Journal of the Ergonomics Society of Korea
Vol.16, No.2, 1997

Virtual Displays and Virtual Environments

ROBERT H. GILKEY*
SCOTT K. ISABELLE** · BRIAN B. SIMPSON**

ABSTRACT

Our recent work on virtual environments and virtual displays is reviewed, including our efforts to establish the Virtual Environment Research, Interactive Technology, And Simulation (VERITAS) facility and our research on spatial hearing. VERITAS is a state-of-the-art multisensory facility, built around the CAVE™ technology. High-quality 3D audio is included and haptic interfaces are planned. The facility will support technical and non-technical users working in a wide variety of application areas. Our own research emphasizes the importance of auditory stimulation in virtual environments and complex display systems. Experiments on auditory-aided visual target acquisition, sensory conflict, sound localization in noise, and localization of speech stimuli are discussed.

* Wright State University, and Armstrong Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio
** Wright State University

INTRODUCTION

Although the concept of virtual environments dates back to the work of Sutherland in the 1960's(e.g., Sutherland, 1970), rapid developments in technology have only recently made virtual environments and virtual displays viable tools for a wide variety of applications. Virtual display technologies are now both more effective and more economical than they were only a few years ago. Although it is still the case that extremely high-end, high-fidelity systems may cost hundreds of thousands or even millions of US dollars, very good quality visual, auditory, or haptic displays are commercially available for only a few tens of thousands of US dollars. In fact, low-end virtual displays can be purchased for only a few hundred US dollars, making them viable for home use. There seems to be little doubt that within a few years, these technologies will directly or indirectly play a significant part in many aspects of our lives.

In recent years, our work in the Signal Detection Laboratory at Wright State University has become increasingly oriented toward virtual environments and virtual displays. We have formed collaborative relationships with a number of other research groups, primarily in the state of Ohio. This paper describes our efforts, through the Ohio Consortium for Virtual Environment Research, to establish a facility for university-based research on virtual environments. We also review our recent research relevant to the design and use of virtual auditory displays.

THE VIRTUAL ENVIRONMENT RESEARCH, INTERACTIVE TECHNOLOGY, AND SIMULATION FACILITY

The Ohio Consortium for Virtual Environment Research

Realizing the tremendous potential for virtual environment technologies, six Ohio universities(Air Force Institute of Technology, Kent State University, Miami University, University of Cincinnati, University of Dayton, and Wright State university) formed the Ohio Consortium for Virtual Environment Research (OCVER) to determine the value of virtual environment technologies in various application areas, to improve these technologies and increase their utilization, to provide training for psychologists and engineers who will work in these technology areas, and to leverage virtual environment technologies for basic research on human performance. The members of OCVER include both psychologists and engineers. Specific research interests of the consortium members include: basic studies of sensory, motor, and cognitive performance; display and control design; computer-aided design and manufacturing; endoscopic surgery; innovations for people with disabilities; and complex data visualization.

Capabilities of the VERITAS facility

Initial efforts have focused on the development of a state-of-the-art facility to support research on virtual environments. An award from the Ohio Board of Regents was used to establish the Virtual Environment Research, Interactive Technology, And Simulation(VERITAS) facility. The facility, which is ouned and operated by wright State University, but housed in the Biodynamics and Biocommunications Division, Crew Systems Directorate, of the Armstrong Laboratory (AL/CFB) at Wright-Patterson Air Force Base, was opened in early 1997. The VERITAS facility consists of a highly immersive visual display subsystem and an integrated spatialized auditory display subsystem; an integrated haptic display subsystem is planned. The visual display subsystem is built around a CAVE^{TM}(Pyramid Systems, Inc.). The CAVE^{TM} technology was first developed at the Electronic Visualization Laboratory at the University of Illinois at Chicago(Cruz-Neira, Sandin, and DeFanti, 1993) and modified for our applications. The VERITAS CAVE includes a set of four rear-projection screens forming a cubical room, about 3.3 m in each dimension. High-resolution stereoscopic images are rear-projected onto the four walls

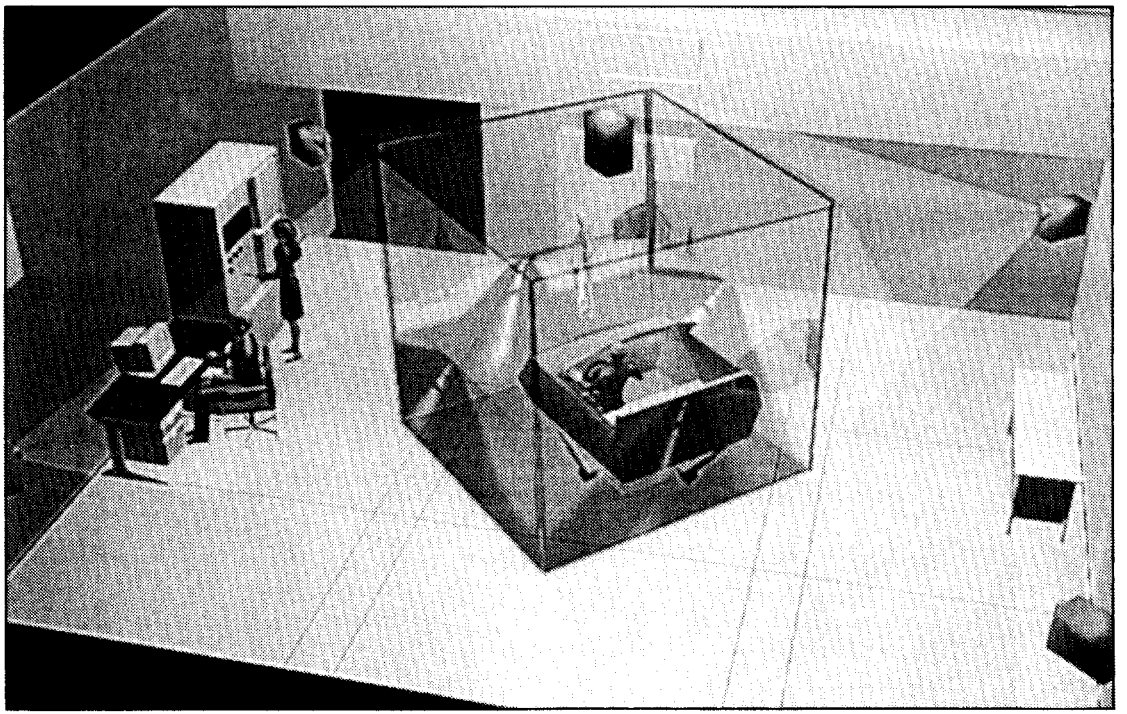

Figure 1. An artist's rendering of the CAVE showing the 4 rear-projected walls and the top-projected floor. As implemented, the arrangement of the projectors in the VERITAS CAVE^{TM} is physically different, but conceptually equivalent.

