the general-purpose communication medium sought by members of MIT's media lab (Brand, 1988). This emerging communication medium has been heralded as the "ultimate form of the interaction between humans and machines" (Krueger, 1991, p. vii) and "the first medium that does not narrow the human spirit" (Jaron Lanier quoted in Rheingold, 1991, p. 156). Looking at the outlines of new virtual environments from inside their head-mounted displays, the developers see a transformation in the way we communicate and think.

But the promise is not yet proven. At the time of this writing, VR technology still has the look and feel of a prototype, a jumble of wires, liquid crystal displays (LCDs), and artful technical compromises (Brooks, 1988), just a portal looking out on a more mature technology to come. These improvised introductory systems remind us of the very early television set, a low-resolution array of black and white lines barely sketching a snow-spotted image. Just as with the early television sets, the image of a new, radically different communication technology is visible. Inside the low-resolution head-mounted displays we can barely see the outlines of a new way to produce and experience mediated information. When fully implemented and diffused, the medium could be the catalyst for a revolutionary change in the way we communicate. At least, this is the vision.

[3] If HDTV is combined with the information-carrying capacity and interactive capability of fiber optic cable, then there may be a significant increase in the range of information that can be seen on the HDTV set. But this possibility is not an inherent property of HDTV, nor would the communication experience be significantly different from what is now experienced during the use of personal computers and television.

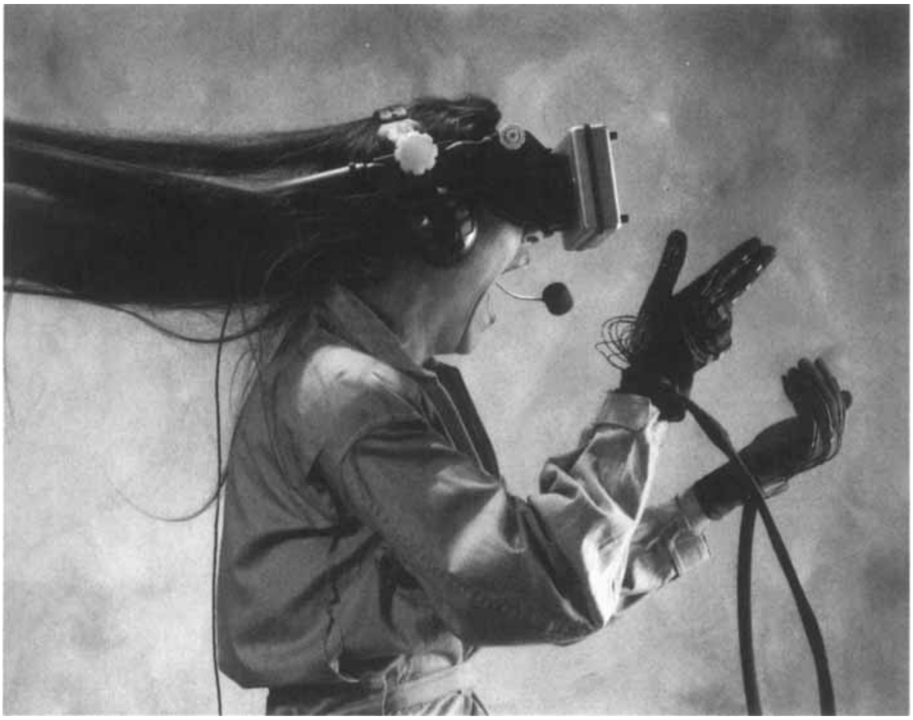

Figure 1. An example of a virtual reality system produced for NASA. Providing rich multisensory environments, this technology may emerge as the platform for a general-purpose communication medium. Visible components include a head-mounted display (output), data gloves (input), "Convolvotron" three-dimensional audio system (output, headphones only), and a magnetic position tracker (input, sensor only). (Courtesy of NASA Ames Research Center.)

The Logic of Virtual Reality

This medium takes many forms. Like the computer itself, it is a protean technology. There will be no single type of VR system and no paradigmatic virtual environment. We are more likely to see tailored combinations of components and applications, each capable of producing various types of experience.

Figure 2 presents an array of components characteristic of highly *immersive*[4] systems. These highly immersive systems tend to envelop the senses with virtual stimuli. The components are discussed below, but for now we should describe VR technology as a whole. Considered as a system, the technology can be said to possess a *developmental logic* that circumscribes the various versions of virtual reality. This logic is a set of goals for the

[4] *Immersive* is a term that refers to the degree to which a virtual environment submerges the perceptual system of the user in virtual stimuli. The more the system captivates the senses and blocks out stimuli from the physical world, the more the system is considered immersive.

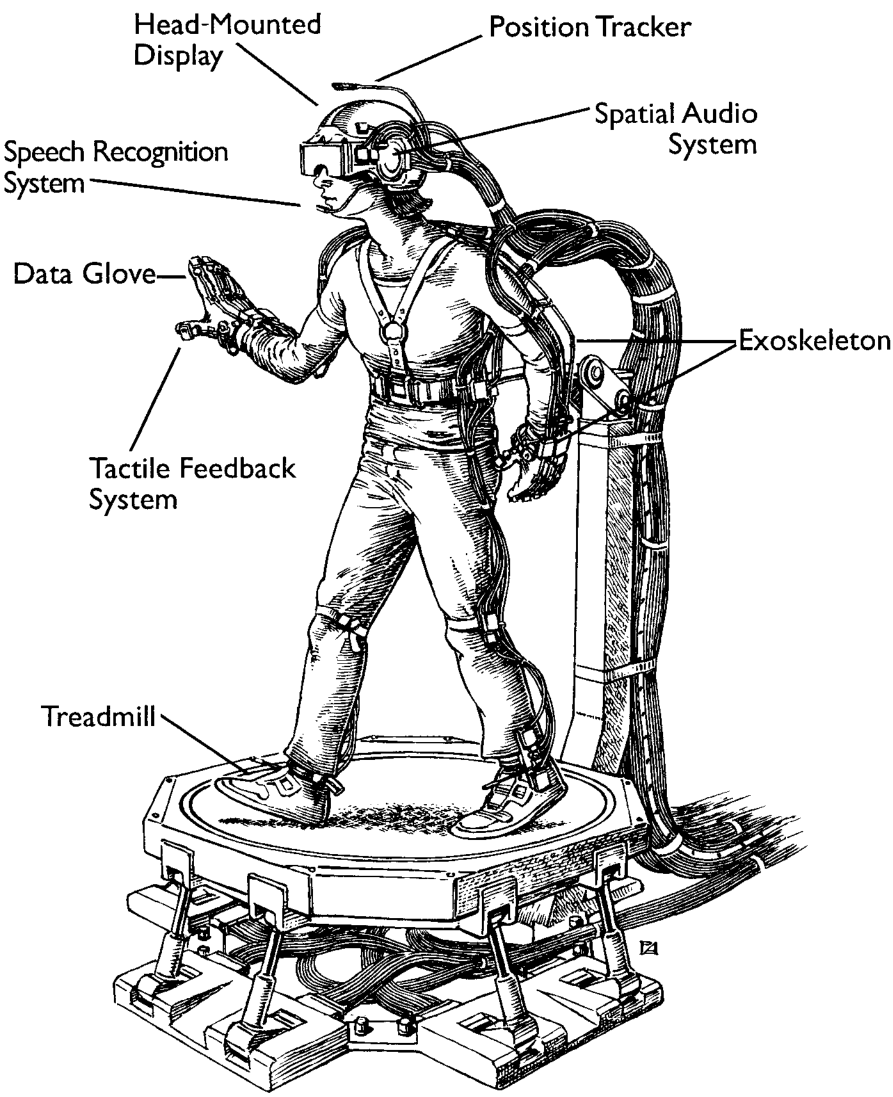

Figure 2. This artist's rendition of a highly immersive virtual reality system shows a full array of input and output components. Only a few systems come close to combining all of these components together (see, for example, Figure 8).

future of the technology. Looking at the limitations of video monitors back in 1965, an inventive computer scientist named Ivan Sutherland, an often-cited pioneer of VR technology, issued this challenge: "The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real" (Sutherland, 1965, p. 507).

His challenge has since become the research agenda for a rapidly growing community of researchers and industries. The long-term developmental goal of the technology is nothing short of an attempt to have our perceptual systems accept the reality of a computer-generated illusion, "to fool eye and mind into seeing . . . worlds that are not and never can be" (Brooks, 1988, p. 1). According to Nat Durlach of MIT, the development of the technology involves solving "questions about how you can map human perception onto virtual worlds" (quoted in Rheingold, 1991, p. 389).

The mapping of mediated information to the needs of our perceptual systems is psychologically complex. Ellis (1991) reminds us that the sense of physical reality is "a consequence of internal processing rather than being something that is developed only from the immediate sensory information we receive" (p. 323). Therefore, he says, the successful creation of perceptually engaging virtual environments "depends on the extent to which all of these constructive processes are triggered" (p. 323). The medium provides the cues to trigger the psychological constructive processes of the user (see Biocca, 1992c). An array of light on a visual display becomes a lush landscape in the mind of the viewer.

Creating a Sense of Presence

The strong perceptual illusion sought is often referred to as an engaging sense of *presence*. The word *presence* immediately suggests that the user will have sensations of being present in an environment, and will perceive objects as equally present. The process of creating a strong sense of presence begins by coupling the sensory organs of the user to the output devices of the computer. The output devices are orchestrated by one or more computers to generate a convincing simulation of the look, feel, and sound of another environment, a virtual reality. The eyes, ears, hands, and inner proprioceptive senses receive electromechanical stimuli that attempt to simulate a world pressing upon the senses. Some of the perceptual illusions used are as old as perspective drawing, some are as new as an electrotactile illusion of virtual sandpaper.

It is fitting that the term *presence* proposes not a goal but a destination, a psychological place, a virtual location. *Presence* has been enshrined as the name of a new journal devoted to virtual environments. While some skeptics doubt that the medium can achieve anything more than a schematized "virtual surreality" (Dennett, 1991, p. 7), it is likely that a strong illusory presence is only the last stop on a long trip to come. *Presence* is a construct, a variable with various levels and dimensions. Few are under the illusion that the technology can ever pass some Turing test of reality. But even the imperfect illusions of presence might appear so convincing in their perceptual realism that they may influence the reactions and behaviors of users (Shapiro & McDonald, 1992).

In an attempt to define the illusive destination of VR technology, MIT's Sheridan (1992) places the concept of presence on a continuum, a matrix defined by three axes: (a) the extent of sensory information, (b) the control of sensors, and (c) the ability to modify the environment (see Figure 3). On

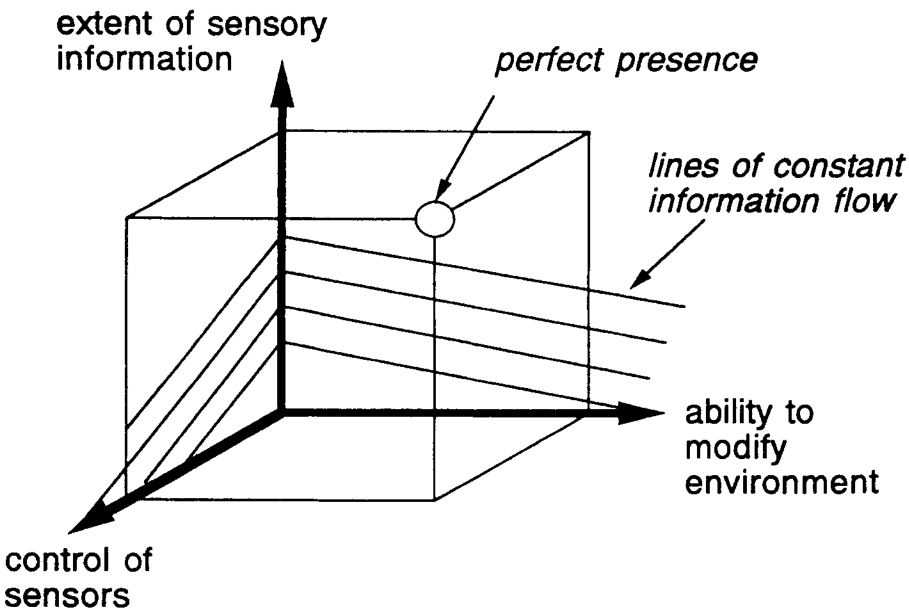

Figure 3. The principal components of the sense of presence (Sheridan, 1992).

a vertex of this matrix, Sutherland's vision becomes a point called *perfect presence,* a phrase that curiously suggests a religious state rather than a technical goal. It is perhaps telling of the state of the art that in Sheridan's diagram the arrows extend beyond the limits of the matrix, suggesting that the technical search for perfect presence will always be outside the grasp of the technology, though the sense of presence may steadily increase.

Another defining goal in the developmental logic of virtual environments is the attempt to turn the whole body of the user into an input device (Krueger, 1991). As Sheridan (1992) suggests, the user's control of the sensors and ability to modify the environment are essential to a strong sense of presence. Body movements, eye movements, facial expressions, and even unconscious physiological processes may be attached to the sensors of the computer.

In virtual environments the computer makes use of the natural way we interact with the physical world. Intuitive movements and actions become computer commands. All kinds of conscious and unconscious bodily movements are potential computer input. The intent of all this input is to sensitize the computer and the user, to turn every movement into a creative tool and a means of communication. Like a foot on an accelerator pedal, small physical movements can be linked to large movements or effects. The goal is a computer interface that is fully responsive to actions of the user.

The current VR systems already may have crossed a threshold. It is a psychological threshold, a point at which our perceptual systems are so immersed in the simulation that the user already begins to feel some of the sense of "being there,"[5] the early flushes of a powerful presence. Virtual reality technology can be defined as the sum of the hardware and software systems that seek to perfect an all-inclusive, immersive, sensory illusion of being present in another environment, another reality, a virtual reality.

Viewing older technologies through McLuhan's (1966) "rearview mirror," we see virtual environments as the latest development in the evolution of media technology and techniques toward greater perceptual verisimilitude. Past media have sought to refine the perceptual illusions within each sensory channel. For example, we can trace the evolution of illusions in visual media from perspective painting to photography, to dioramas, to moving pictures, to wide-screen cinema, to earlier experiments with "headsight television" (Comeau & Bryan, 1961) and "movie maps" (Lippman, 1980).

We can also discern the tendency to include more senses into media illusions. The coordination of aural and visual illusions in film and television is a good example. Predating the present technology are the pioneering multisensory systems of Morton Hellig, such as his famous '60s VR prototypes, "Sensorama" and the prescient "Telesphere Mask" (see Rheingold, 1991, pp. 49–67). Because media are often seen as tools for thinking, there is the underlying assumption that "we can build yet more powerful tools by using more senses" (Brooks, 1977). Even in 1965, Sutherland argued that the computer "should serve as many senses as possible" (p. 507).

Through McLuhan's rearview mirror, all past media can be characterized as our previous attempts to provide engaging simulations of not only perceptual sensations of an environment separated from us by space (e.g., the telephone) or time (e.g., the photograph), but simulations of another human's perceptions, thought processes, or fantasies. Virtual environments may become *the* communication crucible, a challenge for communicators and designers ready to employ the tools of this medium. We now turn to a discussion of those tools.

Virtual Reality Hardware

Unlike most developments of television hardware, changes in the hardware of virtual environments have direct bearing for communication, information design, and the social impact of the technology. Communication designers and researchers will not want to ignore the technical dimensions of this medium. Technical decisions are decisions about communication; they are decisions about how we communicate with the computer and with each other in this medium.

