VIRTUAL ACOUSTICS DISPLAYS

Elizabeth M. Wenzel
NASA-Ames Research Center, MS 262-2
Moffett Field, CA 94035

Scott S. Fisher
11571 Buena Vista Drive
Los Altos Hills, CA 94022

Philip K. Stone
Sterling Software
NASA-Ames Research Center, MS 262-6
Moffett Field, CA 94035

Scott H. Foster
Crystal River Engineering
12350 Wards Ferry Road
Groveland, CA 95321

Abstract

This paper describes the real time acoustic display capabilities developed for the VIrtual Environment Workstation (VIEW) Project at NASA-Ames Research Center. The acoustic display is capable of generating localized acoustic cues in real time over headphones. An auditory symbology, a related collection of representational auditory "objects" or "icons," can be designed using ACE, the Auditory Cue Editor, which links both discrete and continuously-varying acoustic parameters with information or events in the display. During a given display scenario, the symbology can be dynamically co-ordinated in real time with three-dimensional visual objects, speech, and gestural displays. The types of displays feasible with the system range from simple warnings and alarms to the acoustic representation of multidimensional data or events.

Introduction

Recent years have seen many advances in computing technology with the associated requirement that users manage and interpret increasingly complex systems of information. As a result, an increasing amount of applied research has been devoted to a type of reconfigurable interface called the virtual display. Some of the earliest work in this area was done by Sutherland [30] at the University of Utah using binocular head-mounted displays. Sutherland characterized the goal of virtual interface research, stating that, "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." As the technology has advanced, virtual displays have gone beyond the flat CRT screen, assuming a three-dimensional spatial organization which, it is hoped, provides a richer and more natural means of accessing and manipulating information. A few projects have taken the spatial metaphor to its limit by directly involving the operator in the data environment [5], [19], [21]. Thus, the kind of "artificial reality" once relegated solely to the specialized world of the cockpit simulator is now being seen as a next step in interface development for all types of advanced computing applications [20].

Auditory Icons & Symbologies

As with most research in information displays, virtual displays have generally emphasized visual information. Many investigators, however, have pointed out the importance of the auditory system as an alternative or supplementary information channel, particularly when the visual channel is overloaded and visual cues are degraded or absent [12], [13], [27]. Most recently, attention has been devoted to the use of non-speech audio as an interface medium [1], [2], [8], [23]. Auditory signals are detected more quickly than visual signals and tend to produce an alerting or orienting response. Consequently, non-speech audio has been most frequently used in simple alarm or warning systems, as in aircraft cockpits or the siren of an ambulance. Another advantage of audition is that it is primarily a temporal sense and we are extremely sensitive to changes in an acoustic signal over time. This feature tends to bring any such acoustical event to our attention and conversely, allows us to relegate sustained or uninformative sounds to the background. Thus audio is particularly suited to monitoring changes over time, for example when your car engine suddenly begins to malfunction. Non-speech signals have the potential to provide an even richer display medium if they are carefully designed with human perceptual abilities in mind. Just as a movie with sound is much more compelling and informationally-rich than a silent film, so could a computer interface be enhanced by an appropriate "sound track" to the task at hand. If used properly, sound need not be distracting or cacophonous or merely uninformative. Principles of design for auditory icons and symbologies can be gleaned from the fields of music, psychoacoustics, and psychological studies of the acoustical determinants of perceptual organization. For example, one can think of the audible world as being composed of a collection of acoustic "objects." Various acoustic features, such as timbre, intensity, and temporal rhythm, specify the identities of the objects and perhaps convey meaning about discrete events or ongoing actions in the world and their relationships to one another. One could systematically manipulate these features and create an auditory symbology which operates on a continuum from "literal" everyday sounds, such as the clunk of mail in your mailbox (e.g., Gaver's 'Sonic Finder' [23]), to a completely abstract mapping of statistical data into sound parameters [4], [28].

Such a display could be further enhanced by taking advantage of the auditory system's ability to segregate, monitor, and switch attention among simultaneous sources of sound. One of the most important determinants of acoustic segregation is an object's location in space.