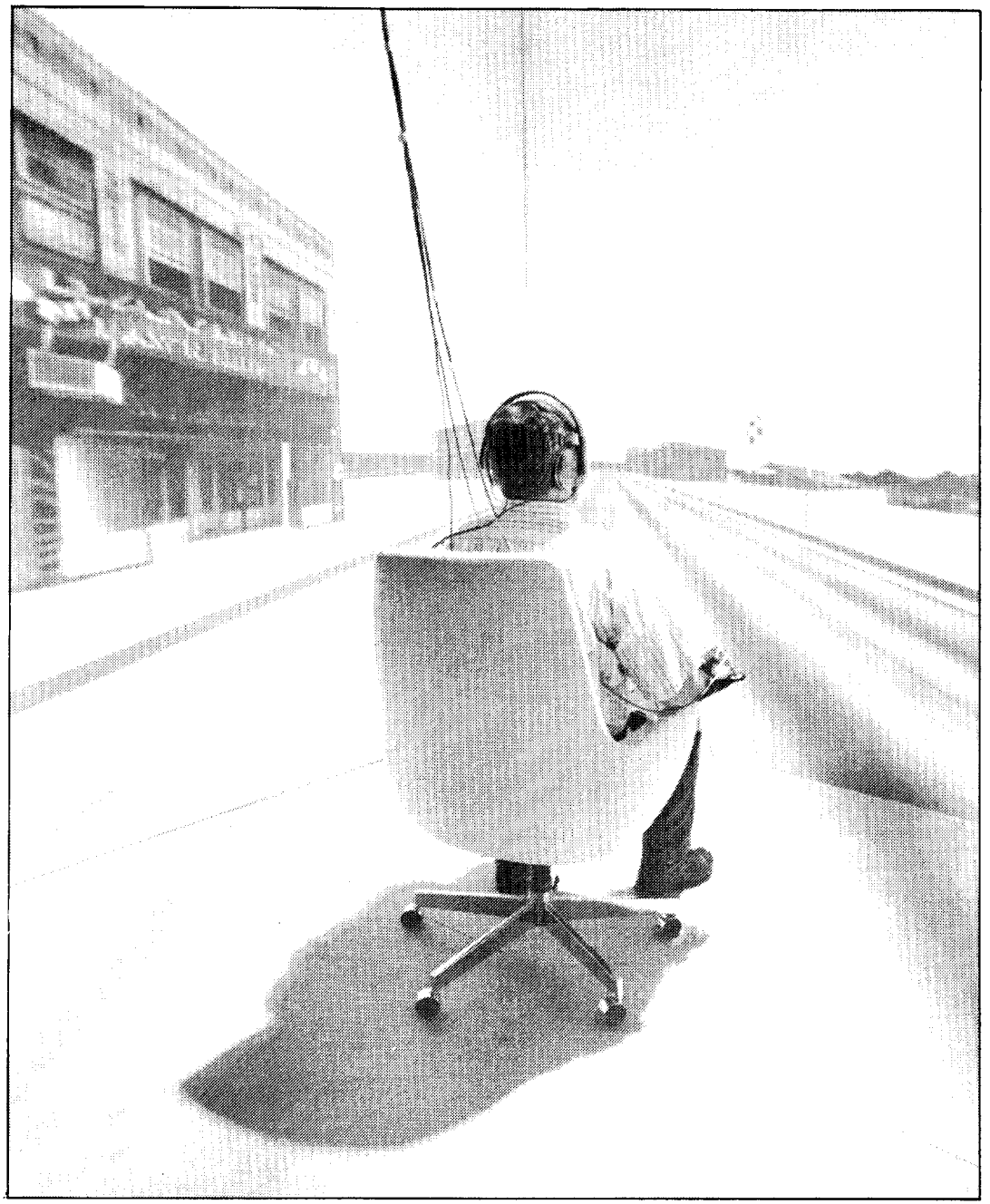

Figure 2. A photograph taken inside the VERitas CAVE™. The imagery depicted is from Performertown by Silicon Graphics. Incorporated. Unfrotunately. the 2-dimensional photograph does not reveal the stereographic capability of the CAVE™.

and top-projected onto the floor by five CRT projectors(Marquee 8500, Electrohome Ltd.). The user is nearly completely immersed, surrounded on all sides and from below with interactive stereoscopic images. The stereo field- sequential technique is used, in which LCD shutter glasses (CrystalEyes, Stereo Graphics Corp.) alternately block image transmission to each eye while the image for the unblocked eye is synchronously drawn on the screen; right and left eye fields can alternate at a rate of up to 160 Hz. Stereoscopic images appear to come out of the walls and can "fill the room" with virtual objects. The user's field of view is limited only by the frames of the shutter glasses, which provide an approximately 105° horizontal field of view. similar to conventional eyeglasses. Imagery on the CAVE walls is generated by a Silicon Graphics(SGI) Onyx computer with multiple CPUs and rendering pipes. The VERITAS CAVE^{TM} projectors can support a maximum of 4 Mpixel per screen, for a total on the five screens of 20 Mpixel. However, for 120-Hz stereo displays(i.e., 60 Hz per eye), projector bandwidth constraints limit the resolution to about 1.25 Mpixel per screen, which is nearly matched to the capabilities of the current graphics subsystem (two SGI InfiniteReality pipes). Figure 1 is an artist's rendering of the CAVE, showing the 4 rear-projected walls and the top-projected floor. A driving simulation is depicted. The automobile "cockpit" shown in the CAVE could be a "real" mock-up, or could be a virtual image produced via the stereographic imagery. Figure 2 is an actual photograph from inside the VERITAS CAVE with a similar scene depicted.

The user can move around the entire interior of the CAVE^{TM}. A 6DOF magnetic tracker (Flock of Birds, Ascension Technology Corp.) is used to monitor the user's head position and orientation in order to compute properly the viewing perspective. The user's hand position and orientation are also magnetically tracked to provide a means for gestural control and interaction with virtual objects. The head and hand tracking subsystem reports position data at a maximum rate of 144 Hz (and can be synchronized to the 120-Hz field rate of the projectors), with accuracy of translation coordinates of better than 2.5 mm and of orientation coordinates of better than 0.1° .

The current virtual auditory display subsystem is based on the PowerSDAC (Tucker-Davis Technologies). This subsystem uses head position and orientation information from the magnetic tracker to compute the appropriate acoustic perspective and present spatialized sounds over headphones. Information about the position of the user, as well as information about the position and movement of virtual objects, is communicated over a high-speed link from the SGI Onyx to the auditory display subsystem (via its host computer), allowing synchronization between the visual and auditory attributes of virtual

objects in the simulated environment The current auditory display subsystem can render up to four simultaneous independent moving sound sources, each with up to 6 first-order reflections, using anechoic Head Related Transfer Functions (HRTFs) of 2.6 ms duration.[2] However, the reverberation times of rooms can range from a few hundred milliseconds, for small well-damped conference rooms, to tens of seconds for cathedral-sized rooms (Beranek, 1954). So. when simulating real rooms, filter-lengths and latency become important considerations. We are currently evaluating systems the could replace the PowerSDAC and provide sufficient computational capacity to support more accurate physical acoustics models, longer filters, lower latency, and more sources(e.g., Huron, Lake DSP).