Virtual reality technology can be considered as an array of possible input and output devices, each device serving a sensory channel or linked to the user's body movements and responses (see Figure 4). The following sec-

[5] I am grateful to Byron Reeves's discussion of his concept of being there and the evolution of media technologies.

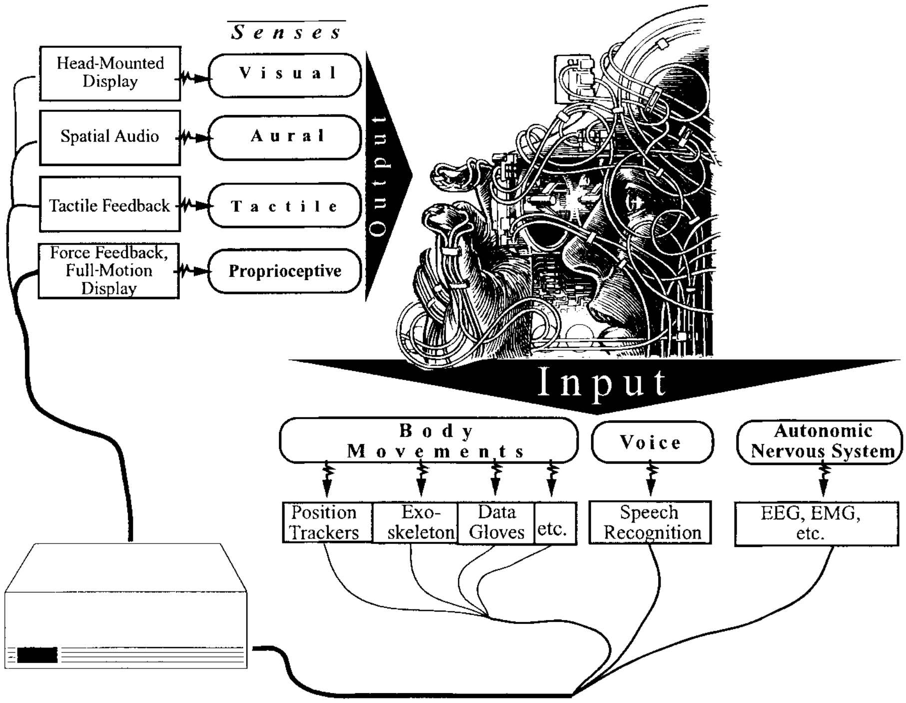

Figure 4. Virtual reality technology can be thought of as an array of possible input and output devices. Each output device serves a sensory channel. Each input device is linked to a user's body movements or responses.

tions discuss the perceptual and communication ramifications of technical options and trends.

Both input and output hardware are essential to an immersive illusion of presence, of being inside a simulated world. I begin with a discussion of the output devices because they more closely resemble other mass media technology and may be more instrumental to creating the illusion of "being there."

Output Devices

If the output devices try to couple the computer to the human senses, we should consider each device in relation to the characteristics of the senses (see Figure 4). In an attempt to realize the overall goal of presence, each device is being refined so that it comes closer and closer to optimally matching the parameters needed for powerful perceptual illusions in each sensory channel. In this continuous process of design the engineer looks for interface devices that are so transparent as to become "second nature" and in the long run invisible, what Winograd and Flores call "ready-to-hand" (1987, pp. 32–33).

We live in a visual culture, and our eyes are an important gateway for information. A great deal of our brain is dedicated to processing the information we absorb through our eyes. When we want information, we "*look* into it": We glance around us, read, gaze at pictures, watch television, view graphs, etc. It is not surprising that a significant part of virtual reality development has tried to create better illusions for our eyes.

Head-mounted Displays

The head-mounted display is the latest technical twist to a centuries-long dream of creating an image that is perceptually convincing (Gombrich, 1961). When, in 1965, Ivan Sutherland wanted to make the virtual world behind the video screen look, act, sound, and feel real, his first step was to work on the video image. His goal was no different from that of the 15th-century painter, Alberti, who thought he could simulate a scene so perfectly that his painting would appear to dissolve and become a window on another world (Alberti, 1966/1458). Alberti's and Sutherland's goals were to successfully simulate the visual array[6] (Gibson, 1966, 1979) a viewer would see when looking at a real scene. Alberti's result was a treatise on the technology of perspective painting. Sutherland's result was a design for a head-mounted display and a means of painting perspective illusions with stereoscopic images.

The head-mounted display builds on the ability of earlier visual media to give the viewer a slight sense of presence (see Ellis, Kaiser, & Grunwald, 1991). In pictures and paintings, the still image simulates the array of light in a "real" visual scene with such three-dimensional cues as linear perspective—for example, trees get smaller in the distance (Sedgwick, 1980). Other cues of depth in the virtual space of a picture come from overlap of pictorial objects (e.g., objects close to the viewer occlude objects farther away, as when a glass in a still life painting overlaps a bottle); changes in color and texture at points in the image that are farther away (i.e., aerial perspective, texture density gradients); as well as other two-dimensional depth cues (see Gibson, 1966, 1979; Hagen, 1980; Nagata, 1992).

Film and television add another level to our sense of presence in simulated visual space. Motion and motion-related depth cues, especially passive, motion parallax cues (Gibson, Olum, & Rosenblatt, 1955) give us a sense of the depth of a scene. When the camera moves, stationary objects

[6] The term *visual array* refers to the pattern of light rays that reach the eye when a viewer looks at a scene. Pictures and film record the pattern of light. Information contained in the recorded visual array is reconstituted as a perceptual experience when light is shown on a printed picture or passed through a celluloid film. A somewhat distorted simulation of the visual array of the original scene (Hochberg, 1986) is etched on the retina and is experienced as an information-rich representation of the original scene. For a history of head-mounted displays and their development, see Rheingold, 1991; Sutherland, 1968. For a review of the technology, see Aviles, Durlach, Held, Pang, and Spain, 1992; Robinette and Rolland, 1992.

slide across the screen at a speed directly proportional to their distance from the camera.[7] The perceptual system has evolved to use this cue along with motion cues to separate the distances of objects in the video scene (Hochberg, 1986).

But the video and film image fails to provide all the cues necessary for a truly convincing perceptual illusion. The perceptual system immediately detects the simulation, a perceptual impostor. For example, each eye sees the same scene, instead of the slightly different perspective on the scene projected on each eye when one views a physical scene—that is, "stereoscopic vision," binocular parallax. Second, the viewer can't really get a different perspective on the virtual video scene by moving. When we move our heads in the physical world, our viewpoint changes. The relation between the objects in a scene before us changes—that is, there are motion parallax cues derived from the motion and occlusion of the projected object surfaces on the retina (Gibson, 1979). But in film and video, these motion parallax cues are connected to the viewpoint of the camera rather than the viewpoint of the viewer. Only camera movements and cuts change the spatial characteristics of the image, and these are not in the control of the viewer. We are not inside the space of the video image, only the camera is. We are spectators, not actors.

That is why the many immersive virtual displays are head-mounted. A head-mounted display (HMD) provides greater verisimilitude to the image by adding two perceptual cues to the standard video image: (a) binocular (stereographic) disparity and (b) head-coupled motion parallax cues. With an HMD the viewer ceases to be a voyeur and comes closer to being an actor in the visual world.

HMDs use an old stereo imaging trick that goes back to the 19th-century Brewster stereoscope used to view stereo photographs. The head-mounted display puts a slightly different video image in front of each eye. Because our eyes are set apart (interpupilary distance), each sees a slightly different view of any scene (binocular disparity). Most HMDs simulate the different viewpoints of each eye by placing a slightly different view of the same image on each of two small cathode ray tubes (CRTs) or small LCDs in front of the eyes.[8] A series of lenses increases the size of the image to create a larger viewing angle and to fill more of the user's visual field (Robinette &

[7] If the concept of motion parallax is not clear, a classic simple demonstration will make the idea more understandable. Stretch your arm in front of you with your thumb up as if you were sighting something in the distance. Now move your head slightly from side to side. Notice how your thumb appears to sway back and forth more rapidly than objects slightly farther away, and these more rapidly than objects far on the horizon. As we evolved to walk, run, and hunt, our visual system evolved to make use of motion parallax cues to determine the distances between us and the objects around us and to generate three-dimensional models of the world around us inside our brains.

[8] An exception is the CAE system (CAE, 1986; see Figure 6).

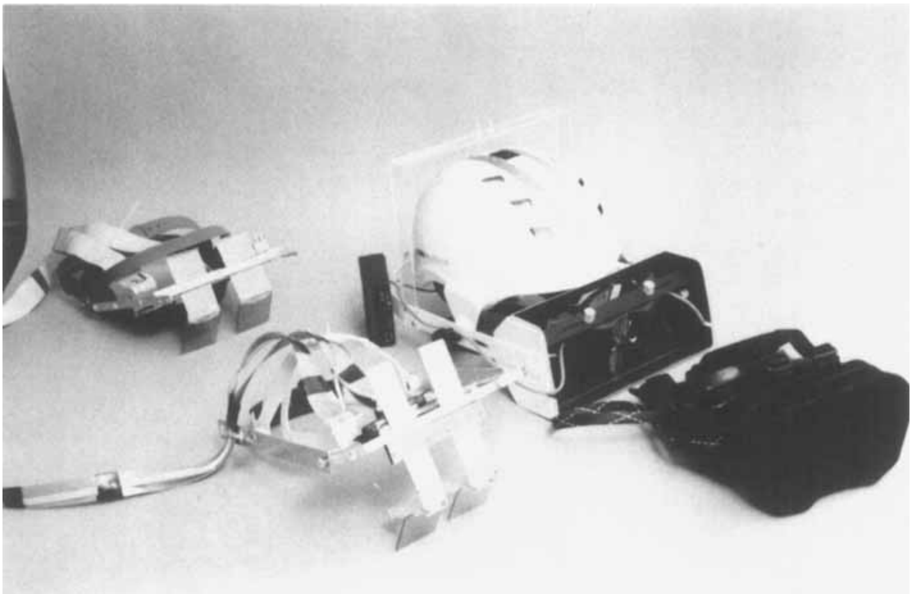

Figure 5. Examples of four different types of head-mounted displays that provide a three-dimensional stereoscopic image. The two on the left are see-through displays that allow a virtual image to be superimposed on a real-world scene. The two on the right (the helmet-mounted University of North Carolina system, and the VPL Eye-phone) provide a color, stereographic illusion of being immersed in a three-dimensional visual world that surrounds the user. (Courtesy of Computer Science Department, University of North Carolina.)

Rolland, 1992).[9] The tiny monitors and lenses may be mounted into head gear that looks like a scuba mask (VPL, 1989) while others are mounted onto a helmet (see Figures 1, 5, and 6) (CAE, 1986; Chung et al., 1989; Fisher, McGreevy, Humphries, & Robinette, 1986).

Graphic software calculates the geometrical difference corresponding to the views the user would receive in each eye if standing in front of the physical scene. The different images are displayed on the tiny screens in front of the eyes of the user wearing the HMD.[10] The computer graphic objects are suddenly three-dimensional. The computer graphic world appears to recede in all directions onto a virtual horizon.

Most head-mounted displays add the second critical visual cue to create the illusion of a physical scene: head-coupled motion parallax. When the

[9] To compensate for distortions caused by the lens, the image may sometimes be systematically distorted (anamorphic projection) to assure the right proportions when the magnifying lenses stretch the image.

[10] It should be noted that on most systems the interpupilary distance assumed by the optics and the image projection system do not match the user's (Robinette & Rolland, 1992). Also, the virtual binocular disparity may sometimes be exaggerated to produce desirable perceptual depth effects.

viewer moves his or her head, the stereographic image changes just as it would if the viewer were looking at a physical scene. This is accomplished by computer computations triggered by head movement information provided by the head-tracking input devices (see section titled "Tracking Body Orientation and Position in Space"). Sophisticated graphic software rapidly changes the images to mimic the spatial relations of objects in the physical world. The result is the illusion that one is looking at a stable, three-dimensional world. With a stereographic display and head-centered motion parallax cues, the user is an actor in the virtual world. The user is immersed in a visual environment and can look around and visually explore the computer graphic world that surrounds the user.

While head-mounted displays provide a strong sense of presence, the illusion is far from optimal (Robinette & Rolland, 1992). There are a number of visual perceptual cues and parameters (ocular convergence, accommodation, user-adjusted interpupilary distance, and visual lag)[11] that are not adequately simulated. While not all visual cues are equally important to the perception of depth and presence (Kilpatrick, 1976), more would probably have to be coordinated before head-mounted displays get closer to making a virtual scene look as perceptually real as a physical scene. We can observe various trends in attempts to overcome the limitations of the current head-mounted displays.

[11] If head-mounted displays are to be truly coupled with human vision, there are other design issues that must be perfected. Displays involving images suspended in front of the eyes can conflict with other cues used by our vision to construct a three-dimensional scene. For example, our eyes rotate in their sockets. The visual system uses information from the efferent muscle receptors of the eye as a spatial depth cue (binocular convergence). The eyes converge closer together when looking at close objects and diverge to point straight ahead when looking at objects on the horizon (infinity). The location of represented objects on the two screens in head-mounted displays need to be placed correctly so that the vergence of the eyes is either closer together or farther apart to match their represented location (close or far) in space. But this is rarely the case (Miyashita & Uchida, 1990; Robinette & Roland, 1992). Muscles, which focus the eye, also provide incorrect spatial cues because the eye is focused to clearly resolve the image on a screen only a few centimeters from the eye. If truly viewing a physical scene rather than a virtual scene, the eyes' accommodation would vary depending on the depth of the focused object. Other cues that need to be optimized to minimize subtle distortions in the stereoscopic image include software adjustments and HMD displays that account for interpupilary distance, correct geometry for the field of view, as well as other changes in HMD and computer optics (see Hochberg, 1986, for a discussion of the perception of visual diplays; and Robinette & Roland, 1992, for a discussion of these issues as they pertain to the design of HMDs).

There are other problems with current head-mounted displays that have their sources in other points in the system. There is a noticeable visual lag between user movements and the image. This may give the users a feeling that they are moving through a liquid-like environment. The images may also jump and jitter. Both problems, as well as some of the others mentioned above, may contribute to reports of simulation sickness (Biocca, 1992b).

Like the ubiquitous television screen, the present family of head-mounted displays are not yet optimized to fit the functioning of our visual perception. They come closer to making the image "look real" but still have some ways to go. But as with a television set, the imperfections in the displays may be acceptable for many uses while refinements in display technology bring visual displays closer and closer to Sutherland's long-dreamed "ultimate display."

Trend Toward Higher Resolution

Displays are composed of picture elements (pixels), little squares or dots that make up the image. The present LCD systems have very low resolution, especially when placed close to the eye (approximately 720 × 480 pixels).[12] CRTs can achieve higher resolutions (1280 × 1024 pixels) but it may be at a higher cost (Aviles et al., 1992). One CRT system used for military flight simulation and instrument display (VCASS) can achieve very high resolutions but only in monochrome. The multiple scanning guns and tight pixel arrays used in color CRTs make it physically difficult to produce very small, high-resolution monitors. LCDs can display color on very small screens but at lower resolutions. The width of the field of view (assuming a fixed pixel density) also affects resolution. One high-resolution system (CAE, 1986) gets around these problems by using a very expensive fiber-optic delivery system (see Figure 6).

The ultimate display ideally would match the resolving power of the eye. One can see a trend toward the development of higher resolution systems to achieve the goal of higher verisimilitude. It is unclear whether the drive toward higher resolution images will force the abandonment of LCDs or CRTs in favor of the alternative display techniques discussed below.