A true three-dimensional auditory display could potentially improve information transfer by combining directional and iconic information in a quite naturalistic representation of dynamic objects in the interface. Borrowing a term from Gaver [23], an obvious aspect of "everyday listening" is the fact that we live and listen in a three-dimensional world. Indeed, a primary advantage of the auditory system is that it allows us to monitor and identify sources of information from all possible locations, not just the direction of gaze. This feature would be especially useful in an application that is inherently spatial, such as an air traffic control display for the tower or cockpit, or even in a two-dimensional interface which has adopted a spatial organization, such as the desktop metaphor. A further advantage of the binaural system, often referred to as the "cocktail party effect" [10], [16], is that it improves the intelligibility of sources in noise and enhances the segregation of multiple sound sources. This effect could be critical in applications involving encoded information as in scientific "visualization," using the acoustic representation of multi-dimensional data [4], [28], or the development of alternative interfaces for the visually impaired [15], [28]. Another aspect of auditory spatial cues is that, in conjunction with other modalities, they can act as a potentiator of information in the display. That is, visual and auditory cues together can reinforce the information content of the display and provide a greater sense of presence or realism in a manner not readily achievable by either modality alone [1], [7], [11], [26], [29], [31]. This phenomenon will be particularly useful in telepresence applications, such as advanced teleconferencing environments, shared electronic workspaces, or monitoring telerobotic activities in remote or hazardous situations. Thus, the combination of direct spatial cues with good principles of iconic design could provide an extremely powerful and information-rich display which is also quite easy to use.
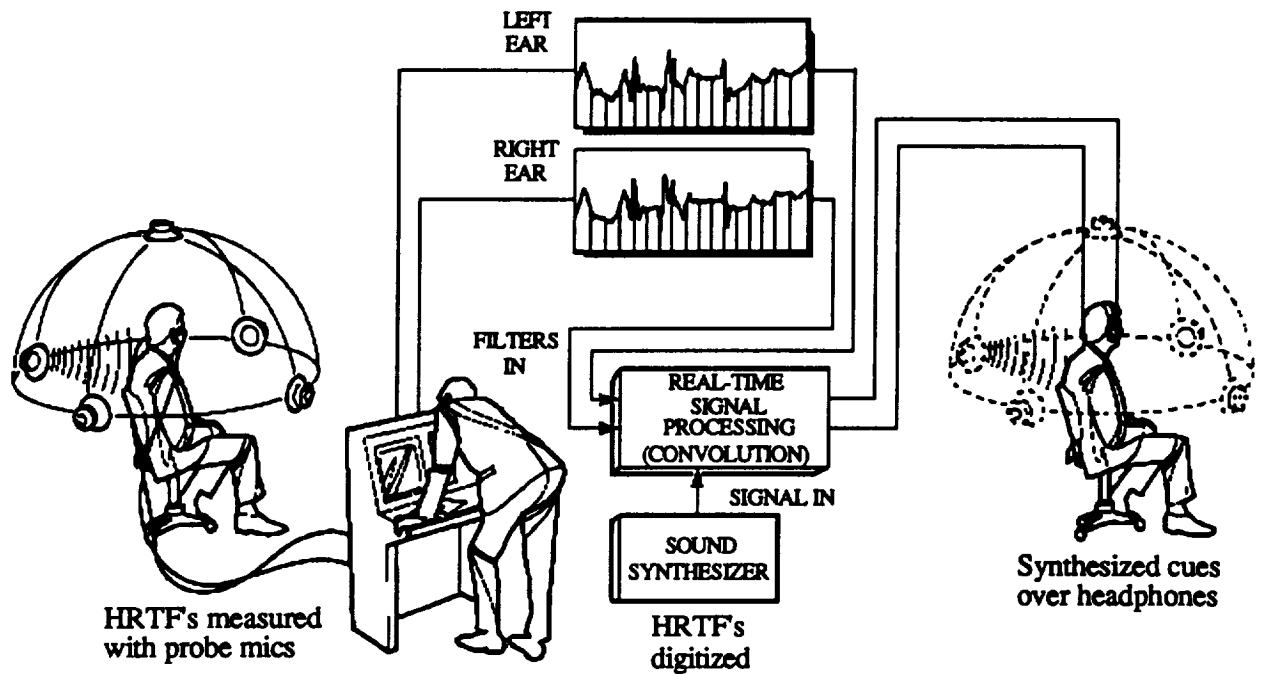
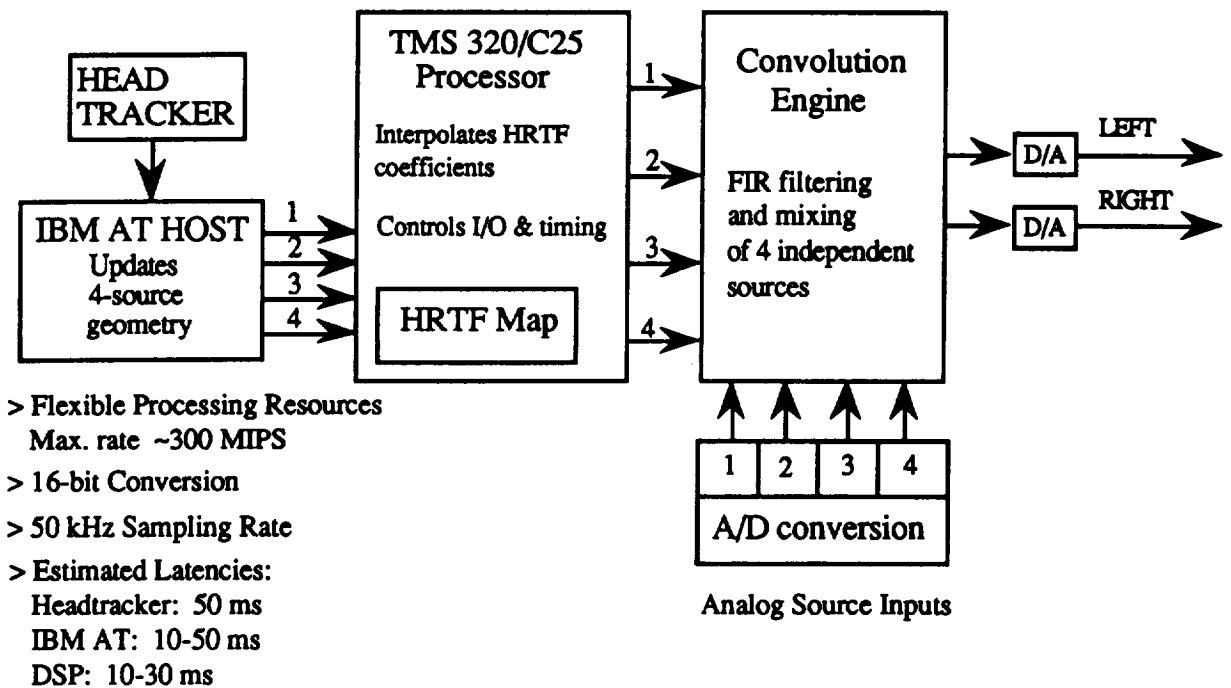**Figure 1: 3D Auditory Display; Synthesis Technique**