Because the user's hands are magnetically tracked, hand gestures can be used as a means of control in the virtual environment. To facilitate the use of gestural control, we have integrated a pair of simple contact-closure gloves (PinchGloves, Fakespace); contact between thumb and forefinger can be mapped to a particular control operation in the virtual environment (e.g., moving the user's viewpoint or grasping an object).

Unfortunately, interaction with virtual objects using our current gloves provides no haptic feedback about the properties of the objects. In order to provide haptic feedback, we plan to integrate both a 2DOF force-feedback manual control stick (IE-2000, Immersion Corp.) and a 6DOF force-reflecting handcontroller (CyberImpact, Cybernet Systems Corp.). The 2DOF force-feedback stick can supply forces up to 8.9 Nm and has a bandwidth in excess of 9 kHz. The 6DOF handcontroller can supply forces up to 50 Nm. Although these devices do not provide a completely general haptic interface (e.g., per-finger joint force and torque), they are well-suited to our planned studies of manual control in aerial and other vehicles, force-reflection in molecular interaction simulations, and tool and object manipulation in assembly sequencing studies.

In order to maximize the productivity of the VERITAS facility, we have selected software tools with a high-level user interface and straightforward integration that still provide impressive real-time performance. Our image generation software (Vega, Paradigm Simulation Inc.) has an easy-to-use high-level user interface (LynX editor and Vega API), supports simulation standards (e.g., Distributed Interactive Simulation, DIS), is built on real-time optimized graphics (SGI performer OpenGL, and Infinite Reality), and has a well-established developer community (i.e., The Solution Group) and user community (e.g., archived listservers such as info-vega@paradigmsim.com and info-performer@sgi.com). Off-the-shelf software components are available for a wide variety of peripherals and user interface devices (gloves, trackers, audio DSP engines, etc.), and for a wide variety of visual simulation applications (military tactical,

flight, marine, driving, manufacturing, surgery, etc.). The graphical user interface (GUI) of the simulation editor (LynX, Paradigm Simulation Inc.) should enable even non-programmers to develop, test, and deploy sophisticated visual simulation applications rapidly. A specific example is the reconfigurable high-fidelity flight dynamics simulation package FLSIM (Virtual Prototypes, Inc.). FLSIM is designed to provide real-time simulation of the dynamics of fixed-wing aircraft in a highly flexible manner. Both the specifics of the equations of motion and those of the aerodynamic database particular to the specific simulated platform can be readily modified using graphical tools. FLSIM is also easily integrated with Vega (Virtual Prototypes, Incorporated is a member of The Solution Group).

The CAVE™ approach has a number of advantages and advantages and disadvantages relative to order display systems. As implemented in the VERITAS facility, the CAVE™ includes a number of the properties that previous researchers have argued are critical for determining the sense of presence in virtual environments. The CAVE™ includes high-resolution, high-fidelity displays (as suggested by Sheridan, 1992, and by Zeltzer, 1992), with a wide visual field of view (as suggested by Kalawksy, 1993). Held and Durlach (1992) argued for the importance of a responsive virtual environment generation system with minimal delays; although the CAVE™ is subject to delays from the magnetic tracker similar to those in other generation systems, the impact of these delays is reduced because the walls are pre-drawn, with information that will be approximately correct for the new viewpoint, based on the head position at the old viewpoint. Sheridan (1992) argued that it is important for the users to be able to control their point of view with respect to the environment; within the 3.3 m by 3.3 m confines of the CAVE™, subjects are able to move in a natural fashion (e.g., walking) and are able to change their viewpoint to examine objects as they would in the real world. Durlach and Mavor (1995) argued that isolating the subject from the real environment helps to increase the sense of presence; the VERITAS CAVE™ effectively isolates users from the real environment beyond the walls. However, real objects within the CAVE walls are clearly visible, so cumbersome haptic displays, treadmills, etc. may be difficult to integrate without disrupting the sense of presence. On the other hand, subjects are able to directly view their own body in the virtual environment(a characteristic suggested by Heeter, 1992, to be important for the sense of presence).

The fact that real objects can be readily integrated into the CAVE™ is often useful because real objects are likely to have higher-resolution visual and haptic properties, which can add to the overall realism. For example, in our work on aircraft cockpits, we will be integrating touchscreen monitors into the CAVE™ to present simulated instruments

and multifunction displays (including pushbuttons). However, the mixing of real objects (including the user's hand) and virtual objects within the CAVE™ environment can have undesirable and anomalous visual consequences. For example, real objects can occlude virtual objects, but not vice versa. In addition, real and virtual objects that are close to each other and close to the user will not, in general, be simultaneously in focus. That is, a real object, like the user's hand, which is 15 cm in front of the eye, is in focus at 15 cm, but the virtual object the hand is holding could be in focus more than 3 m away, at the CAVE wall. These kinds of anomalous effects are likely to work against the sense of presence when they occur.

Other disadvantages of the CAVE™ as a display system include its cost and physical size. A standard CAVE™ from Pyramid Systems, consisting of 3 walls and a floor, lists for more than $200,000 US (this does not include the cost of the Silicon Graphics computer to drive the displays). Our CAVE™ is housed in a 7.5 m by 9.0 m room and uses essentially all of the available floor space. At least 4 m ceiling clearance is required.

Planned Research on Uninhabited Aerial Vehicles

A large part of our initial research funding for work in the VERITAS facility comes from the United States Air Force Office of Scientific Research. Projects supported by these funds are concerned with developing effective display and control interfaces for piloting Uninhabited Aerial Vehicles(UAVs). This work illustrates the utility of the CAVE™ system. We plan to use the CAVE™ both as a high-fidelity simulator to mimic the experiences an operator would encounter while controlling a real UAV and as a rapid prototyping environment in which complex display and control representations can be implemented and evaluated in a cost-effective manner.

The task of piloting a UAV is quite different from that of piloting a traditional aircraft. Because of the difficulty in maintaining a reliable communication channel in actual operational settings, there will be noise, bandwidth limitations, long delays, and dropouts in the control loop. That is, much of the visual information about the surrounding environment, vestibular information about the g-forces acting on the aircraft, and auditory information about the plane's status that is directly available to the pilot of a traditional aircraft is likely to be absent or distorted for the operator of a UAV. Similarly, control inputs from the operator may not be reliably received by the UAV. Because of these limitations, most UAV systems are designed to have considerable autonomy resident in the aircraft and relatively little low-bandwidth information to and from the operator. Such a system is likely to be adequate for current UAVs, which are typically not high-performance aircraft and are utilized primarily for reconnaissance missions. However, future UAVs are likely to have

wider performance envelopes and may be used in more varied missions, which will often benefit from a greater level of human-in-the-loop control. Our research focusses on the design of displays and controls that will allow the human operator to maintain effective situation awareness at the remote site and evoke the desired actions by the UAV, thereby maximizing overall system performance. Our approach is to develop meaningful displays and controls that allow higher-order information to be transmitted between the operator and the aircraft. Multisensory displays will provide redundancy and will be used to direct the operator's attention to relevant events.