When we look at a physical scene, our eyes give us a field of view (FOV) that measures approximately 160° horizontally by 120° vertically (degrees of visual angle). Verisimilitude calls for displays that come closer to matching the width and height of the visual field. Previous media technologies like Cinerama and existing display systems such as those in IMAX theaters obtain powerful perceptual effects by filling much of the users' visual field with a vivid moving image. When the stimulus filling the visual field is in motion, viewers may have the illusion of self-motion, called "vection" (Dichgans & Brandt, 1978).

Our peripheral vision—that is, ambient as opposed to focal processing—is sensitive to certain types of information (Held, 1970). Peripheral vision is particularly tuned to detect motion, such as an animal running in the bushes or a waving hand in the corner of our eye. Peripheral vision is also sensitive to luminance changes, such as a blinking light or a mirror flashing in the distance.

While a wide viewing angle may seem well suited to producing a greater sense of presence, head-mounted displays vary greatly as to the amount of the visual field they cover (Aviles et al., 1992). Popular head-mounted displays such as the one manufactured by VPL cover a field of view of approximately 110° × 75°. Some expensive head-mounted displays report FOVs as high as 160° × 80° (e.g., CAE's Phase V HMD). When the visual angle presented by a head-mounted display is small, users may feel as if they are

[12] Reports of resolution on LCDs can sometimes be misleading. Picture elements (pixels) are sometimes reported as the total number of color elements (red, blue, yellow) as opposed to each picture element that the three colors comprise. In such cases, the resolution numbers have to be divided by three before they can be compared to other technologies like CRT (G. Bishop, personal communication, June 3, 1992).

Figure 6. The system supporting the CAE head-mounted display (CAE, 1986) begins with a pair of stereoscopic images from high-resolution video projectors. The images are condensed and "piped" through thick fiber-optic bundles (at back) where they are reflected to the eyes through a reflector and lens system. Though the result is a high-resolution virtual image, the current cost of this head-mounted display is very high (approximately $1,000,000). (Courtesy of CAE.)

looking at the world through a portal or scuba mask. This may lessen the sense of presence.

Why not just fill the visual field? Designers are faced with a trade-off between visual angle and resolution. Most systems have a limited pixel density. As lenses stretch this image to fill more of the visual field, the resolution of the image drops. So the trend toward greater visual angle tends to be limited by technical obstacles. Questions are asked as to which is more important under which circumstances—resolution or field of view (e.g., Aviles et al., 1992). It remains an open question, as it was for Cinerama, whether the greater field of view is worth the additional cost to obtain it.

Back in the '60s, Sutherland's original head-mounted display was so heavy it had to be suspended from the ceiling. The display was nicknamed the "Sword of Damocles." Like the legendary sword, the mechanical system attached to the display hung precariously overhead and might have killed the user if it fell. Users of the early head-mounted displays at NASA and the University of North Carolina reported fatigue in their necks and shoulder muscles from the weight of the display (Chung et al., 1989). VPL's popular head-mounted displays have gotten lighter since their introduction. Most displays have been made lighter by using lighter materials in the helmets and lenses that make up the system. Some of the electronic parts that were part of the early displays have been removed from the display and placed elsewhere in the system so as to lighten the load. Better balancing of the weight makes head-mounted displays feel lighter.

Extrapolating this trend into the future, one can foresee the development of head-mounted displays that may be as small and as light as a pair of glasses, allowing either a full virtual image or a superimposition of virtual images upon physical scenes (see Lanier & Biocca, 1992). But a number of design issues would have to be solved before that level of miniaturization could be reached. Some of the alternative display technologies might help accomplish this (see below).

Alternative Display Techniques

What if instead of assembling the image line-by-line on the picture surface, you were to assemble the image line-by-line on the retina itself? This is the basic principle behind one family of visual display technologies known as virtual image displays, retinal scanners, and direct retinal write (DRW) displays (Holmgren, 1992). The term "retinal write" can be misleading, because in the optical sense all images that we see are "written" on the retina. The systems differ from CRTs and LCDs in that the scanners do not use an intermediate display surface like a screen. The light that constructs the image is directed at the eye and quickly scanned to create a virtual image that is suspended in space. Anyone who has quickly twirled a light source like a sparkler to create a circle has in a somewhat similar way created the "virtual" image of circle from a single point of light.

A family of systems produced by Reflection Technologies (Wells, 1992), which has been used in military and industrial applications, constructs a virtual image line-by-line using mirrors (see Figure 7). In a system called Megapixel, the image begins as a single, magnified line of 1,120 tiny light-emitting diodes (LEDs). In continuous motion, a mirror swivels (50 Hz) to reflect the light pattern of the line of LEDs. As the LED line changes, the

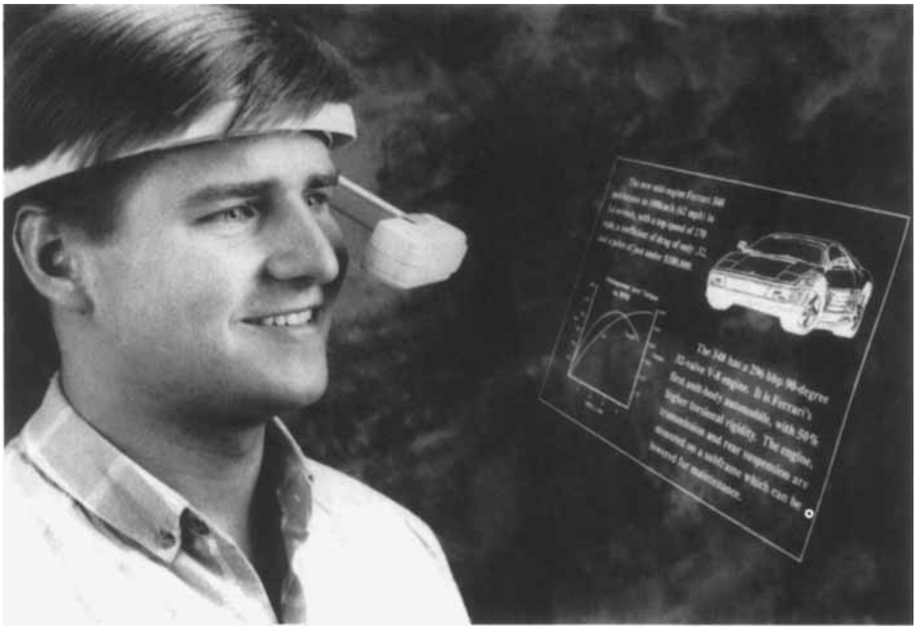

Figure 7. Example of a virtual image display. These displays assemble the image line-by-line on the retina. In this system image begins as a single column of 280 tiny light-emitting diodes (LEDs). In continuous motion, a mirror swivels horizontally (50 Hz) to reflect the light pattern of the LEDs. As the LED line changes, the mirror projects the new column. Doing so for each column, the mirror rapidly assembles the image without using an intermediate display surface like a screen. The image, which is assembled on the retina, appears to float in space about 2 feet in front of the user. (Courtesy of Reflection Technologies.)

mirror projects the new line lower down on the retina. Doing so for each line, the mirror rapidly assembles the image on the retina without the use of an intermediate display surface. The image appears to float in space about 2 feet in front of the user's eyes. The system's limitations include a narrow viewing angle (approximately 30°) and monochrome monocular displays. But resolution levels are good (1120 × 900 pixels) and the HMDs are light (2.25–10.8 ounces, or about 63.8–306.2 grams).

Lasers may also paint the image on the retina. A University of Washington research team headed by Tom Furness, a leading researcher in the development of air force applications of VR technology, is working on the development of a laser-based, retinal scanning system. Laser light can be narrowed into a very fine beam. Luminance levels can be well within safe ranges. Various ways of scanning the laser light—for example, a spinning multifaceted mirror—can be used to build the image on the retina, turning the retina into the ultimate screen. The first stage of a prototype using acousto-optic deflection of the laser beam has been developed at the Human Interface Technology Lab.

A number of potential properties recommend this technology. Laser retinal scanners offer the possibility of achieving resolution levels commensurate with the resolving power of the eye itself. The laser and mirror mechanism could be made very light and tiny; it can lessen the weight and bulk of head-mounted displays. Laser diodes are not very expensive. Unobtrusive scanners might be a means of developing virtual overlay system, a system that overlays virtual images on top of natural scenes.

Informed skeptics have doubts about this route to the ultimate visual display (Holmgren, 1992). Some doubt that the higher resolution levels can be achieved, or that lasers are the best way to proceed. Others have doubts about the various ways a laser beam could be deflected to produce high-resolution scan lines. There are also a number of technical problems in creating laser-based, color imaging systems.

Other alternative visual display technologies vying as possible roads to the ultimate visual display include high-speed monochrome CRTs with an electronic shutter that switches color filters rapidly to build a full-color high-resolution CRT image (manufactured by Textronics) and a mirror-reflection system that has 1 million tiny mirrors arrayed in a 1000 × 1000 matrix slightly bigger than a postage stamp (manufactured by Texas Instruments).

While all these options sound interesting, no one should underestimate the technical difficulties of achieving a lightweight, full-color, high-resolution, wide-angle viewing system. Many applications may not require this level of verisimilitude. Flat-panel or boom-mounted stereographic displays may be more than acceptable for many applications.[13] But if developed successfully, high-performance HMDs could go a long way toward achieving a high sense of "presence," of truly coupling the visual system to a computer graphic illusion.

Aural displays

Fully experiencing a virtual environment means hearing it as a space (Aviles et al., 1992; Wenzel, 1992). Hearing is three-dimensional; it is one of the ways you model the space around you (Blauert, 1983). Think of how we rely on our hearing to track the sound of a mosquito buzzing around our head. The aural realism of virtual spaces requires replicating the spatial characteristics of sounds like the changing intensity of a race car engine as it approaches a listener and screeches past; or the tapping of footsteps as they echo in a dark, empty corridor; or the chatter of a conversation off in the corner of a room.

[13] If one is willing to forgo the push toward a higher and higher sense of presence and settle for a modern-day version of Alberti's window, then a number of other displays may be considered. For some applications where optimal immersion either is not necessary or desired, we are likely to see the merger of other VR technology such as advanced input devices with various forms of stereographic monitors and display systems. For example, the University of North Carolina has a VR system that uses polarized glasses that display disparate images to the eyes to achieve the three-dimensional effect. One can see the coupling of elements of the VR technology with the host of stereographic display systems such as lenticular glasses, anagraphic systems, and boom-based stereographic displays (e.g., Fake Space Labs).

Our minds extract the spatial characteristics of sound from differences in sound patterns reaching our outer ears (pinnae). One spatial cue is the difference in the sound intensity from ear to ear and the slightly different delay between the time a sound reaches one ear and then the other (duplex theory). But a more discriminating set of cues comes from the way our outer ears generate different spectral distributions of frequencies when sounds differ in location, be it differences in azimuth (left–right, front–back) or elevation (up–down) (Shaw, 1974). If spatially distinct, audio "objects" are to be part of the illusion of presence in virtual reality, they should ideally (a) be stereophonically matched to the acoustic properties of our two ears, including the spectral mix of frequencies, delays, and distortions; (b) change in relation to the acoustics of a space, such as the size of the space, its shape, and the sound-absorption properties of its surfaces; and (c) smoothly change in relation to the virtual location and position of the user's ears in the space.

Like the changing geometry of VR graphics when the user is in motion, an aural virtual space has to mathematically model the changing properties of the acoustic space as the users walk about and swivel their heads to and from virtual voices, musical instruments, clanging metal objects, and roaring virtual engines.

An audio system developed for NASA (Foster, 1988; Wenzel, 1992) attempts to mathematically map the time and frequency properties of a sound source as its virtual position from the user's two ears varies. Sound sources are mathematically modified using a set of calculations called a "head-related transfer function" (HRTF) to make the sounds appear as if they come from a specific location relative to the user. The mathematical formulae at the basis of the virtual acoustic display are based on research using tiny microphones inside the ears of a sample of subjects. The research team mapped the changing acoustics as the ears were bombarded by various sound frequencies from 144 differently positioned speakers in a soundproof, dome-shaped, anechoic chamber (Wenzel, Wrightman, & Foster, 1988a, b). The results have been translated into a software and hardware package called the Convolvotron (Foster, 1988) that can give spatial properties to a limited number of sound sources in a virtual space. The system can also model the echoic properties of a small volume such as a room. Similar commercial systems exist for the Macintosh (Focal Point System), and other systems are under development.

The cheapest and most flexible way to couple acoustic systems to our sense of hearing remains the ubiquitous pair of headphones (circumaural earphones are visible in Figure 1). Headphones are head-mounted audio devices well suited to combination with the head-mounted visual displays. High-quality headphones used in VR help shut out sounds from the physical space and replace them with the high-resolution sounds from the virtual space. Insert earphones (ear plugs) also have desirable sound qualities and can be lighter and less noticeable (less skin pressure) than more bulky headphones.

The key difference between the aural experience on a standard set of headphones and the aural experience produced in a virtual system is the element of user motion. In a standard stereo recording, the properties of the audio space are determined at the time of the recording and mixing. The listener's position in the music hall (aural spatial position) is fixed at the time of recording by the sound engineers. But in a head-centered virtual audio space, the sound is dynamic and interactive; it changes as the user's head swivels toward the virtual sound source. The sound is engineered for the listener's position relative to the virtual sound sources at the moment of listening rather than fixed in space at the time of recording. Move your head closer to the virtual drum and the sound changes.

There is little doubt that audio imaging is an essential part of generating the full illusion of presence, of being there. Fully coordinated with other perceptual illusions, spatial audio can enhance the verisimilitude of visual and tactile illusions. Powerful audio imaging also can provide valuable spatial information about objects. Spatial audio can alert a user to objects and virtual beings that are visually occluded.

Audio illusions are quite advanced in their development because of a history of experimentation in audio engineering (see Blauert, 1983; Wenzel, 1992). But the optimization of the aural imaging systems must face a number of design challenges to optimize the verisimilitude of the illusion.

Each one of us hears slightly differently, and our individual differences can affect the success of the illusion. Each pair of human ears has slightly different acoustic properties, but an acoustic imaging system needs to be designed for an "average" set of ears.[14] In some cases, audio spatial illusions are subject to distortions such as perspective reversals—the user perceives that a sound source is behind him or her when the intention is to create the illusion of a sound source in front of the user. Users also may sometimes misperceive the elevation of the sound source. The use of headphones and anechoic sound models sometimes leads users to perceive the sound as localized "inside" their heads, rather than outside. But those perceptual distortions are rare and are more common when the user has no visual cue to match the aural location of a sound source. Distortions are also less likely if the user's head is in motion. Dynamic cues help resolve spatial ambiguities in both visual and aural perception as well as environmental audio cues such as echoes and reverberations in a simulated, asymmetric room (see Wenzel, 1992).

The calculations necessary to tailor various sound sources to the location of the user's head in space take time. Users report the perception of lag as the sound "catches up" to their current locations (Rheingold, 1991). Lags can be as long as 90 msec on current systems (Wenzel, 1992). Whether the

[14] It may be possible to calibrate systems in the future to adapt them to individual users. Our perceptual systems are also highly adaptable and this may help smooth over imperfections in the audio-imaging systems. It is possible that users will adapt and adjust for spatial mismatches in the audio system by adjusting the mental program that calculates the location of a sound source in space.

lag is perceptible varies from user to user. Higher head and body velocities (e.g., dancing) make the lag more perceptible. One can only imagine that the computational demands will continue to increase as VR systems must accommodate whole groups of moving, bobbing, and swiveling heads in multiuser spaces and as the number of sound sources that must be modeled in a space increases.