Fig. 2:   The Convolvotron

Implementing Three-Dimensional Sound

Localized acoustic cues could be realized with an array of real sound sources or loudspeakers [9], [13]. An alternative approach, recently developed at NASA-Ames Research Center, generates externalized, three-dimensional sound cues over headphones in real time using digital means [32], [33]. This type of presentation system is desirable because it allows complete control over the acoustic waveforms delivered to the two ears and the ability to interact dynamically with the virtual display. The synthesis technique, illustrated in Figure 1, involves the digital generation of stimuli using Head-Related Transfer Functions (HRTFs) measured in the ear-canals of individual subjects (see [36], [3]). The advantage of this technique is that it preserves all of the interaural temporal and level differences over the entire spectrum of the stimulus, thus capturing the effects of filtering by the pinnae which are critical for the veridical simulation of externalized sound sources.

In the real time system, the Convolvotron, up to four moving or static sources can be simulated in a head-stable environment by digital filtering of arbitrary signals with the appropriate HRTFs. Motion trajectories and static locations at greater resolutions than the empirical data are simulated by linear interpolation of the four nearest measured transforms. Also, a simple distance cue is provided via real time scaling of amplitude. Figure 2 shows the functional components of the Convolvotron system designed by Scott Foster.

Such an interface not only requires the development of special-purpose display technology, it also necessitates the careful psychophysical evaluation of listeners' ability to accurately localize the virtual or synthetic sound sources. The working assumption of the synthesis technique is that if, using headphones, one can produce ear-canal waveforms identical to those produced by a free-field source, the free-field experience will be duplicated. A recent study [36] confirmed the perceptual adequacy of the basic technique for static sources for experienced subjects localizing stimuli in the free-field compared with stimuli synthesized from their own HRTFs. Source
azimuth was synthesized nearly perfectly for all listeners while synthesis of source elevation was less well-defined, e.g., more variable with a compressed range of responses. Elevation was also the source of the most obvious individual differences in localization for both free-field and synthesized signals.