Several overall formats will be considered for the control environment. One possibility would be to make the remote control environment for a UAV operator similar to the cockpit of a traditional aircraft. In the CAVE™, we would provide an "out-the-window" view on the CAVE™ walls and integrate a "real" cockpit mock-up with instruments and multi-function displays, throttle, pedals, and joystick. A heads-up display could be integrated into the cockpit mock-up or projected on the CAVE™ walls.

A second representation would emulate a command and control center. In this representation, the walls of the CAVE™ could serve as a "data wall" where maps, videos, and status displays for one or more UAVs could be projected. An operator workstation would allow the operator to "drag" images using a 3D mouse or wand from the data wall to their monitor for high-resolution viewing. More than one operator could work in this environment, and team members could interact in a natural fashion.

A third representation will use object-resolved controls and a god's-eye view of the airspace surrounding the UAV. With this representation, the operator would use a 6DOF force-reflecting hand controller as a proxy to the UAV, delivering high-level commands about the position and orientation of the aircraft, rather than low-level commands to the control surfaces. This representation assumes considerable intelligence in the UAV, which would translate these higher-level commands into the appropriate control surface commands. Information about the status of the aircraft, relative to the mission requirements or relative to its own performance envelope, could be relayed back to the operator using force feedback. That is, if the operator tried to move the plane in a inappropriate manner, the hand controller would resist the operator's command, trying to force the "stick" back to a safer or more desirable position.

We anticipate that each of these representations will have certain advantages and disadvantages, depending on the specific limitations of the control channel between the operator and the UAV, and depending on the specific requirements of the mission. Indeed, we anticipate that within the same mission, different representations may be more effective or less effective in specific situations.

For example, a command and control center view may be good for navigation, whereas a pilot-centered view may be good for ordnance delivery. One advantage of the CAVE™ is that it allows us to switch rapidly between representations in order to make comparisons, and would even allow us to switch representations within a particular mission in order to increase overall performance.

In summary, the VERITAS facility provides a multisensory virtual environment and prototyping system. It can be used to support a wide variety of research, including our work on UAVs.

AUDITORY DISPLAYS

The VERITAS facility provides state-of-the-art capability for the study of multisensory determinants of human performance. However, until recently, most of our work in the Signal Detection Laboratory has focused on spatial hearing and auditory displays. Although many people think of visual displays exclusively when they think of virtual environments or cockpit interfaces, the auditory system has many useful properties to recommend it as a display channel. Obviously, the auditory system is the natural communication channel for speech, but in addition the auditory system seems to be naturally wired to alert and to direct the eyes to relevant information. Whereas the visual system can be cut off from the world by merely closing one's eyes, we have no "earlids" to close our ears, so the auditory system is always available to monitor the environment. The auditory system is a 360° -surround sense, so it can provide spatial information about events behind the listener, as well as events in front of the listener.

Design considerations for auditory displays in complex systems frequently do not go beyond providing a monaural channel, which is used for communications and perhaps to present a few tones or buzzers as warning signals. However, introducing spatial information can further enhance the capability of auditory displays. First, spatially separating communication channels has been shown to increase speech intelligibility via the "cocktail party effect"(see Zurek, 1993, for a review). Second, a spatialized auditory warning can be used to direct the eyes to the source of a problem. Third, attaching auditory signals to objects that are outside the field of view allows the user to monitor their position and status. Finally, a spatial cue can be used to deliver non-spatial information. For example, auditory warnings about high-urgency information might appear at a spatial location proximal to the user, while information of lesser importance might appear at more distant locations. Applications of spatialized auditory displays include aviation and other vehicular systems, virtual environments, architectural acoustics, entertainment, and data "visualization".

Auditory stimulation and the experience of presence

Most researchers working on virtual environments have focused on the visual channel. For example, Kalawksy(1993) states that, "The visual channel is the most important interface in a virtual environment system". The National Science Foundation Workshop on Virtual Environments(1992) states that, "Because of the pervasive, dominant role of vision in human affairs, visual stimuli are without question the most important component of the computer-based illusion that users are in a virtual environment" (p.157). This emphasis on the visual system is also evident in expenditures on hardware, software, and research. It is not unusual in a virtual environment system to spend 10 to 100 times the amount of money on the visual channel as is spent on the auditory channel.

In contrast to these views, Gilkey and Weisenberger(1995) suggest that audition, rather than vision, is fundamental for determining the experience of presence in virtual environments (the sense of "being there" in the virtual environment). They build on the work of Ramsdell (1978) who argued that audition plays a unique role in connecting us to the world around us, which is distinct from its role as a communication channel and as an alerting system. Ramsdell based his arguments on his interactions with adults who had experienced a sudden profound hearing loss. Obviously, Ramsdell's patients suffered because of their greatly diminished capacity to communicate with other humans and because of the loss of auditory stimulation to warn of danger and other important events. However, Ramsdell argued that the most devastating effect of deafness was the inability to hear the incidental background sounds of everyday life. According to Ramsdell, it is these sounds that provide us with a sense of connectedness with the "living world". His patients routinely described the world as "dead" or "unreal". That is, these suddenly-deafened adults did not have a good sense of "presence" in the real world. The experiences of these suddenly-deafened adults send a somewhat discouraging message to virtual environment researchers. Ramsdell's patients experienced a loss of presence in the real world, even though all of the other sensory information was perfectly rendered. In comparison, virtual environment generation systems are likely to have poorly-rendered (cartoon-like) visual displays, limited haptic and tactile feedback, conflicting vestibular information, and no olfactory or gustatory input. These results imply that good experiences of presence may always elude virtual environment researchers unless care is taken to represent the auditory environment adequately.

In contrast, our own anecdotal experience with high-quality binaural recordings suggest that auditory information, in some circumstances, may be sufficient to produce a compelling sense of presence. Listeners to our binaural recordings sometimes reach for

virtual telephones and answer questions asked by virtual talkers, but may simultaneously ignore events in the real world, mistaking them for virtual. That is, our listeners have difficulty discriminating between real and virtual events. Moreover, some of our listeners report multisensory illusions, such as feeling the breath of a virtual talker whispering in their ear, noting a shadow as a virtual talker passes in front of their closed eyes, or feeling a vibration when a virtual coin hits a real table on which the listener's hand is placed. That is, a striking sense of presence can be achieved with auditory stimulation alone, and thus the auditory channel may be the most direct and readily achievable means to increase the experience of presence, even in multisensory virtual environments. Moreover, because auditory displays are typically less costly than displays for other sensory systems, this approach may also be the most economical solution.