Tactile Output

A number of existing communication media provide stimuli for the visual and aural sensory channels. For example, you can experience the look and sound of a violin. But few media attempt to represent the feel of an object, such as feeling the weight of the violin, the resistance of the bow on the strings, or the vibration in the body of a violin when a note is played (see Rheingold, 1991). The presence of the object in one's hands is communicated to the user by the surface skin sensations and pressures such as the smoothness of the varnished surface of the violin. Signals from muscle proprioceptors communicate the pull of gravity when the violin is picked up and the resistance of pushing the violin against the chin is felt.

It is part of the ambition of the VR technologies to direct information to our tactile and proprioceptive sensors. A virtual environment that truly "feels real" should be able to simulate the sensation of surface textures like sandpaper or velvet, the resistance of surfaces like rocks or pillows, and the sensation of physical resistance, like moving an oar or stick through water, mud, oil, or glue. Systems that attempt to produce these illusions are sometimes called haptic interfaces and are of significant interest to VR researchers (Salisbury & Srinivasan, 1992). The term *haptic,* which contains the Greek root meaning "to fasten," suggests the interactive nature of the sensations. Part of the body is "fastened" to part of the world, and feedback results from the active exploration of a surface or object by the limbs, hands, and skin of the user. For example, we make full use of our haptic sensory channel when our hand searches for a set of keys in a bag or pocket.

Why the need for haptic and tactile imagery? In an age when most information is fed to us through visual and aural media, we forget the importance of touch in exploring our world. Hill's 1976 study (cited in Burdea, Zhuang, Roskos, Silver, & Langrana, 1992) found that haptic feedback reduced completion times by 50% on difficult manual tasks. Tactile imagery and force feedback are major sources of information about objects and the environment. To "put your hands" on something is to truly feel that the thing is present; it is so real "you can feel it." That kind of language suggests the essence of the concept of presence.

Tactile images can be used to communicate very subtle information in telepresence applications and can even provide for new insights in more abstract displays such as three-dimensional, tactile graphs. For example, a team at the University of North Carolina has demonstrated that the incorporation of haptic information in computer models of physical systems such as molecules can significantly enhance the problem-solving ability of chemists who use the system (Brooks, Ouh-Young, Batter, & Kilpatrick, 1990; Minsky, Ouh-Young, Steele, Brooks, & Behensky, 1990). If an engineer designing an automobile could not only draw the engine but also feel how the parts worked together, his or her insights during the design stage could be greatly enhanced. There is no doubt that tactile or haptic displays venture into new ground in the evolution of communication media and present significant new challenges.

If the goal is to create tactile illusions, then a good place to start is with that part of our body we use to explore the world of surfaces and textures: the skin and pressure-sensing organs on our fingertips. The fingertips are only a small percentage of our skin surface, but a significant part of our brain is dedicated to them. When users of a VR system bend down to pick up a computer graphic cube, the illusion (and information!) would be greatly enhanced if they could rub the smooth surface of the cube, or feel the edge of the cube pressing on their fingertips (Johnson, Cutt, & Ng, 1992).

Various shape-changing devices have been used to simulate textures and the slight pressures of surfaces against the skin. Electrotactile and vibrotactile devices (Kaczmarek, Webster, Bach-y-Rita, & Tompkins, 1991) attempt to produce surface illusions on a user's fingertips by stimulating skin (cutaneous) receptors with electrical currents or tiny vibrating mechanical prods.

The concept behind these devices is a simple one. If visual images on a video screen are composed of the varying illumination of a matrix of tiny dots, then tactile images might be created by a matrix of tiny pin-like rods ("piezoelectric vibrotactile actuators," TiNi, 1991), tiny balloons ("balloon-o-glove," Rheingold, 1991; "g-seat," Matthews & Martin, 1978), or other micromechanical systems (see Salisbury & Srinivasan, 1992). Each fingertip can be covered by one of these matrices. Simulations of edges and bumps or surface textures like sandpaper might be created by varying the height of these rods or balloons and vibrating them to maintain the illusion. Satisfactory tactile illusions of texture and "contact" might be generated when users bend down and pick up a computer graphic cube or pass their hand over the bumpy surface of a rock. A similar device, the "g-seat," even simulates the pressure of gravitational forces (g-forces) on a pilot's buttocks by inflating air cells in the pilot's seat (Mathews & Martin, 1978; Rolfe & Staples, 1986). Tactile displays could also be used to communicate (translate) visual or auditory information into tactile sensations for the blind or hearing impaired (Bach-y-Rita, 1982; Reed, Durlach, Delhorne, Rabinowitz, & Grant, 1989) or for enhancing "pressing" information (Bliss, King, Kotovsky, & Crane, 1963) in situations where the user's attention may be engaged or other sensory channels may be overloaded with information.

Force Feedback Devices

While tactile surface images are one way of simulating the look and feel of things, it is clear that to get to a point that satisfies Sutherland's quest for a world that "feels real" (Sutherland, 1968) requires more than tactile illusions on the surface of one's fingertips. When you reach out and touch the virtual world, the full illusion of presence requires a world that "pushes back at you."[15] What is needed is a full haptic illusion. When we slide an object like a cube across a surface or pick up a rock, part of the information of the physical sensation of the object comes not just from the surface of the skin but from various other sensors attached to our muscles and joints. These are part of the proprioceptive systems that tell us about the location and movement of our limbs in space.

Many force feedback devices use some mediating object to transmit the "feel" of hitting a hard surface, moving through a liquid, or swizzling a stick in a bucket of ice. Just as a blind man may use a cane to feel surfaces and objects he cannot see, a user may grasp a mediating object like a joystick (Minsky et al., 1990; Rheingold, 1991), a steering wheel, or a mechanical hand grip (Brooks et al., 1990) to "probe" the virtual world. Mathematical models run by the computer are used to apply forces to the joystick or another object to create haptic illusions described by such words as springiness, bumpiness, hardness, or viscosity. For example, the user might see a stick in a virtual bucket of ice, grab it, and feel the sensation of stirring the ice in the bucket, feeling the stick press and slide past virtual ice cubes. Illusions of the feel of a virtual world can be very startling and engaging. "Haptic illusions like the ones Margaret Minsky demonstrated," reported Howard Rheingold, "shook my reality sense far more than the cartoon like visual worlds I had explored . . . I'll always remember that as a particularly weird moment in my personal history of reality" (1991, p. 313).

Another way of obtaining haptic illusions is to encase the hand or, more radically, most of the body in an external brace-like device known as an exoskeleton (see Figure 8). As the name suggests, an exoskeleton is like an external skeleton with joints at the same locations as joints in the hands, arms, or other points of movement. When fully anchored and connected to hydraulic motors controlled by a computer, an exoskeleton can apply forces to the arm to simulate sensations like grasping and using a hammer, pushing a large rock across a table, or picking up a heavy object from the ground. The development of exoskeletons for force feedback has been fueled by telepresence applications. Examples are the remote manipulation of robot arms in dangerous environments such as nuclear plants or remote areas such as deep sea or space. Because an exoskeleton can also monitor the motion of the body, it can also be used as an input device (see section titled "Exoskeletons").

Some of the most sophisticated force feedback systems, such as the one developed at the Naval Ocean Systems Center (see Figure 8), incorporate the so-called Utah/MIT arm and hand—the Dextrous teleoperator system master (Jacobsen, Iversen, Knutti, Johnson, & Biggers, 1986). This device

[15] This is a kind of phrasing that Brenda Laurel has used to describe criteria for a satisfactory world (Laurel, 1991; see also Rheingold, 1991, p. 298). In her case, the meaning extends beyond satisfactory tactile illusions and metaphorically suggests the kind of interactive realism that ultimately will characterize the more engaging and satisfying virtual worlds.

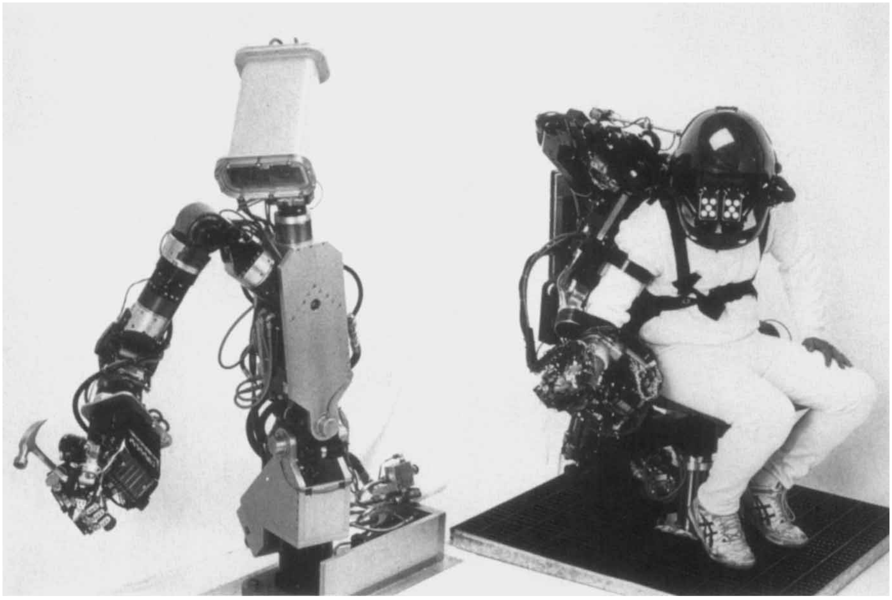

Figure 8. This force-reflecting exoskeleton is used to give the operator the sensation of being present in some distant location (telepresence). The unit on the right controls the actions of the robot on the left. The user sees the three-dimensional world of the robot and experiences some of the sensations of touching the objects the robot touches. This particular system is used to control and experience the actions of a robot in remote-sensing, underwater applications. A version of the unit on the right can also be used to see and touch purely virtual objects created by a computer. (Courtesy of Naval Ocean Systems Center.)

can register and apply forces to a number of the joints with up to 7 degrees of freedom and with high levels of sensitivity ("high bandwidth"). While this is a top-shelf system costing between $100,000 and $300,000, there are less expensive but lower resolution means of delivering force feedback to the hand (Burdea et al., 1992; Venkataraman & Iberall, 1990).

Experience with early exoskeletons—for instance, GE's Handyman—showed there was potential danger if they transmitted the full force of the push of some virtual object. In most force feedback systems (e.g., University of North Carolina's GROPE system) the transmitted forces are attenuated so that they do not accidentally injure the user by overly straining or bending a limb. Force feedback systems have been used to create the delicate feel of virtual musical instruments (Cadoz, Florens, & Luciani, 1984; Cadoz & Ramstein, 1988; Rheingold, 1991), the electric field of surrounding atoms (Brooks et al., 1990), or the resistance of moving massive objects.

Displaying Whole Body Movements

Some virtual environments attempt to simulate the feeling of the user's body moving through a large space. This can be the illusion of actively walking around a space or being passively transported in a vehicle such as car or

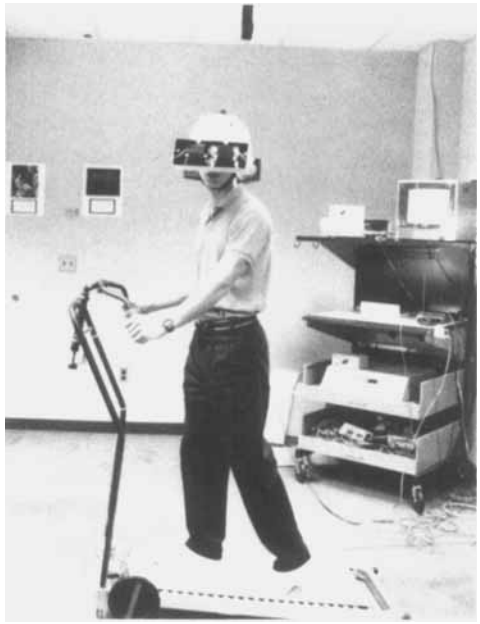

Figure 9. In this architectural "walk-through" application, the user walks on a treadmill. This creates the illusion of strolling through the virtual building seen on the head-mounted display. The treadmill acts as an input device and records the direction (using handle bars) and extent of the user's movement. Distance walked in the virtual world is matched to distance walked on the treadmill. (Courtesy of Computer Science Department, University of North Carolina at Chapel Hill.)

plane. But most virtual equipment is confined to a small space like a room or to the even smaller space covered by the position tracker. It is usually not possible for the individual to actually walk around the physical space to simulate walking around the virtual space. Treadmills (Figure 9) are sometimes the solution (see also DiZio & Lackner, 1986; Lackner & DiZio, 1992).

Some simulations require that the user have the illusion of significant motion through the virtual space—illusions like acceleration, deceleration, or the rocking motion of a ship, etc. Morton Hellig's now famous '60s arcade ride, Sensorama (Rheingold, 1991), included a vibrating seat and handlebars to simulate the "feel" of a motorcycle. The most widely experienced examples of such illusions are the well-known fantasy rides at Disney World and Universal Studios. These are large-scale simulations that require that the user have the vivid illusion of acceleration, deceleration, and other motion and inertial forces.

Some of the most advanced simulations of strong gravitational and inertial forces are found in flight simulators, a form of virtual environment (Flexman & Stark, 1987; Haber, 1986; Rolfe & Staples, 1986). These strong, visceral physical illusions are produced by *motion platforms* (see Figure 10). Users are usually seated, though they may, in some cases, lie or stand on the plat-

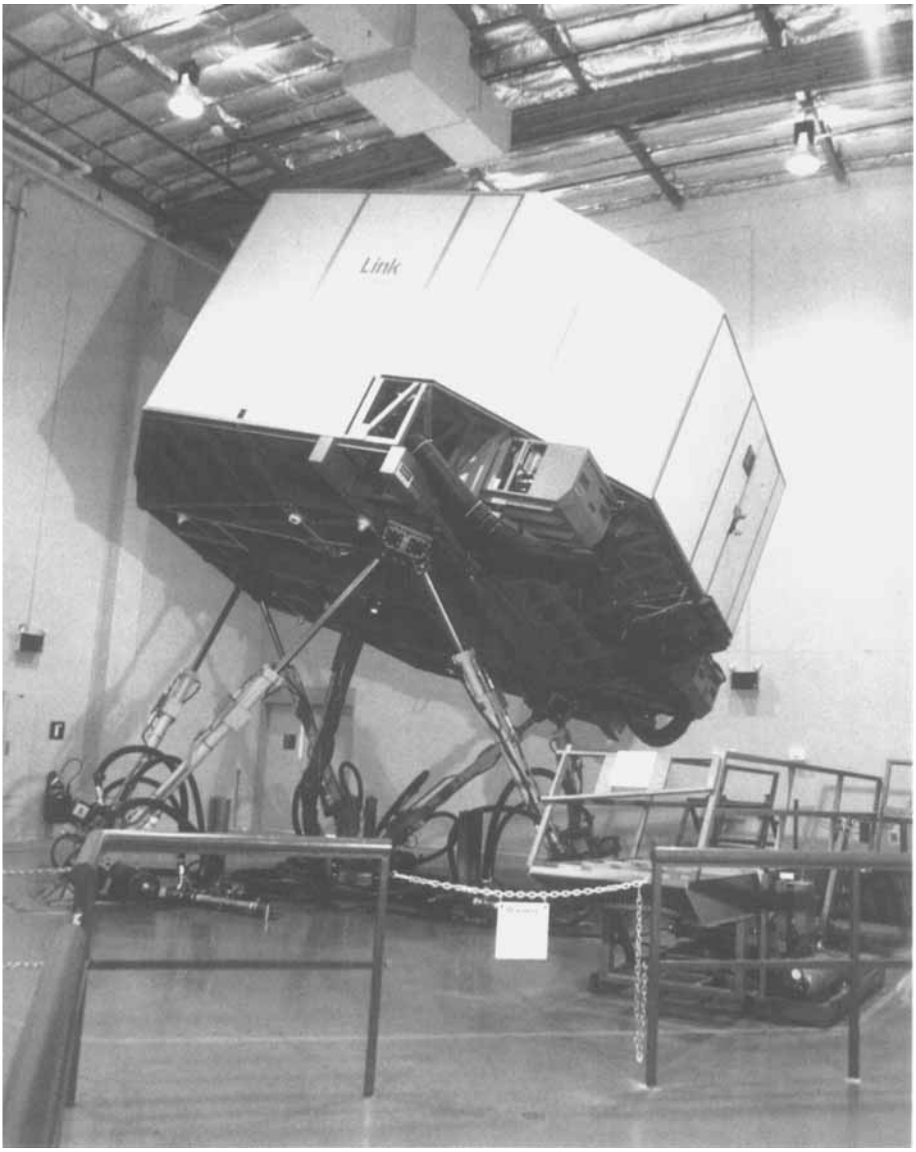

Figure 10. Motion platforms are used to simulate gravitational and inertial forces. They can give the user strong, visceral illusions of motion. This very large motion platform is used in simulation studies at NASA. (Courtesy of NASA Ames Research Center.)

form. Carefully calibrated movements, vibrations, and jolts are applied to the platform by hydraulic or pneumatic cylinders as well as motor-driven mechanical devices. By carefully applying the inertial forces in conjunction with visual and audio illusions, the user may have the illusion of having moved far greater distances than the actual short movements of the platform. There are many applications in the areas of entertainment and training where the forces of motion may be major components of the sense of presence.