Implementing auditory displays

Although for some applications it is reasonable to use loudspeakers to present spatialized sounds (e.g., home theater), in many applications using loudspeakers is impractical (e.g., in fighter cockpits). In these situations, it is necessary to present sound through headphones that will invoke the desired virtual spatial image. Note that conventional stereo recordings do not achieve this goal; that is, the spatialized images are heard as inside the head, not at the desired location in the external world. To achieve externalized virtual auditory images, current state-of-the-art headphone-based auditory display technology attempts to recreate the physical stimuli at the listener's eardrums that would be present if the listener were positioned in a real sound field with a real sound source at the desired location relative to the listener. In this way, the listener should have the same perceptual experience as if they were actually in that sound field. A fundamental technique is to impose on the sound source the directional filtering that occurs due to the listener's head, pinnae, and torso. For example, the head provides acoustic "shadowing"(reduction in amplitude) of sounds propagating to the ear away from the sound source, such that an interaural level difference is introduced. Similarly, the difference in propagation path length from the source to the two ears produce differences in the time-of-arrival (interaural time differences). These interaural differences are important cues for the determination of the azimuthal (i.e., right/left) direction of the sound source (Kistler and Wightman, 1992). In addition to these cues, the folds and cavities of the pinnae introduce direction-dependent spectral colorations that are important cues for the determining the elevation of the sound source and for determining whether the sound arose from the front or the rear (Kistler and Wightman, 1992). For a given position in space, the effects of the propagation paths to the two ears are captured by a pair of linear

filters (one for each ear) called Head Related Transfer Functions, or HRTFs (e.g., Wightman and Kistler, 1989). Typically, HRTF pairs are measured on a subject for a fixed set of positions on an azimuth-elevation grid. An arbitrary source signal can then be convolved with the pair of HRTFs that correspond to the desired position, using digital filter techniques. When presented to the listener over headphones (with suitable equalization), the stimulus contains the directionally dependent cues that cause the listener to hear the source signal as coming from the desired virtual location. Usually, special purpose signal processing hardware is used for the convolution. This hardware allows input from a magnetic head tracker such that the HRTFs are changed to correspond to the appropriate new position when the listener's head is moved. That is, in a virtual auditory display, the HRTFs must be updated based on head position in order to provide the auditory stimulation appropriate for a fixed source. If head movements are not properly correlated with the simulated source positions, then a "stationary" virtual sound may appear to move with the head or the virtual illusion may collapse, such that the sound is heard as inside the head (Wallach. 1940).

Auditory-aided visual target acquisition

Headphone-based virtual audio spatialized auditory images to be superimposed on visual targets. In this way, spatialized auditory cuing signals can be used to direct attention to relevant visual information. A number of laboratory studies have shown substantial improvements in visual target acquisition times when a spatially coincident signal is present. Perrott, Cisneros, McKinley, and D'Angelo(1995), for example, had subjects locate an isolated visual target, which could appear in locations that ranged 360° in azimuth and 160° in elevation, against a dark field. In all conditions, target acquisition times were shortest for targets occurring within the central visual field, and became increasingly longer as the distance between the target and the central visual field increased. Specifically, when no spatially correlated auditory cue was provided, target acquisition times ranged from 1000 ms in the central visual field to over 2000 ms for positions in the rear hemifield. However, when the spatially correlated cue was provided, target acquisition times for the rear hemifield were reduced to less than 1250 ms. Even in the central visual field search times are substantially reduced (in some cases search times were as short as 750 ms).

Nelson, Hettinger, Cunningham, Brickman, Haas, and McKinley(1997) compared auditory-aided visual search performance for stimuli presented on a large projection screen (150° wide by 70° high) to performance for stimuli presented on a helment-mounted display (HMD) that provided only a 60° Horizontal by 40° vertical field of view. In both conditions, the same 150° by 70° front-hemifield search field was used.

Broadband auditory signals were convolved with HRTFs to produce spatialized images that, when presented through headphones, appeared at a virtual location coincident with the target. Their results indicated that target acquisition speed was improved when spatialized auditory cues were provided under both viewing conditions, but that the greatest increase in speed occurred in the HMD condition. In addition, subjective measures of perceived workload were significantly reduced when a spatialized auditory cue was present. That is, the auditory cue simultaneously increased task efficiency and reduced task workload.

In these experiments, the more difficult the visual search task, the greater the benefits of spatialized auditory cues. In a pilot study, we measured target acquisition times for a very difficult visual search task. The target (the letter "R") could occur anywhere in a field filled with distractors (the letters "P" and "Q") that extended 360° in azimuth and 60° in elevation. The field was projected on a helmet-mounted display, which limited the field of view to 40° in azimuth and 20° in elevation. Each letter, whether target or distractor, occupied a space that subtended approximately 5° in azimuth and 4° in elevation, so that 40 letters were visible to the participant at any given time. On half of the trials, the participants received no auditory cues. On the other half, the auditory stimulus was convolved with the appropriate head-related transfer functions for each ear so that the sound would appear to come from the direction of the visual target. The position of each participant's head was monitored with a magnetic head tracker so that the auditory filters could be updated appropriately when the head moved. The spatialized auditory cue decreased visual target acquisition times by more than a factor of 8 compared to visual search without the auditory cue.

There is also evidence that these laboratory results are likely to generalize to real-world settings. McKinley and Ericson (1997) reported the results of in-flight tests employing the similar 3D audio hardware to that used by Nelson et al. (1997). Under task conditions similar to those of Nelson et al., pilots using 3D audio in a T-1 AV-8B Harrier attack aircraft reported that visual targets were acquired more readily when spatialized auditory cues were provided than when visual symbology was provided as an aid. Moreover, pilots indicated that the 3D auditory display improved communication and decreased workload, thus improving overal situational awareness.

Problems with auditory displays

Auditory displays have tremendous potential in a number of application areas, as indicated by our work on presence and our work on auditory-aided visual target acquisition. However, there are still issues that need to be resolved to realize this potential fully. Systematic misperceptions are often reported in auditory virtual environments. Sounds are

often poorly localized in elevation, and sounds presented in front of the listener are often biased towards higher elevations. Typically, sounds are heard as closer to the listener than was intended by the rendering software. Perhaps most troubling is the fact that sounds presented in front are often heard as coming from the rear, whereas sounds presented in the rear are sometimes heard as coming from the front. Although these front/back reversals are often experienced with real sounds (particularly in anechoic environments), they are more common with virtual auditory displays. The absence of individualized HRTFs (i.e., transfer functions based on measurements of the subject's own head and ears) is often used as an explanation for the poor perception in auditory displays. However, the data do not clearly demonstrate whether individualized head-related transfer functions are necessary or sufficient to eliminate these problems (e.g., Wenzel, Arruda, Kistler, and Wightman, 1993).