The use of such platforms for training raises some difficult questions regarding the relationship of the forces exerted in the simulation and the actual physical forces modeled (Flexman & Stark, 1987; Lackner & DiZio, 1992). For example, the sensations generated by gravitational and inertial forces are used by pilots to make judgments about their position in space or the correctness of a maneuver (e.g., "flying by the seat of your pants"). There is a question as to whether the user's physical learning, which involves the subtle calibration of muscles, will transfer positively to the actual world where the environmental forces are different.

The motion platforms used in the military and major amusement parks are large, expensive, and relatively inflexible, that is, they tend to be dedicated to a limited class of illusions. One can foresee a cheaper class of small, single-user motion platforms intended for a wide range of VR entertainment applications (Rheingold, 1991). In the development of motion platforms we can already see two very distinct trends. The most common will probably continue to be the development of motion platforms in entertainment. As with the 25-cent pony ride outside grocery stores, the verisimilitude of the proprioceptive illusion may have a very low criterion for acceptability. Higher end developments are likely to be restricted to those rarer training applications where an accurate proprioceptive modeling of the physical forces is key to preparing the mind and body for the actual experience of those forces (e.g., flight and battle training as well as remote exploration and other telepresence applications).

Olfactory Displays

Compared to the other sensory channels, the olfactory and oral[16] senses are rarely engaged or participating in virtual environments. There are, obviously, clear reasons for this. Both senses are chemical interfaces with the physical world. Chemicals must be applied to the senses to obtain the sensation. It is easier to have computer control over light, sound, and motion than over odors and flavors.

But olfactory sensory illusions have been part of past simulations and media. A precursor to present VR systems, Morton Hellig's '60s Sensorama, simulated the smell of exhaust, pizza, and flowers by releasing puffs of artificial aromas into the nasal passages of the user at the appropriate points in the ride. Some of the amusement simulations found at Disney's Epcot Center use aromas to increase the illusion of displays such as an orange grove, a smoke-filled prehistoric scene, and a humid underwater simulation.

It is clear that stimulation of the olfactory sense erases somewhat the dis-

[16] In the case of oral displays, it could be argued that junk food is already a "simulation" of real food; nothing could be more artificial. The author is not aware of any stimulation of the oral sense that does not involve the ingestion of some form of food.

tinction between real and virtual stimuli. Real physical scents must be used to simulate the smell of the real environment. Sometimes those chemicals are derivatives of the actual object being simulated (e.g., perfumes for flowers, orange distillations for the smell of oranges, etc.). Illusions for these senses are still clumsy to control and deliver. For example, once released into the air, aromas linger in the air and in nasal passages, so it is hard to quickly change from one olfactory illusion to another.

While the addition of this sensory input can add to the sense of presence in certain environments, there seems to be limited interest in pursuing the development of VR illusions in this sensory channel.[17] While arguing for the inclusion of more senses, Ivan Sutherland dismissed the olfactory and oral senses: "So far as I know, no one seriously proposes computer displays of smell, or taste" (1965, p. 507). This was the case in 1965, and appears to be the case today. But the history of perfumes suggests that as a species we have long been interested in simulating pleasant aromas (Ackerman, 1990). Sensations of smell may play an important role in developing a feeling of presence in certain simulated environments (i.e., a rain forest) . While olfactory displays may increase presence or provide some hedonistic value, the lack of interest in this perceptual channel may be due to the fact that the sense is rarely used by humans to encode the higher level symbolic information often found in the visual, aural, and tactile sensory channels.

Inputting the Body

To create a powerful virtual illusion, the computer must be able to sense the location and actions of the user's body in space. This is necessary to (a) accurately represent the user's body in the virtual space and (b) turn specific body movements and actions into commands for the computer.

The input devices of highly immersive virtual environments try to conform to the way we interact with the physical world by making use of things such as the movement of our limbs, head, eyes, and other motions in physical space. The difference is best illustrated by an example. Say you want to move a computer graphic representation of a cube. In a nongraphic system you might type: Move cube, Location x = 10, y = 55, z = 42. In virtual reality you simply bend down and pick up the cube with your hand and place it on a computer graphic table. The floor, the cube, the table, and the graphic representation of your hand are all data entities in a program, as is the computer's representation of your movement. To you it appears as a naturalistic perceptual event.

Virtual reality technology tries to make greater use of the natural skills we have acquired from our interaction with the physical world, evolutionary skills that are coded in the operation of our perceptual and motor systems.

[17] A meeting of VR experts convened by the National Science Foundation put little emphasis on developments in this channel. A major VR training proposal (Durlach, 1992) covering various forms of input and output makes no mention of olfactory displays.

It is clear that the wider range of input devices used in VR systems changes what we think of as computer input, and especially the associated concept of "data entry." When the computer monitors the movement of the user's head, hand, and body, the user is inputting information into the computer and, in a general way, "entering data." But this is, of course, unconscious, unwilled, and, in most cases, passive. The computer responds both to these passive data entry methods and to active entry methods such as the use of a coded hand gesture. For example, in some VR programs a set of pointing gestures is used to signal to the computer the user's desire to fly in a specific direction.[18]

But it is still easy to overwhelm the computer's current capacity to process information. As Krueger (1991) points out:

> *The design of a perceptual system (the sensors that comprise the input for the computer) requires a trade-off between the need for the computer to know as much as possible about a participant's behavior and its commitment to respond in real time.* (p. 103)

The intuitive simplicity of input devices that conform to everyday actions has clear and revolutionary implications for the training, communication, and data-manipulation applications that will be discussed later.

Let us consider some of the input devices sometimes found within VR systems. In interpersonal communication, we learn a lot about the intentions and motivations of others by observing how they move their bodies: how close they stand to us (proxemics), how they walk (gait analysis), how they move their hands (kinesics), the tiny movements of their facial muscles, and where they look (eye movements). If this is valuable information for communication with humans, it may be valuable information for communication with computers. Kinematic input devices turn the movement of some body part into computer input. There is a large family of such devices. Each concentrates on a limited class of movements or uses different means to capture and digitize the movement of a body part.

Tracking Body Orientation and Position in Space

When we move in the physical world, the perspective relations of a scene or a room change. As we move closer to a cup, for example, its image becomes larger on our retina. Also, as we turn our head, the sound of an object like a horn will change as each ear either faces or turns away from the sound. To create a convincing virtual space, the computer must keep track of where the user is looking and where the user is positioned so that the appearance of visual scenes is changed appropriately as the user moves.

[18] This gesture language was first developed by Warren Robinette of the University of North Carolina while working on NASA's VIEW project (Rheingold, 1991). Some form of this gestural code has become a part of many glove-based VR programs, including those used by VPL and labs at the University of Washington and the University of North Carolina.

Position trackers are often used for this purpose (Meyer, Applewhite, & Biocca, 1992). Position trackers may use mechanical, optical, magnetic, or acoustic means to keep track of the physical location of a user's head, hands, or other parts of the body. One group of systems works by calculating the orientation and distance between emitters and the sensors. Depending on the type of system, the emitters or sensors may be worn by the user. In one popular magnetic tracking system (Polhemus trackers) the sensors are worn by the user and are quite small, about 1 cubic inch (16.4 cubic centimeters). Position trackers can report position along three dimensions of some small space, such as a room, as well as the orientation of the device in space along three axes. In this way, the computer can keep track of the location and movement of key parts of the user's body.

Keeping track of the location and orientation of the head is particularly important. The computer must be aware of what direction the user is looking to calculate the perspective relations of a scene that will be displayed to the user on the head-mounted display (Liang, Shaw, & Green, 1991). The computer must continuously sense movement and change the visual scene to match the new position and orientation of the user's point of view.

Current position trackers can foster some limitations in virtual environment systems. Most trackers can only keep track of a user's movements in a relatively small space, and there are problems of accuracy, report lag, and other sources of distortion.

We can evaluate trends in the development of this technology using the evaluative scheme proposed by Meyer, Applewhite, and Biocca (1992).

Resolution and accuracy. Current systems have adequate levels of resolution (the smallest change a system can detect) and accuracy (the range within which the reported position is correct). Good resolution and accuracy are desirable to detect such things as small movements of the head. Even small head movements can lead to significant changes in the visual scene. As VR systems must integrate real and virtual objects in a single illusion, the demands for higher resolution and accuracy are likely to increase.

Responsiveness. Current systems are often plagued by lags, the time between the moment a user makes a movement and the moment the computer responds to the movement. Part of this lag is attributable to the slowness of the position tracker. As lag increases, the sense of presence decreases to the point where it simply breaks down (Held & Durlach, 1991; Held, Efstathiou, & Greene, 1966). We can already see a tremendous emphasis on increasing the responsiveness of not only position trackers, but VR systems as a whole.

Registration. Does the position reported by the position tracker really match the position of the object in the physical space? Misregistration of your body parts—such as your hand, for example—can contribute to sensory conflicts and even simulation sickness (Biocca, 1992b) when the virtual visual location of your body parts do not match their "felt" position (proprioceptive location). Like resolution and accuracy, good registration will become essential to the interaction of virtual and real objects in a single illusion.

Sociability. How many emitters (i.e., people or body parts) can a system keep track of? If VR systems are to become interpersonal communication environments where a number of people communicate in a task or game, then position trackers must be able to keep track of all the individuals and their movements. Some systems, like people, cannot "see" (detect) an object or person when they are occluded. But some communication applications, like virtual theaters (Lanier & Biocca, 1992) will require that the system keep track of multiple users engaged in rapid action. A "sociable" system also needs to have a large range of operation so that it can track positions in a large space. The systems most commonly used today have a limited range.

Robustness. Like snow on your TV set, position trackers are subject to interference from other environmental sources. Position trackers are subject to distortion from sources of energy similar to the sources used for position tracking by that particular system, be it magnetic, acoustic, or optically based.

While positioners keep track of the general position and orientation of body parts such as the user's head and hands, more information is needed by the computer. It is also valuable for the computer to know when the users flex their arms and fingers to pick up objects, or flex their legs to walk. Objects in the virtual world need to move when pushed by virtual hands, grasped by virtual fingers, or kicked by virtual legs. Users may also want to communicate with the computer through gestures. And, finally, in telepresence applications the computer might use the digital record of a person's movements to guide the operations of a distant robot who then replicates the gestures in places such as deep under the sea, in a nuclear power plant core, or in space.

Exoskeletons. One way a computer can keep track of our body movements is to have us wear something that can detect the flexing of our limbs when we walk or pick up objects. An exoskeleton is such a device. Earlier I introduced the use of exoskeletons as output devices (see Figure 9). But an exoskeleton can simultaneously be an input device (Little, 1988).[19] Strapped to the user's limbs, the exoskeleton's joints move when the user's joints move. By digitizing this information, the computer can know when we flex our fingers, arms, and legs (see Figure 11).

While exoskeletons can be bulky, they can be used to both sense the movement of a user's limbs (input) and provide force feedback during movement (output) to help create illusions such as the resistance of hard surfaces like walls, springy surfaces like cushions, or movement through water. Exoskeletal devices are particularly important to applications involv-

[19] Various exoskeletons are potential candidates for use with various VR systems. Exos (Burlington, MA) produces such systems as the "dextrous hand master," "grip master," and their "exoskeletal arm master." Sarcos Corp. (Salt Lake City, UT) produces expensive teleoperator exoskeletons, including their "dextrous hand master."

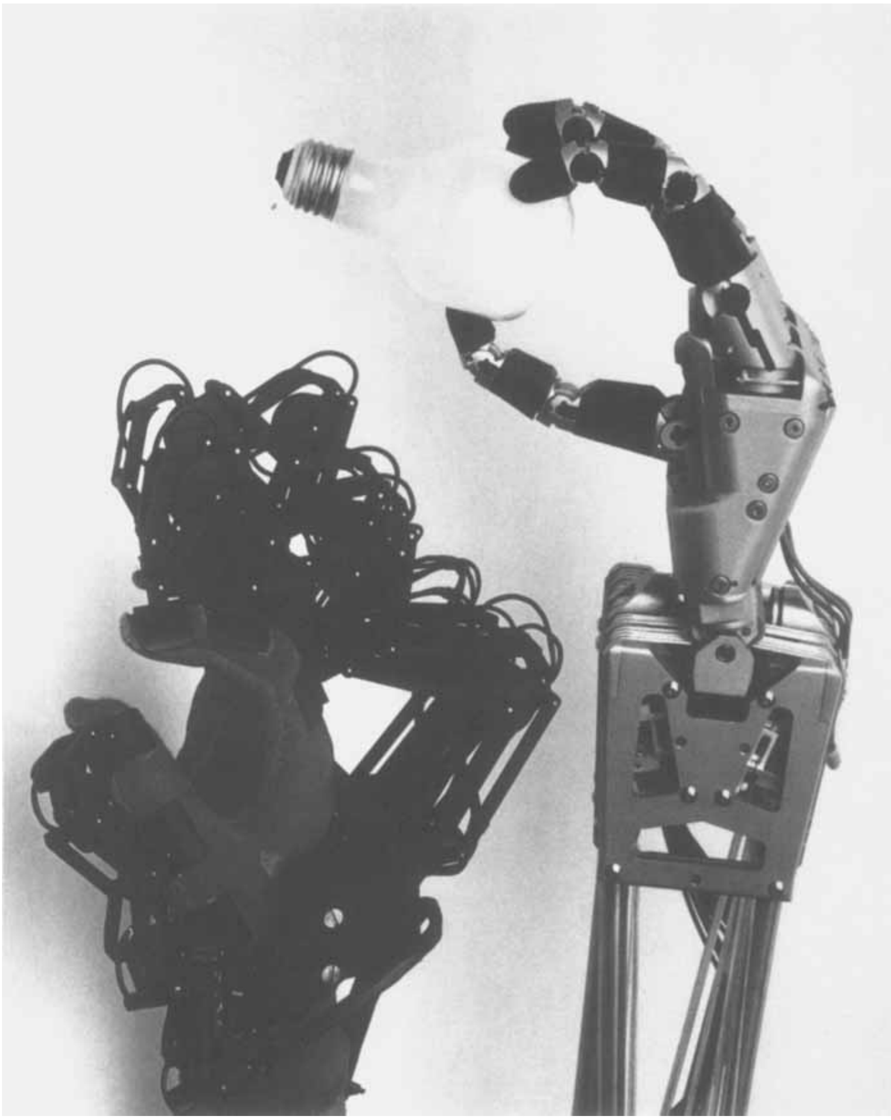

Figure 11. This exoskeletal hand device (left) provides a means of inputting gestures in virtual environments. In purely virtual environments, it controls a computer graphic representation of the user's hand. In telepresence applications, it also controls the movements of a remote robotic hand (right). (Courtesy of Sarcos.)

ing telepresence applications, such as the remote manipulation of robotic devices in space, military combat, microenvironments, or hazardous environmental applications.