Conflicting stimulation and the experience of presence

Misperceptions in auditory displays may be exacerbated by multisensory interactions, notably the ventriloquist effect, where a sound is drawn toward a visual target that seems to be a likely source of the sound. A similar effect may underlie anecdotal reports from casual listening to binaural recordings that indicate that the recordings are most compelling when the listener hears them while seated in the room in which the recordings were made. Simpson, Hale, Isabelle, and Gilkey(1996) hypothesized that when the actual room in which the subject was seated "matched" the virtual room to which the subject was listening, the greatest sense of presence would be achieved. They made binaural recordings of "real-world" stimuli(e.g., keys, telephone, speech, etc.) in three rooms that varied in size from 16 m^3 to 194 m^3 and had subjects listen to the sounds over headphones. In general, the virtual sounds(i.e., the recordings) were rated as more realistic(and more like they were in the room with the subjects) when the subjects listened to the recordings in the same room where the recordings were made. That is, a sound recorded in the 16 m^3 room. Sounded less realistic when heard in the 194 m^3 room than when heard in the 16 m^3 room. The effects were relatively small, but significant. These results may indicate that in order to achieve the best experience of presence, it may be important to be certain that the room acoustics model used in a virtual environment matches the visually depicted room. However, it is uncertain at this time whether the observed effects are driven by the mismatch between visual and auditory stimulation or by a mismatch between current and previous auditory experience in the room.

Loclization in Noise.

Most work relevant to 3D spatial auditory displays has been conducted in laboratory settings. Typically these studies have used simple stimuli, such as tones, clicks, or

broadband noise, in simple environments with no reflections and no interfering stimuli. In contrast, applied settings are likely to be noisy, the required stimuli are likely to be spectrally and temporally complex(e.g., speech), and the simulated environments are likely to require multiple sources and multiple echoes. Good and Gilkey (1996) considered the impact that noise might have on performance with spatialized audio. Their experiment was conducted in the free field(i.e., not with a virtual display). The subject's task was to localize a brief (268 ms) wideband(0.53~11.0 kHz) click-train in the quiet, or in the presence of a single, 468 ms broadband (0.41~14.2 kHz) masker, which was always presented from directly in front of the subject within the horizontal plane. On each trial, the location of the signal was randomly chosen from 239 possible locations, which ranged from -45° to + 90° in elevation and surrounded the subject(360°) in azimuth. The subject responded using the God's Eye Localization Pointing(GELP) technique, as described by Gilkey, Good, Ericson, Brinkman, and Stewart(1995). In this technique, subjects point to a 20 cm spherical model of auditory space with a magnetic stylus in order to indicate the perceived direction of the sound. Across trials, subjects made 6 localization judgments for each signal position under each signal to noise ratio condition. For analysis, the target directions and judgment directions ware converted to the 3-pole coordinate system as described by Kistler and Wightman(1992). In this condition, the left-right(L/R) coordinate is the angle between the median plane and a vector from the center of the subject's head pointing in the judged or target direction, such that 90° L/R is directly to the right, and -90° L/R directly to the left. Note also that 45° azimuth and 135° azimuth in the horizontal plane both have the samp L/R coordinate, 45° . In a similar manner, the front/back(F/B) coordinate is computed relative to the frontal plane, and the up/down(U/P) coordinate is computed relative to the horizontal plane. This coordinate system is useful for analysis : 1) because it distinguishes between people's ability to judge laterality(which is typically good) and their ability to determine front from back (which is often poor) : and 2) because our current understanding of sound localization suggests that different acoustic cues determine performance in the L/R dimension compared to the U/D and F/B dimensions.

Figure 3 shows the proportion of variance(r^2) in the subject's judgments that is accounted for by a linear relation between the judged and target angles. The value of r^2 is shown as a function of signal-to-noise ratio, where 0.0 dB corresponds to a signal that is just barely detectable when it is presented from the same speaker as the masker. The values of r^2 have been averaged across subjects. As can be seen, subjects are quite accurate in their L/R judgments in the quiet, showing a very high correlation between