Data gloves. As mentioned above, some exoskeletons can be heavy, awkward, or bulky. While they can be precise, the bulk may not be desirable if high precision or force feedback are not part of an application. Light, clothing-like input devices may be a better means of keeping track of limb and hand movements.

The ubiquitous data glove (Zimmerman, 1987) is a popular means of inputting hand movement. VPL's data glove replaces the mechanical devices of an exoskeleton with thin fiber-like devices than can be stitched into a piece of clothing. When the hand is flexed, the fibers in the glove are bent and stretched at the joints.

There are various ways to measure the flexing of the joints to get a computer graphic model of the user's moving fingers. Using small inexpensive light sources (LEDs), VPL's popular data glove shines light down a fiber optic cable stitched onto the back of the glove. The fiber optic cable is altered slightly to be more sensitive to the flexing of the joints. When the hand is flexed, the cables are bent and less light appears at the end of the optical cable. A photosensor detects changes in magnitude of light and the computer translates this information into a estimate of the flexion of a particular set of joints ($\pm 1^{\circ}$) (Zimmerman, 1987).

When combined with a position tracking device, the data provided by the data glove are translated into a three-dimensional representation of the user's hand. The virtual model of the hand moves and flexes in synchrony with the movement and flexing of the user's physical hand.

Other data glove designs measure the flexing of the user's hand using other means, such as tension-sensitive switches at the joints. Mattel's PowerGlove detected the flexing of the hand by monitoring changes in resistance in an electrical current running down a path of conductive ink printed on a flexible plastic strip attached to the back of each finger in the glove (Rheingold, 1991, p. 163–164).

Data suits. If the developmental logic of the technology is to further increase the user's sense of presence in a virtual world, it is not surprising that the engineering principles behind the data glove have been applied to other body parts as well. VPL's work on the development of a data suit reflects the trend toward detecting and using the input from the movement of the whole body into the creation of VR illusion. The greater immersion is attained at the cost of "suiting up." But critics like Krueger (1991) argue that this cumbersome need to put on so much equipment will restrict the use of devices like the data suit to a limited range of applications. Problems have led VPL to suspend development of their data suit.

Facial Expressions and Eye Movements

Facial expressions play a major role in interpersonal communication (Ekman, 1974). In a telecommunication application, a greater sense of "being there" (Brittan, 1992) might be achieved if VR systems could display representations of facial expressions.

While two-dimensional filmic representations of facial expressions can be achieved (e.g., picture phones), it is more difficult to provide a three-dimensional graphic model of facial expressions mapped inside a fully immersive virtual environment. There seems to be a trade-off between the more immersive head-mounted displays and the full capturing of facial expressions. Currently, most systems can display only a user's gross body movements. Output devices like head-mounted displays obscure much of the user's face, especially the user's eyes. Any simple input recording of facial expression is made difficult. With flat panel displays, this is less of a problem. But here, too, we can see some obstacles. Special glasses used to see some stereographic displays give all users that "sunglasses" look. The expressive eyes are masked. It is possible that some of the alternative display technologies may alleviate this problem.

The picture phone and teleconferencing facilities are examples of previous attempts to get facial cues into telecommunication between individuals. Some researchers are not willing to surrender this important communication channel when we go to three-dimensional displays. Rheingold (1991) discusses a very ambitious project at the Advanced Telecommunications Research Institute International in Japan. A goal of the institute is a large telecommunications environment in which VR equipment may be used to create virtual meeting places—teleconferencing in 3-D. A key component of this project is to find ways to capture and communicate facial codes and display them in 3-D in these networked cyberspaces. Facial codes could be carried on replicas of a user's face or used to animate computer graphic "masks," which may or may not resemble the face of the user. The exact means of incorporating and displaying facial codes in the current generation of virtual environments is not clear. Some point to the use of shape-acquisition cameras (see section titled "Gloveless VR?") as a solution to this problem (Krueger, 1991; Rheingold, 1991). But the refinement of this technology for use in real-time VR applications is still a long-term proposition.

Back in 1965, Ivan Sutherland said: "Machines to sense and interpret eye motion data can and will be built. It remains to be seen if we can use a language of glances to control a computer. An interesting experiment will be to make the display presentation depend on where we look" (p. 506). This is exactly what has occurred.

Eye movement trackers have been part of military applications of virtual environments and are considered valuable input for future applications (Aviles et al., 1992; Rheingold, 1991). Eye movements can reveal where we allocate our attention over the visual space (Rayner, 1984). Just as another person may be very aware of the location of your gaze when the two of you are speaking, a computer with eye-movement sensors can also make use of these behavioral indicators of a user's visual attention. According to Richard Bolt, a researcher in this area:

> *One reason for the system to watch the eyes is to open a new channel through which we can detect where the user's attention is directed. The effect can be compared to what children gain when they discover that where the parent is looking is useful to them in comprehending what is transpiring between them and their parent, and, in turn, with the world about them.* (quoted in Rheingold, 1991, p. 231)

Eye movement can be digitally inputted in a variety of ways (Young & Sheena, 1975). Some of the more advanced eye-location sensors use the corneal reflection of infrared light to track eye movements. Other less expensive techniques use eye-muscle potential measures—electro-oculographic techniques. Both types of eye tracking have been used for various forms of computer input (Jacob, 1990) and flight simulation (Rolfe & Staples, 1986), and are likely to be part of high-level virtual environments (Aviles et al., 1992).

As Sutherland had suggested, one way to use eye-movement input in virtual environments is to change the kind of visual display presented within the foveal region of the visual field. These are called *gaze-responsive displays.* Some gaze-responsive displays make use of the fact that the eye's fovea needs a clear, crisp image, while our peripheral vision can make do with a fuzzier image. A maximum-resolution display is presented at the points where the user's vision is focused; a less resolved image is used elsewhere. This helps conserve the limited computing capabilities of the hardware and software. The computer saves computing time by not having to generate a highly resolved image over the full display but only where it is needed by the user, in the user's focal area.

Eye movements are voluntary and can be used to signal intention. Eye movements have been used to issue computer commands in the cockpits of fighter planes. MIT's famous "Put that there" demo combined eye movement and voice input to tell a computer to move items around a screen (Bolt, 1984; Brand, 1988). Japan's Advanced Telecommunications Research Institute is actively exploring the use of eye movement for both computer commands and input for interpersonal communication within advanced virtual telecommunication systems (Rheingold, 1991).

Talking to Computers

The idea of human computer interaction as a kind of conversation has been, until recently, the dominant metaphor in the design of human–computer interfaces (Walker, 1988). You type in a verbal command that approximates a human language like English (in the best of cases!), and the computer responds with language-like symbols. The science fiction literature of the 1950s to the 1970s is filled with images of computers that take voice commands and respond with metallic voices in a kind of pidgin syntax and monotonous rhythm. The conversant computer is best exemplified by the character HAL in the movie *2001*.

Voice-input technologies have been developed to objectify this conversational vision of human–computer interaction (Waibel & Lee, 1990). A number of universities have major speech-recognition projects funded by the Defense Advanced Research Projects Agency (DARPA) and there are significant commercial research projects at IBM, Texas Instruments, and elsewhere. There are a number of commercial voice-input systems that potentially can be used for VR applications (Duchnowski & Uchanski, 1992).

While some speech recognition systems are quite inexpensive, only a few VR systems make significant use of voice input. For example, NASA's early VR system included voice input (Rheingold, 1991, p. 141).

Voice input serves two purposes in a virtual environment: (a) to converse with other humans present in the virtual environment (a kind of virtual "walkie talkie"), and (b) to converse with (command) the computer. The former use is technologically quite simple. The latter use of speech input is more troublesome. The following discussion will concentrate on voice input as a form of computer command.

Developing Speech Recognition Systems

According to Duchnowski & Uchanski (1992), speech recognition systems for VR and other applications face three developmental hurdles to optimal performance.

Continuous speech input. When we speak, we make a continuous stream of sounds. Some systems can identify isolated words (discontinuous speech), but it is more difficult for computer systems to pick out the words in regular speech unless the user clearly stresses the words and pauses (100–250 msec) between words. It will be a challenge to develop a system that does not involve the compromises of discontinuous speech or special speaking styles for human-to-computer speech. But many would consider such compromises acceptable.

Speaker independence. Each one of us sounds different. We pronounce words in different ways. It is a challenge for a computer to recognize the common word among the many variations of sound produced by different speakers. Some speech recognition systems must be trained to recognize one or a few voices. While speaker-independent systems exist with acceptable error rates, the gain of speaker independence is achieved with a loss in vocabulary size or continuous speech abilities.

Vocabulary size. Systems exist that can recognize from 10 to 50,000 words. While the higher numbers seem impressive, we must remember that "recognize" does not mean the semantic understanding of words. Recognition often means nothing more than the ability of the system to successfully translate the sound of a word into its printed (graphemic) version. Speech recognition systems use dynamic programming algorithms to rapidly cut speech sounds (acoustic wave form) into speech units (words, syllable, phonemes, etc.), filter out a specific sound pattern, and match the pattern to stored patterns from prerecorded exemplars of the word ("dynamic time warping"). Statistical or stochastic models based on information theory and Markov models sometimes provide the rules for the matching process.

We are likely to see more complex models as faster processors with greater memory are available for speech recognition and incorporation of speech recognition capabilities into VR environments. Gains in accuracy and flexibility should develop from models that can use the discursive context to assist in speech recognition and understanding (Duchnowski & Uchanski, 1992), something that humans use to disambiguate speech. The use of neural network models may be useful for this and acoustic–phonemic modeling.

Computers have always used a wide variety of input devices, though many of these have been primitive (see Greenstein & Arnaut, 1988). For example, the standard computer mouse is a system that detects two-dimensional position by monitoring the extent and direction of the movement of a track ball located in the underside of the mouse. The standard mouse has two or three buttons which can be used to issue all kinds of commands in tandem with options offered by the user's software.

The three-dimensional mouse[20] used with some VR systems is a similar device equipped with a position tracking sensor to register three-dimensional position information. Such devices are very useful when the user is engaged in three-dimensional design applications such as those found in engineering, architecture, and computer animation. For example, a VR system at the University of North Carolina uses a custom-designed device created by hollowing out a billiard ball and inserting two buttons and a position tracking device inside the ball.

It follows that buttons and position sensors can be attached to all manner of VR props, including guns (e.g., Mattel), racquetball rackets (e.g., Autodesk, see Figure 12), wands (e.g., Lincoln Labs), or any other object that may be a physical prop with a graphic equivalent inside a virtual environment. Input commands may be issued by virtue of the physical location of the prop (e.g., a virtual racket hitting a virtual ball), the pressing of a button on the prop (e.g., a trigger on a gun), by moving a lever or knob (e.g., joysticks and dials), or by pointing a prop in a specific direction (e.g., pointing a wand). It is likely that we will see the development of all types of creative, useful, or frivolous input devices to be used in entertainment and training applications of VR technology.

Psychophysiological Inputs

Psychophysiological input has been proposed for use in virtual environments (Krueger, 1991; Rheingold, 1991; Warner, 1992). While psychophysiological measures have long been used to monitor user responses to mediated experiences (e.g., Cacioppo & Tassinary, 1990; Reeves et al., 1985; Stewart & Furse, 1982), rarely have such responses been used as "feedback," altering the nature of the mediated presentation. But most biofeedback applications involve tying psychophysiological inputs to some sensory display, be it differing tones, lights, or other indicators. Companies have

[20] A number of companies, such as Simgraphics, have introduced a three-dimensional mouse device.

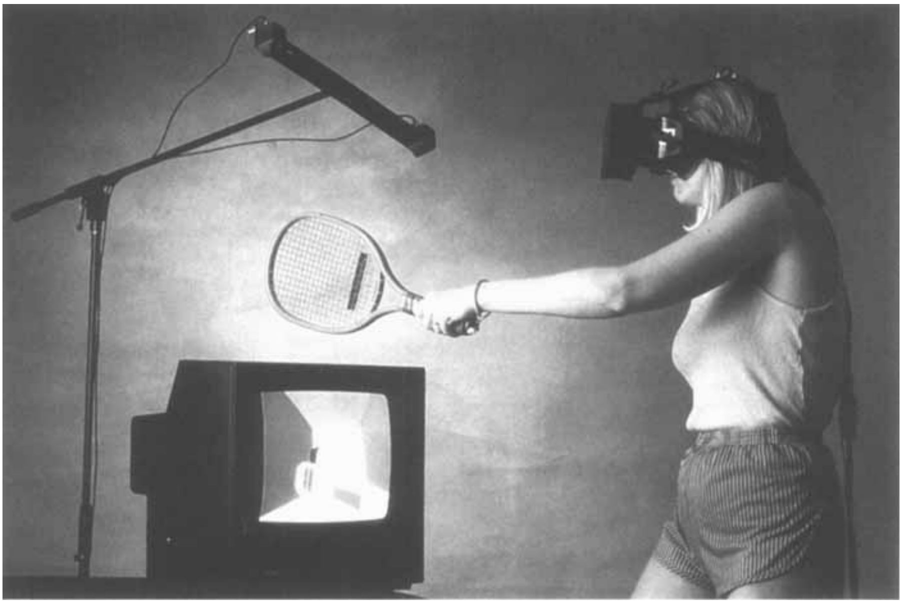

Figure 12. Example of an entertainment application in which a prop, a racket, is used as an input device. Position tracking sensors and/or buttons are likely to be attached to all types of props and input devices to facilitate the virtual illusions of entertainment applications (e.g., wands, pistols, levers, etc.). (Courtesy of Autodesk.)

even marketed so-called therapeutic systems that link a psychophysiological response to patterns on a computer display.

Virtual reality applications in training or entertainment can potentially take this one step further and tie various parameters of the environment (e.g., color, spatial characteristics, motion, activity of agents, etc.) to the digitized values of a psychophysiological measure. Such input could be used for medical reasons, personal mood management, interpersonal communication, or to evaluate training effectiveness. Consciously guided psychophysiological responses can also be used to control external devices (e.g., Keirn & Aunon, 1990). Some psychophysiological measures under consideration for use in virtual environments include heart rate, blood pressure, muscle tension [electromyelography (EMG)], electrocardiogram (EKG), skin resistance [electrodermal activity (EDA)], and brain waves [electroencephalography (EEG)] (see Warner, 1992).

Gloveless VR?

As immersive virtual environments are diffused to various segments of society, two critical technical problems are likely to become an issue: (a) How do we quickly import detailed models of physical objects into virtual worlds, and (b) How do we lessen and finally eliminate some of the devices (such as gloves, suits, etc.) necessitated by current immersive uses of the technology?

The first question is based on the notion that virtual worlds could be furnished by "three-dimensional pictures" of real-world objects like chairs, cars, elephants, etc. Most virtual objects must be built: Their shape (polygon structure) must be drawn using computer-assisted design (CAD) tools, the shading and texturing of surfaces specified, and physical and behavioral properties defined. The building of objects and worlds is still labor intensive, though the goal is to make it easier and accessible to the least trained user. At the moment, there are few objects to populate virtual worlds and the marketing of virtual objects has already begun (Lanier & Biocca, 1992).