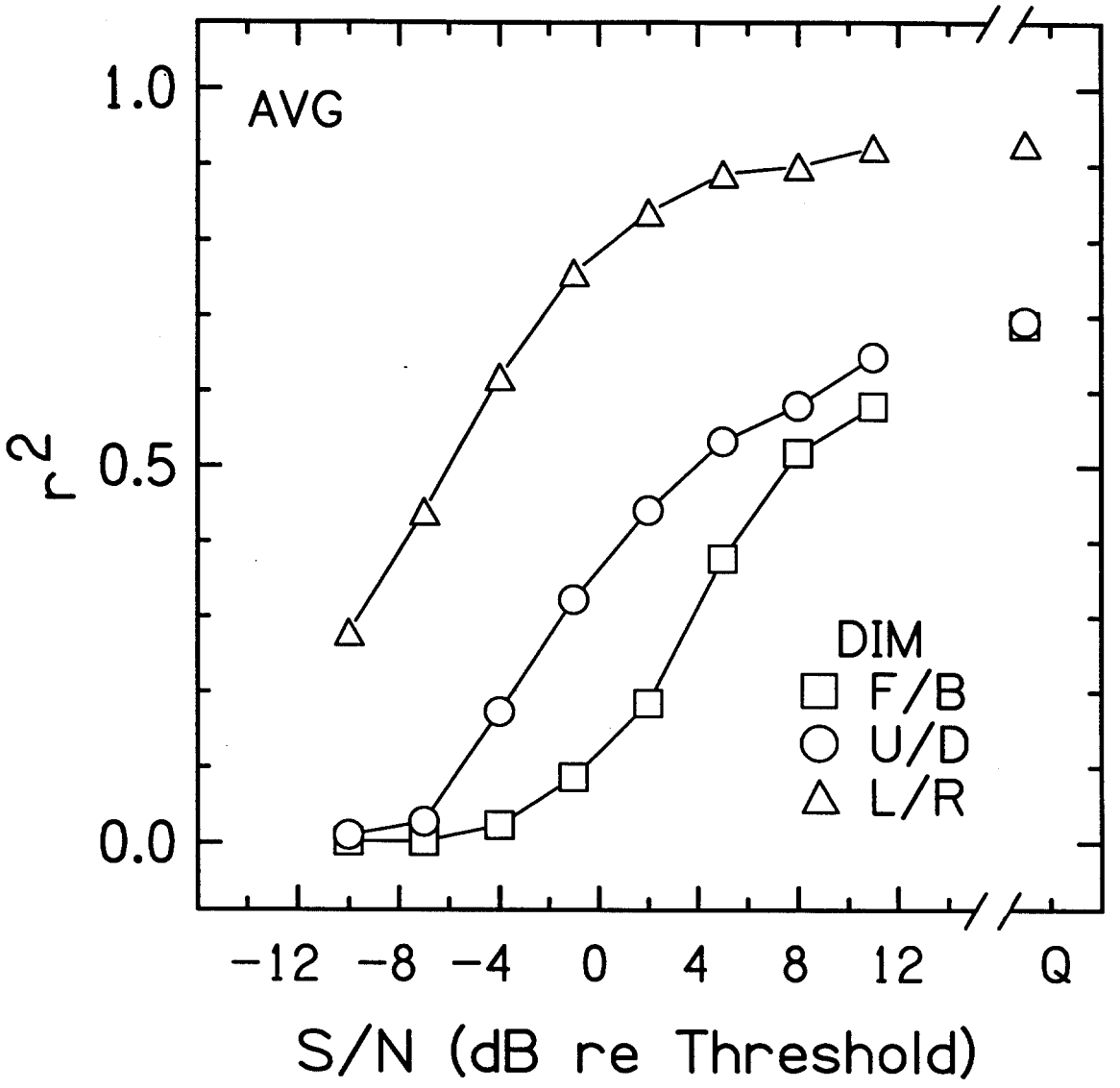

Figure 3. The proportion of variance accounted for(r^2) by the relation between the judged angles and the actual target angles is shown as a function of signal-to-noise ratio for the L/R [△], F/B [□], and U/D[○] dimensions. The results have been averaged across subjects. A 0.0 dB signal-to-noise ratio is defined as the signal level that produces a just-detectable signal when the signal and masker are presented from the same speaker, directly in front of the subject. (From Gilkey, Simpson, Isabelle, Anderson, and Good, 1997, with permission : based on the results of Good and Gilkey, 1996).

judged and target responses. Performance in the F/B and U/D dimensions is noticeably worse, but still relatively good, in the quiet. When the signal-to-noise ratio is lowered, performance in the L/R dimension remains near that in the quiet, until the signal-to-noise ratio is reduced to quite low levels. In contrast, performance in the F/B and U/D dimensions is affected by noise at much highter signal-to-noise ratios, with the F/B dimension being the most affected.

As expected, sound localization performance degrades systematically, and nearly monotonically, as signal-to-noise ratio is reduced. However, the effects of noise are greater for the U/D dimension and greatest for the F/B dimension.

Localization of Speech

A similar pattern of results is observed when listeners localize speech stimuli in the quiet. Gilkey and Anderson(1995) used similar procedures to those of Good and Gilkey(1996). Subjects were to localize signals that could arise from any of 239 locations, which surrounded the subject in azimuth(360°) and ranged from -45° to +90° in elevation. The subjects used the GELP technique to indicate the perceived locations of the sounds. In separate blocks of trials, the signal was either a broadband(0.40 ~11.0 kHz) click-train, or monosyllabic speech tokens taken from the Modified Rhyme Test(House, Williams, Hecker, and Kryter, 1965), which were spoken by a male or a female talker. No masker was present.

Figure 4 shows the difference in localization accuracy between speech and click signals. As can be seen, in the L/R dimension, performance for speech and click signals is approximately equal. In the U/D dimension, for 3 out of 4 subjects, performance is moderately worse with speech signals. In the F/B dimension, performance is substantially warse for all 4 subjects when speech is used as a signal.

Both the localization in noise experiment and the speech localization experiment indicate that localization accuracy observed in applied settings is likely to be poorer than that which has previously been demonstrated in the laboratory. Moreover, different aspects of the localization judgment are likely to be differentially affected by more demanding stimulus conditions. Performance in the L/R dimension is usually quite good. The U/D dimension is more easily disrupted, and localization performance in the F/B dimension can be quite poor even in situations where performance in the L/R dimension is close to optimal. It should be noted that in both of these experiments, the subject's head was stationary. Perhaps head movements would have reduced these effects. Nevertheless, designers of auditory displays should seriously consider what type of information they wish to convey and how the stimulus situation encountered by the user is Likely to limit the ability of the display to transimit this information. In situations where F/B or U/D information is critically important, care should be taken to make sure that sound localization cues are

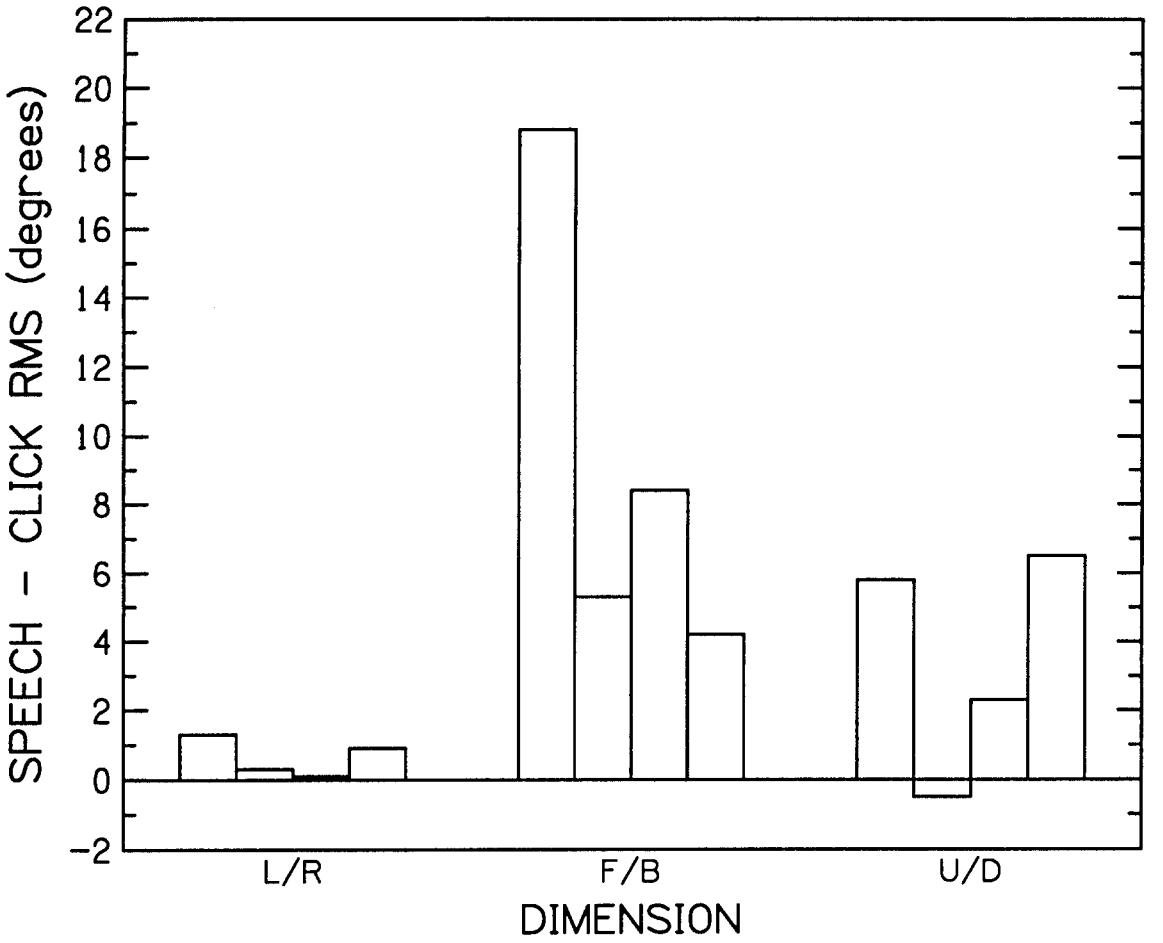

Figure 4. The difference in rms error between the speech target and the click target conditions for the L/R, F/B, and U/D dimensions. Separate bars in each cluster show the results for individual subjects. RMS error is computed between the judged and target angle in degrees. (From Gilkey et al.. 1997, with permission ; based on the results of Gikey and Anderson, 1996)