The idea that one could take a three-dimensional "picture" of an object and import it into a virtual world appeals to many (e.g., Rheingold, 1991). Many commercial VR systems have the means of importing two-dimensional images, such as textures, and mapping them on a two-dimensional or three-dimensional surface. But the problem of importing pictures is obvious. Pictures are two-dimensional representations of objects. Virtual worlds need three-dimensional digital representations of objects. Shape-acquisition cameras that generate a digital three-dimensional map of the surfaces of an object would be of great value within VR systems.

Technologies exist for mapping a three-dimensional surface using shape-acquisition methods based on visual pattern recognition, ultrasound imaging, and magnetic resonance imaging. A shape-acquisition system that could also map colors, textures, and shapes would be most valuable for VR applications. Shape-acquisition cameras would become the essential tool for quickly constructing models of physical environments for training and other applications. While prototypes of such systems can be found in military and medical technologies, they remain very slow, very expensive, and limited to objects and surfaces of certain sizes.

A number of VR commentators and researchers hold out the hope for what one calls "gloveless VR" (Rheingold, 1991) and another calls "come as you are" technologies (Krueger, 1991). The idea is based on the desire for input devices that use natural body movement without cumbersome equipment like gloves, suits, and exoskeletons—a system that anyone can simply walk in and use. Shape-acquisition cameras able to generate digital models of real-world objects in real time would become the ideal input device. Body movements and facial expressions could "simply" be digitally captured, transported, and represented anywhere else.

The telecommunications value of a full three-dimensional representation of a distant colleague, object, or environment is obvious. For example, NASA is particularly intrigued by the possibility of having ground-based specialists like geologists act upon virtual models of distant environments (McGreevy, 1991). If, in the more distant development of this technology, researchers could successfully combine real-time shape-acquisition cameras with holographic display systems, then VR technology would have taken a critical step to creating the kind of science fiction communication system exemplified by the holodeck of the *Star Trek: The Next Generation* television series.

But like holographic display systems, shape-acquisition technologies are still in their infancy, a distant vision rather than a reality. Real-time versions of the prototypes are still off in the future, but there is disagreement about whether the technologies are in the near or distant future. Shape-acquisition input devices remain at the upper and distant end of the input devices discussed here.

Computer Platforms and Architectures

If the output and input devices constitute the "senses" and "limbs" of a VR system, then the computer platforms and software constitute the "guts" and "brains" of the system. While each area of virtual-environment development shows great activity and variety, this area is probably the most dynamic.

At the heart of most VR systems is a graphics computer. The graphics power of these configurations is often judged by rough measures such as the number of polygons produced per second. A specialist in computer graphics technology, Alvy Ray Smith, estimates that the visual simulation of physical reality would require the production of approximately 80 million polygons per second (quoted in Rheingold, 1991). The best commercially available systems, such as the ones manufactured by Silicon Graphics and Hewlett Packard, rarely break the 1 million polygons-per-second threshold. Except for some of the unique stereoscopic computational issues, virtual environments share many of the same graphic-design problems found in other computer graphic and animation applications.[21]

Among the more sophisticated virtual-environment computing systems are some advanced multicomputers that make extensive use of teams of specialized processors. These specialized processors are designed from the beginning with special "virtual construction" tasks in mind. Working in parallel, these multicomputers are like an assembly line of highly-skilled workers, each constructing one aspect of a virtual world.

As an example, let's consider one of the "high-end" computer platforms used in virtual-environment modeling. One of the most advanced computers supporting VR applications is the Pixel-Planes series of graphics engines designed at the University of North Carolina at Chapel Hill ("Pixel-Planes 5," 1991; Fuchs & Abram, 1985). This specialized graphics engine has three main components (see Figure 13): graphics processors, renderers, and a high-bandwidth ring network that keeps data moving between all the busy

[21] Readers interested in technical issues in the creation of visually realistic virtual worlds should consult Foley, Van Dam, Feiner, and Hughes (1990) and Badler, Barsky, and Zeltzer (1991). Zeltzer (1992) provides a very useful review and evaluation of various computer platforms and graphics software suitable for immersive virtual environments.

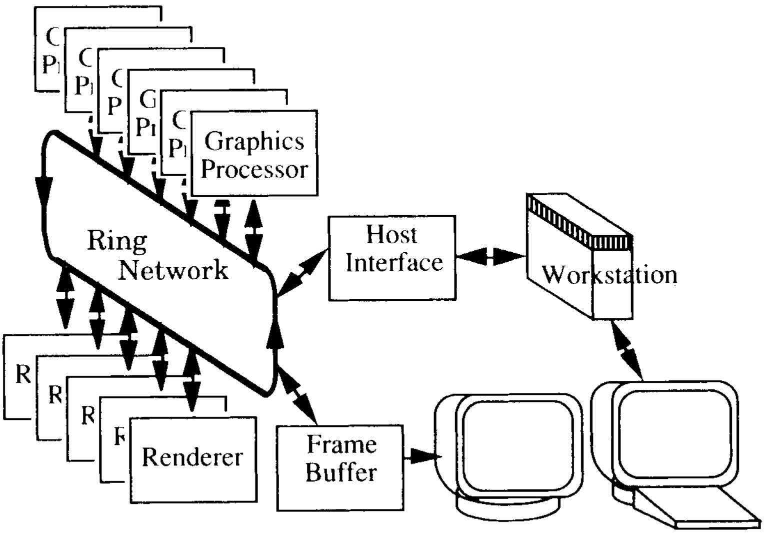

Figure 13. A schematic diagram of the powerful Pixel-Planes 5 graphics engine (see text). (Courtesy of Computer Science Department, University of North Carolina at Chapel Hill.)

components. In VR applications the Pixel-Planes system is linked to a host workstation, which acts as a delivery clerk and feeds input instructions into the graphics engine.

Like all multicomputers, the Pixel-Planes system has an elaborate division of labor. Each group of components can be characterized by the kind of work it performs:

The *graphics processors* are architects; they build the virtual space. The teams of graphic processors calculate the geometric properties of the virtual environment independent of any specific view of that environment. These are math-based processors that calculate graphic primitives such as lines, polygons, and curved surfaces in a matrix-like "image space."

The *renderers* are painters; they paint the image the human viewer sees. Each renderer is in charge of a small section of the screen. The renderers take the calculations from the graphics processors and generate the color and luminance values for a 128 × 128-pixel area of the screen. The renderers send their parts of the image back to the graphics processors for more advanced processing or, more often, directly to the frame buffer. The frame buffer pastes the image together before it is displayed to the viewer.

The *ring network* is a jogging postman rapidly picking up and delivering bits of information. In Pixel-Planes 5, the ring network handles an aggregate data rate of 160 million 32-bit words per second.

The result of all of these components is a very fast graphics engine that rapidly renders objects and environments, handles up to 2.3 million polygons per second, and generates a radiosity lighting model that can handle multiple light sources and elaborate textures. Computer platforms like Pixel-

Planes, composed of mixtures of specialized and general processors, will likely orchestrate the more sensory immersive applications of virtual environments.

The Pixel-Planes system, like all VR systems, is faced with the difficulty of creating vivid, detailed illusions in real time. Virtual worlds are created from an endless series of computations, which can be run on a number of computers. Computations take time. The more detailed and complex the virtual world, the higher the number of computations. Each added layer of sensory vividness, like detailed textures and multijointed simulated figures, takes computing power. Each additional layer of virtual-world modeling increases the complexity of the software.

Because calculations take time, users often experience a perceptual lag in most VR systems. For example, when users rapidly move their heads, they may notice that the visual world lags behind their movements. Because perceptual lags can destroy the illusion of presence, it has been a major concern for developers.

Because computing power is limited and software can support only limited levels of complexity, most VR systems involve compromises. In the following section I discuss the systems that emerge from these compromises.

Making Virtual Reality

Immersive VR setups range from various "garage VR" systems using PCs, improvised software, the cheap Mattel PowerGlove and "home-brewed," head-mounted displays. At the high end we find costly research and military systems running on graphic workstations like the Silicon Graphics Skywriter configurations or powerful graphic engines like Pixel-Planes 5.

In general, the nonspecialist can make some rough distinctions among the dizzying range of platforms and software architectures using a limited set of simple criteria (see Table 1). Each criterion is a construct. These constructs may be influenced by a host of technical variables. The table contains two kinds of criteria: (a) social constructs of user friendliness, such as diffusability and cost, and (b) constructs emphasizing the level of presence, such as sensory vividness and sociability.

Using these simple, basic criteria we can divide the more immersive computer platforms and software combinations into three general classes. Each takes a slightly different approach, trading off cost or performance in the drive to develop more convincing or more available VR illusions (see Table 2).

One approach tries to develop low-end VR systems that can run on a single computer or a pair of commonly available personal computers and workstations such as the high-end Apples, IBM compatibles, or personal workstations like the Silicon Graphics Iris. Present commercial examples of these systems include VPL's Microcosm and W Industries' Virtuality. The advantage of these configurations is their widespread availability and low

Table 1. Basic Criteria for Evaluating VR Platforms and Software

Criterion	Description
Sensory vividness	The capabilities of the base hardware and software, and not just the output devices, contribute to sensory vividness, a key component of the sense of presence. Sensory vividness includes such considerations as (a) the number of sensory channels supported, (b) the sensory resolution within each sensory channel, and (c) the level of coordination between sensory displays and illusions.
Interactivity	Interactivity is also critical to the sense of presence. In VR the criterion can be defined as (a) the number and forms of input and output, (b) the level of responsiveness to conscious and unconscious user actions and states, and (c) the range of interactive experiences (including applications) offered by the system.
Sociability	Sociability is defined as the number of users a system can support. The lowest level of sociability is the single user interacting with the contents of a virtual world. An example of a highly sociable system is the military simulation SIMNET, which supports 200+ sites and hundreds of users.
Diffusability	Diffusability is the likelihood that a system can be adopted and used by various business, educational, and personal users. Variables that can raise or lower diffusability include: (a) Hardware compatibility: Systems that run on computers that are commonly available in organizations (e.g., PCs and low-end workstations) as opposed to those that require highly specialized computers (e.g., Pixel-Planes). (b) Software compatibility: Systems that can import existing two-dimensional and three-dimensional models (e.g., CAD-based models) are more likely to be diffused. For example, this is an important criterion for the CAD-based Autodesk system. (c) Tool integration: Integrated systems with hardware support, modular software (i.e., applications), and catalogs of virtual objects and worlds are more likely to be diffusable. (d) User friendliness: Systems that allow world building with minimal to moderate computing skills.
Cost	Cost is influenced by all of the variables above. We can anticipate the same pattern of development we have seen elsewhere in the computer industry: cost dropping as low-end systems "inherit" high-performance features previously available only on the most costly systems.

cost. Virtual reality applications developed on this hardware could be rapidly diffused throughout society. But the immersive quality of the illusion is less than optimal. The limitation and key disadvantage is the relatively low computing power of the common microcomputer platforms. But, as usual, this may change.

Another approach is to divide the task of creating and maintaining the vir-

Table 2. Likely Range of Immersive Virtual Environments

Level	Characteristics	Examples
Low	*Sensory vividness:* Poor, nonimmersive or low-resolution immersion, mostly limited to visual and aural senses. *Interactivity:* Poor to good, long lags, small movement volume, narrow range of input devices and applications. *Sociability:* Most often restricted to a single user. *Diffusability:* Tends to be good but will vary according to compatibility of the system and its user friendliness. *Cost:* Relatively low, especially as the market matures. *Possible markets and applications:* Arcades, home consumers, "hackers."	"Garage" and "home brew" VR, Sense8, W Industries' Virtuality, VPL's Microcosm.
Middle	*Sensory vividness:* Immersive, fair to good sensory resolution, especially in the visual and aural sensory channels, but poor in tactile and proprioceptive channels. *Interactivity:* Good range of input devices, but with noticeable lag. Broad range of interactive experiences. *Sociability:* Multiuser capability, possible telecommunications capability at later stages. *Diffusability:* Fair to good, depending on market segment, but may be narrowly targeted. *Cost:* Moderate to high. Affordable for institutions but prohibitive for individuals. Rental agreements may make it affordable as an entertainment system. *Possible markets:* Engineering, telecommunication, medicine and rehabilitation, education, high-end consumer or consumer rental (arcade and amusement).	VPL's RB2 and "Voomies", AutoDesk's Trix and Cyberspace Playhouse.[a]
High	*Sensory vividness:* Good to very good, high end of range. Well integrated support for tactile, force feedback, and other proprioceptive illusions. Emphasis on simulation. *Interactivity:* Wide range of input and output devices. Range of interactive experiences may be narrow due to high cost (e.g., high-end flight simulators). *Sociability:* Good to very good but dependent on need in application. *Diffusability:* Poor, likely to require highly specialized hardware, software, and programming support. *Cost:* High to very high. *Possible markets:* Military and civilian flight simulators, military and industrial telepresence applications, high-end teleconferencing applications, scientific visualization, high-end medical and rehabilitation.	NASA, Naval Ocean Systems Center, U.S. Air Force, University of North Carolina's scientific and medical research, NTT lab.

[a] At the time of this writing the VPL "Voomies" and the AutoDesk's TRIX and Cyberspace Playhouse were still at the "promise" stage.

tual environment among a large number of standard microcomputers or workstations (Appino et al., 1992; Bricken, 1990). For example, separate computers may serve the input and output devices, while others handle graphics processing, dialog managing, or other functions. This kind of distributed system is a form of parallel computing where a number of processors work in tandem. While distributed computing and other multicomputing solutions raise the cost of the basic VR platform, some configurations make use of common computers. Therefore, it is economically feasible for organizations to link a group of their computers together to support a multicomputer VR application.

At the higher end of the scale we may see systems using highly specialized components driven by one or more expensive parallel processing machines. While these high-level systems may not be the mass media of virtual reality, they may fuel the most immersive displays and create the location where the sense of presence is most intense.

The computer is a protean technology; virtual reality is a protean medium. As virtual environments begin to diffuse throughout society, the range of these systems will be quite broad. The categories proposed here will certainly increase in complexity. The categories of components and the distinctions among systems will multiply as the virtual environment marketplace bursts into a kaleidoscope of applications and options. Like its cousin, the microchip, a version of this medium may find its way into almost every form of mediated communication. From the low-end to the high-end system, various configurations of these components may grow to simulate every communication channel from a handshake to a book to the video image.

This tutorial has described the Model T of VR technology. The aerodynamic, Formula 1 version may have similar functional parts to the Model T version; both may use steering wheels and rubber tires. But driving the later version may be as dissimilar as the difference between a bumpy chugging over a turn-of-the-century dirt road and a pinned-in-your-seat, white-knuckle blast down some desert straightaway. Some see a distant destination—perfect presence—barely visible on a hazy virtual horizon. Others fear that the destination itself is a virtual image, a mirage. But the medium has left the lab and is already in motion. We will soon be better able to see where and how far this technology may take us.

References

Ackerman, D. (1990). *A natural history of the senses.* New York: Random House.

Alberti, L. B. (1966). *De pictura.* New Haven, CT: Yale University Press. (Original work published 1458)

Appino, P. A., Lewis, B., Koved, L., Ling, D., Rabenhorst, D., & Codella, C. (1992). An architecture for virtual worlds. *Presence, 1*(1), 1–17.

Aviles, W., Durlach, N., Held, R., Pang, X.-D., & Spain, H. (1992). Visual and auditory displays. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-3–III-A-17). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Bach-y-Rita, M. (1982). Sensory substitution in rehabilitation. In L. Illis, M. Sedgwick, & H. Granville (Eds.), *Rehabilitation of neurological patients* (pp. 87–102). Oxford: Blackwell Scientific.

Badler, N., Barsky, B. A., & Zeltzer, D. (1991). *Making them move: Mechanics, control, and animation of articulated figures.* San Mateo, CA: Morgan Kaufmann.

Benedikt, M. (Ed.). (1991). *Cyberspace: First steps.* Cambridge, MA: MIT Press.

Benson, K. B., & Fink, D. G. (1990). *Advanced television for the 1990's.* New York: McGraw-Hill.

Biocca, F. (1987). Sampling from the museum of forms: Photography and visual thinking in the rise of modern statistics. In M. McLaughlin (Ed.), *Communication yearbook 10* (pp. 684–701). Hillsdale, NJ: Lawrence Erlbaum Associates.

Biocca, F. (1988). Opposing conceptions of the audience: The active and passive hemispheres of mass communication theory. In J. Anderson (Ed.), *Communication yearbook 11* (pp. 51–80). Hillsdale, NJ: Lawrence Erlbaum Associates.

Biocca, F. (1992a). Communication within virtual reality: Creating a space for research. *Journal of Communication, 42*(4), 5–22.

Biocca, F. (1992b, May). *The problem of simulation sickness and the diffusion of virtual environments.* Paper presented at the annual meeting of the International Communication Association in Miami, FL.

Biocca, F. (1992c, September). *Communication design and cognition in virtual environments.* Paper presented to the annual conference of Virtual Reality, San Jose, CA.

Blauert, J. (1983). *Spatial hearing: The psychoacoustics of human sound localization.* Cambridge, MA: MIT Press.

Bliss, J., King, B., Kotovsky, J., & Crane, H. (1963). *Tactual perception of visual information. Technical Documentary* (Report No. ASD-TDR-63-732). Wright-Patterson Air Force Base, OH: Air Force Avionics Laboratory.

Bolt, R. (1984). *The human interface: Where people and computers meet.* Belmont, CA: Lifetime Learning Publications.

Brand, S. (1988). *Media lab.* New York: Penguin Books.

Bricken, W. (1990). *Virtual environment operating system: Preliminary functional architecture* (Tech. Rep. No. HITL-M-90-2). Seattle: University of Washington, Human Interface Laboratory.

Brittan, D. (1992). Being there: The promise of multimedia communications. *Technology Review, 95*(4), 42–51.

Brooks, F. (1977). The computer scientist as toolsmith: Studies in interactive computer graphics. In B. Gilchrist (Ed.), *Information processing 77* (pp. 625–634). Amsterdam: North-Holland.

Brooks, F. (1988). *Grasping reality through illusion: Interactive graphics serving science* (Tech. Rep. No. 88-007). Chapel Hill: Dept. of Computer Science, University of North Carolina at Chapel Hill.

Brooks, F., Ouh-Young, M., Batter, J., & Kilpatrick, P. (1990). Project GROPE—Haptic displays for scientific visualization. In *Proceedings of ACM SIGGRAPH 90, Computer Graphics, 24*(4), 177–185.

Burdea, G., Zhuang, J., Roskos, E., Silver, D., & Langrana, N. (1992). A portable dextrous master with force feedback. *Presence, 1*(1), 18–28.

Cacioppo, J., & Tassinary, G. (1990). *Principles of psychophysiology: Physical, social and inferential elements.* New York: Cambridge University Press.

Cadoz, C., Florens, J. L., & Luciani, A. (1984). Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The CORDIS system. *Computer Music Journal, 8*(3), 60–73.

Cadoz, C., & Ramstein, C. (1988). Capture, representation, and composition of the instrumental gesture. *Proceedings of ICMC 90.* Glasgow: ICMC.

CAE. (1986). *Introducing the visual display that you wear.* Quebec: CAE Electronics, Ltd.

Chung, J., Harris, M., Brooks, R., Fuchs, H., Kelley, M., Hughes, J., Ouh-Young, M., Cheung, C., Holloway, R., & Pique, M. (1989). Exploring virtual worlds with head-mounted displays. In *Non-holographic true 3-dimensional display technologies: SPIE Proceedings, 1083,* 15–20.

Comeau, C., & Bryan, J. (1961, November 10). Headsight television system provides remote surveillance. *Electronics,* pp. 86–90.

Czitrom, D. (1982). *Media and the American mind: From Morse to McLuhan.* Chapel Hill: University of North Carolina Press.

Dennett, D. (1991). *Consciousness explained.* Boston: Little, Brown.

Dichgans, J., & Brandt, T. (1978). Visual-vestibular interaction: Effects on self-motion perception and postural control. In R. Held, R. Leibowitz, & H. Teuber (Eds.), *Handbook of sensory psychology, Vol. VIII: Perception* (pp. 756–795). Berlin: Springer-Verlag.

DiZio, P., & Lackner, J. (1986). Perceived orientation, motion, and configuration of the body during viewing of an off-vertical rotating surface. *Perception and Psychophysics, 39,* 39–46.

Duchnowski, P., & Uchanski, R. (1992). Speech recognition. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-33–III-A-53). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Durlach, N. (Ed.). (1992). *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Ekman, P. (1974). *Unmasking the face.* Englewood Cliffs, NJ: Prentice-Hall.

Ellis, S. (1991). Nature and origins of virtual environments: A bibliographic essay. *Computer Systems in Engineering, 2*(4), 321–347.

Ellis, S. R., Kaiser, M., & Grunwald, A. (1991). *Pictorial communication in virtual and real environments.* London: Taylor & Francis.

Fink, D. (1989). The future of high definition television. *SYMPTE Journal, 89,* 89–94, 153–161.

Fisher, S., McGreevy, M., Humphries, J., & Robinette, W. (1986). Virtual environment display system. *Proceedings of 1986 Workshop on Interactive 3D Graphics* (pp. 77–87). Chapel Hill: Dept. of Computer Science, University of North Carolina at Chapel Hill.

Flexman, R., & Stark, E. (1987). Training simulators. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1012–1038). New York: John Wiley & Sons.

Foley, J. D., Van Dam, A., Feiner, A., & Hughes, J. F. (1990). *Computer graphics: Principles and practice.* Reading, MA: Addison-Wesley.

Foster, S. (1988). *Convolvotron user's manual.* Groveland, CA: Crystal River Engineering.

Fuchs, H. & Abram, G. (1985). Specialized computer organization for faster-graphics display. In L. Snyder, J. Jamieson, D. Gannon, & H. Siegel (Eds.), *Algorithmically-specialized parallel computers* (pp. 99–111). New York: Academic Press.

Furness, T. A. (1988). Harnessing virtual space. *Society for Information Display Digest, 16,* 4–7.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

Gibson, J. J., Olum, P., & Rosenblatt, F. (1955). Parallax and perspective during aircraft landings. *American Journal of Psychology, 4,* 27–35.

Gombrich, E. H. (1961). *Art and illusion: A study in the psychology of pictorial representation.* New York: Bollingen Foundation.

Greenstein, J., & Arnaut, L. (1988). Input devices. In M. Helander (Ed.), *Handbook of human factors* (pp. 495–516). Amsterdam: North-Holland.

Haber, R. (1986). Flight simulation. *Scientific American, 141,* 96–103.

Hagen, M. (Ed.). (1980). *The perception of pictures.* New York: Academic Press.

Heim, M. (1991). The erotic ontology of cyberspace. In M. Benedikt (Ed.), *Cyberspace: First steps* (pp. 59–80). Cambridge, MA: MIT Press.

Held, R. (1970). Two modes of processing spatially distributed visual stimulation. In F. Schmitt (Ed.), *The neurosciences: Second study* (pp. 317–323). New York: Rockefeller University Press.

Held, R., & Durlach, N. (1991). Telepresence, time delay and adaption. In S. Ellis, M. K. Kaiser, & A. C. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 232–245). London: Taylor & Francis.

Held, R., Efstathiou, A., & Greene, M. (1966). Adaption to displaced and delayed visual feedback from the hand. *Journal of Experimental Psychology, 2*(3), 171–191.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. Boff, L. Kaufmann, & J. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2* (pp. 21-1–21-55). New York: Wiley.

Holmgren, D. E. (1992). *Laser displays for HMD.* Unpublished technical report, Department of Computer Science, University of North Carolina at Chapel Hill.

Jacob, R. (1990). What you see is what you get: Eye movement-based interaction techniques. *Proceedings of the CHI '90: Human Factors in Computing Systems,* 11–18.

Jacobsen, S., Iversen, E., Knutti, D., Johnson, R., & Biggers, K. (1986). Design of the Utah/MIT dextrous hand. *Proceedings of the 1986 IEEE International Conference on Robotics and Automation,* 1520–1531.

Johnson, A. D., Cutt, P., & Ng, P. F. (1992). *The sense of touch in multisensory interfaces.* Oakland, CA: TiNi Alloy Co.

Kaczmarek, K., Webster, J., Bach-y-Rita, P., & Tompkins, W. (1991). Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering, 38*(2).

Keirn, Z., & Aunon, J. (1990). Man–machine communications through brain-wave processing. *IEEE Engineering in Medicine and Biology,* 55–57.

Kilpatrick, P. (1976). *The use of a kinesthetic supplement in an interactive graphics system.* Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Krueger, M. (1991). *Artificial reality.* New York: Addison-Wesley.

Lackner, J., & DiZio, P. A. (1992). Whole body displays. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-53–III-A-65). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Lanier, J., & Biocca, F. (1992). An insider's view of the future of reality. *Journal of Communication, 42*(4), 150–172.

Laurel, B. (1991). *Computers as theater.* Menlo Park, CA: Addison-Wesley.

Liang, J., Shaw, C., & Green, M. (1991). On temporal-spatial realism in the virtual reality environment. In *Proceedings of the Fourth Annual Symposium of User Interface Software and Technology* (pp. 19–25). Hilton Head, SC.

Lippman, A. (1980). Movie-maps: An application of the optical videodisc to computer graphics. *Computer Graphics, 14*(3), 32–40.

Little, A. (1988). *Sensing human hand motions for controlling dextrous robots.* New York: Academic Press.

Loomis, J. (1992). Distal attribution and presence. *Presence, 1*(1), 113–119.

Matthews, N., & Martin, C. (1978). The development and evaluation of a g-seat for a high performance military aircraft training simulator. In *Piloted aircraft simulation techniques.* Paris: Advisory Group on Aerospace Research and Development (AGARD-CP-249).

McGreevy, M. (1991). *The presence of field geologists in Mars-like terrain.* Unpublished manuscript. NASA Ames Research Center, Moffett Field, CA.

McLuhan, M. (1966). *Understanding media.* New York: Signet.

Meyer, K., Applewhite, H., & Biocca, F. (in press). A survey of position trackers. *Presence, 1*(2).

Minsky, M., Ouh-Young, M., Steele, O., Brooks, F., & Behensky, M. (1990). Feeling and seeing: Issues in force display. *ACM Computer Graphics, 24*(2), 235–243.

Miyashita, T., & Uchida, T. (1990). Cause of fatigue and its improvement in stereoscopic displays. *Proceedings of the Society for Information Display, 31*(3), 249–254.

Nagata, S. (1992). How to reinforce perception of depth in single two-dimensional pictures. In S. Ellis, M. K. Kaiser, & A. C. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 527–545). London: Taylor & Francis.

Pixel-Planes 5. (1991). Chapel Hill: University of North Carolina, Computer Science Department.

Rayner, K. (1984). Visual selection in reading, picture perception, and visual search. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 67–96). London: Lawrence Erlbaum Associates.

Reed, C., Durlach, N., Delhorne, L., & Rabinowitz, W., & Grant, K. W. (1989). Research on tactual communication of speech: Ideas, issues, and findings. In N. McGarr (Ed.), *Research on the use of sensory aids for hearing impaired people.* Washington, DC: A. G. Bell Associates.

Reeves, B., Thorson, E., Rothschild, M., McDonald, D., Hirsch, J., & Goldstein, R. (1985). Attention to television: Instrastimulus effects of movement and scene changes on alpha variations over time. *International Journal of Neuroscience, 25,* 241–255.

Rheingold, H. (1991). *Virtual reality.* New York: Summit Books.

Robinette, W. (1991). Electronic expansion of human perception. *Whole Earth Review, 71,* 16–21.

Robinette, W., & Rolland, J. (1992). A computational model for the steroscopic optics of a head-mounted display. *Presence, 1*(1), 45–62.

Rolfe, J., & Staples, K. (1986). *Flight simulation.* Cambridge, England: Cambridge University Press.

Salisbury, J. K., & Srinivasan, M. (1992). Haptic interfaces. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-27–III-A-32). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Sedgwick, H. (1980). The geometry of spatial layout in pictorial representation. In M. Hagen (Ed.), *The perception of pictures* (pp. 33–90). New York: Academic Press.

Shapiro, M. & McDonald, D. (1992). I'm not a real doctor, but I play one in virtual reality: Implications of virtual reality for judgments about reality. *Journal of Communication, 42*(4), 94–114.

Shaw, E. (1974). The external ear. In W. Keidel & W. Neff (Eds.), *Handbook of sensory physiology, Vol. 1, Auditory system* (pp. 455–490). New York: Springer-Verlag.

Sheridan, T. (1992). Musings on telepresence and virtual presence. *Presence, 1*(1), 120–126.

Shneiderman, B. (1992). *Designing the user interface: Strategies for effective human–computer interaction.* Reading, MA: Addison-Wesley.

Stewart, D., & Furse, D. (1982). Applying psychophysiological measures to marketing research problems. In J. H. Leigh & C. Martin (Eds.), *Current issues and research in advertising* (pp. 25–64). Ann Arbor: University of Michigan Press.

Sutherland, I. (1965). The ultimate display. *Proceedings of the International Federation of Information Processing Congress, 2,* 506–508.

Sutherland, I. (1968). A head-mounted three dimensional display. *FJCC, 33,* 757–764.

TiNi. (1991). *The programmable tactile array.* Oakland, CA: TiNi Alloy.

Venkataraman, S., & Iberall, T. (1990). *Dextrous robot hands.* New York: Springer-Verlag.

VPL. (1989). *VPL EyePhone operations manual.* Redwood City, CA: VPL Research, Inc.

Waibel, A., & Lee, K. (1990). *Readings in speech recognition.* San Mateo, CA: Morgan Kaufmann.

Walker, J. (1988). *Through the looking glass: Beyond "user interfaces."* Sausalito, CA: Autodesk Inc.

Warner, R. (1992). Physiological responses. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-66–III-A-73). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Wells, B. (1992). *A very high-resolution virtual display.* Paper presented at the annual meeting of the SPIE.

Wenzel, E. (1992). Localization in virtual acoustic displays. *Presence, 1*(1), 80–107.

Wenzel, E., Wrightman, F., & Foster, S. (1988a). Development of three-dimensional auditory display system. *SIGCHI Bulletin, 20,* 52–57.

Wenzel, E., Wrightman, F., & Foster, S. (1988b). A virtual display system for conveying three-dimensional acoustic information. *Proceedings of the Human Factors Society, 32,* 86–90.

Winograd, T., & Flores, F. (1987). *Understanding computers and cognition.* Reading, MA: Addison-Wesley.

Young, L., & Sheena, D. (1975). Methods and designs: Survey of eye movement recording methods. *Behavior Research Methods and Instrumentation, 7*(5), 397–429.

Zeltzer, D. (1992). Computational support hardware. In N. Durlach (Ed.), *Virtual environment technology for training* (BBN Systems and Technologies, Report No. 7661) (pp. III-A-74–III-A-94). Cambridge, MA: Virtual Environment and Teleoperator Research Consortium, MIT.

Zimmerman, T. (1987). *Data Glove model 2: Operating manual.* Redwood City, CA: VPL Research, Inc.