faithfully represented or that non-auditory cues are used to supplement the auditory information.

SUMMARY

Wright State University, which is the lead organization in the Ohio Consortium for Virtual Environment Research, has established the VERITAS facility to promote research on virtual environments. The state-of-the-art multisensory facility will be used for a wide variety of applications. In particular, early work in the VERITAS facility will focus on

the development of display and control designs for piloting uninhabited aerial vehicles.

We plan to leverage our history of research in auditory processing to develop better auditory displays. In particular, we are interested in using 3D virtual audio to enhance the sense of presence in virtual environments and to direct attention in complex systems.

ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Air Force Office of Scientific Research(AFOSR-91-0289, F49620-95-1-0106, F49620-97-1-0231), from the Department of Defense (F49620-97-1-0118), from the U.S. National Institute of Deafness and Other Communicative Disorders(DC-00768), and from the Ohio Board of Regents Investment Fund, Action Fund, and Research Challenge program, We would like to thank James W. Kondash, Jeffrey P. Shapiro., and Janet M. Weisenberger for reading and commenting on an earlier draft of this manuscript.

REFERENCES

Beranek, L.L.(1954). Acoustics. New York, McGraw-Hill.

Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A.(1993). Surround-screen projection-based virtual reality : The design and implementation of the CAVE. Proceedings of SIGGRAPH 93, 135~142

Durlach, N. I., and Mavor, A. S. (Eds.) (1995). Virtual reality : scientific and technological challenges. Washington, D.C. : Academic.

Gilkey, R.H., and Anderson, T.R.(1995). The accuracy of absolute localization judgments for speech stimuli. Journal of Vestibular Research. 5, 487~497.

Gilkey, R. H., Good, M. D., Ericson, M. A., Brinkman, J., and Stewart, J.M.(1995). A pointing technique for repidly collecting localization responses in auditory research. Behavioral Research Methods. Instrumentation and Computers. 27, 1~11.

Gilkey, R. H., Simpson, B. D., Isabelle, S.K., Anderson, T.R., and Good, M.D. (1997). Design considerations for 3-D auditory displays in cockpits. AGARD Conference Proceedings 596 : Audio Effectiveness in Aviation, 2.1~2.10.

Gilkey, R. H., and Weisenberger, J. M. (1995). The sense of presence for the suddenly-deafened adult : Implications for virtual environments. Presence : Teleoperators and Virtual Environments. 4, 357~363.

Good, M. D., and Gilkey, R. H.(1996). Sound localization in noise : I. Effects of signal-to-noise ratio. Journal of the Acoustical Society of America, 99, 1108~1117.

Heeter, C.(1992). Being there : The subjective experience of presence. Presence : Teleoperators and Virtual Environments. 1, 262~271.

Held, R. M., and Durlach, N. I. (1992).

Telepresence. Presence : Teleoperators and Virtual Environments. 1. 109~112.

House, A.S., Williams, C.E., Hecker, H.L., and Kryter, K. D. (1965). Articulation- testing methods : Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America. 37, 158~165.

Kalawksy, R.S.(1993). The true science of virtual reality and virtual environments. Workingham, England : Addison-Wesley.

Kistler, D.J., and Wightman, F.L.(1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. Journal of the Acoustical Society of America. 91, 1637~1647.

McKinley, R. L., and Ericson, M. A. (1997) Flight demonstration of a 3D auditory display. In R. H. Gilkey and T. A. Anderson(Eds.). Binaural and spatial hearing in real and virtual environments (pp. 683~700). Mahwah, NJ : Erlbaum.

National Science Foundation(1992). Research directions in virtual environments : Report of an NSF Invitational Workshop. Computer Graphics. 26, 153~177.

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., and McKinley, R. L. (1997). The effects of localized auditory information on visual target detection performance using a helmet-mounted display. Article subvmitted for publication.

Perrott, D. R., Cisneros, J., McKinley, R. L., and D'Angelo, W. R. (1995). Aurally aided detection and identification of visual targets. In Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, Volume, Volume 1(pp. 104~108). Santa Monica, CA : Human Factors Society.

Ramsdell, D. A.(1978). The psychology of the hard-of-hearing and the deafened adult. In H. Davis and S. R. Silverman (Eds.), Hearing and deafness(4th ed., pp. 499~510). New York : Holt, Rinehart, and Winston.

Sheridan, T.(1992). Musings on telepresence and virtual presence, Presence : Teleoperators and Virtual Environments. 1, 120~126.

Simpson, B. D., Hale, D. W., Isabelle, S. K., and Gilkey, R. H. (1996). The experiences of untrained subjects listening to virtual sounds. Journal of the Acoustical Society of America, 100, 2633.

Sutherland, I.E.(1970). Computer displays. Scientific American, 222, 56~81.

Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. Journal of Experimental Psychology. 27, 339~367.

Wenzel, E. M, Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America, 94, 111~123.

Wightman, F. L., and Kistler, D. J.(1989).

Headphone simulation of free-field listening. I : Stimulus synthesis. Journal of the Acoustical Society of America, 85, 858~867.

Zeltzer, D. (1992). Autonomy, interaction, and presence. Presence : Teleoperators and Virtual Environments. 1, 127~132.

Zurek, P. M.(1993). Binaural advantages and directional effects in speech intelligibility. In G.A. Studebaker and I. Hochberg (Eds.), Acoustical factors affecting hearing aid performance (pp. 255~276). Boston : Allyn and Bacon.

FOOTNOTES

[1]Requests for reprints should be sent to Robert H. Gilkey, Department of Psychology, Wright State University, Dayton, OH 45435.

[2]The Power SDAC hardware includes digital delay lines that can provide a room response of up to 655 ms, but the model is sparse in that only a total of seven 2.6-ms time intervals can contain non-zero filter data, for four sources. If a single source is simulated, the model can have up to twentyeight 2.6-ms time intervals containing non-zero filter data.