Unfortunately, measurement of each potential listener's HRTFs may not be possible in practice. It may also be the case that the user will not have the opportunity for extensive training. Thus, a critical research issue for virtual acoustic displays is the degree to which the general population of listeners can obtain adequate localization cues from stimuli based on non-individualized transforms. Preliminary data [34] from
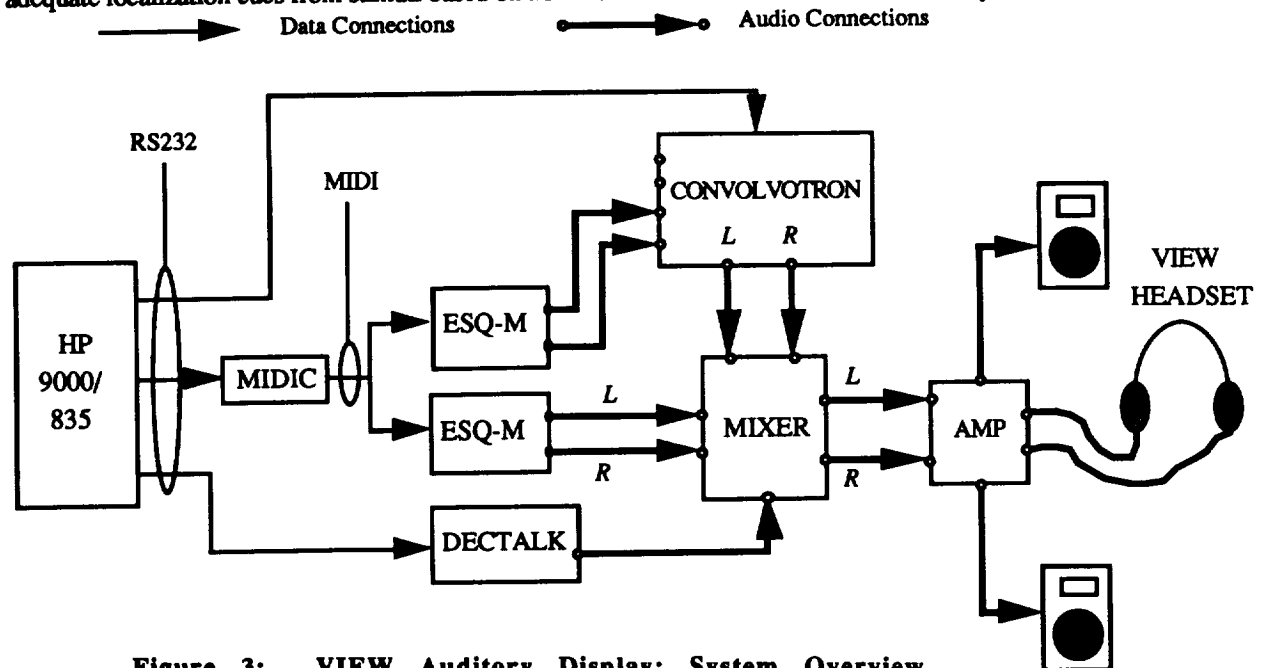

Figure 3:   VIEW Auditory Display; System Overview

three experienced subjects suggest that using non-listener-specific transforms to achieve synthesis of localized cues is at least feasible. Localization performance was only somewhat degraded compared to a subject's inherent ability, even for the less robust elevation cues, as long as the transforms are derived from what one might call a "good" localizer. Further, the fact that individual differences in performance, particularly for elevation, can be traced to acoustical idiosyncrasies in the HRTF spectra, suggests that it may eventually be possible to create a set of "universal transforms" by parametric modeling techniques (e.g., [24]), principal components analysis, or perhaps even enhancing the spectra of empirically-derived transfer functions (e.g., [14]).

VIEW Sound System Architecture

While perceptual studies of individual sensory modalities are clearly needed, it is also important to examine the role of sensory interaction. NASA-Ames' VIEW system provides the opportunity to implement localized auditory icons and assess their contribution to an integrated spatial display. Briefly, VIEW is a multisensory display environment which allows the user to explore and interact with a 360-degree synthesized, or remotely-sensed, world using a head-mounted, wide-angle stereoscopic display controlled by operator position, voice, and gesture. More detailed descriptions of the VIEW visual and gestural displays can be found in [17] - [19].

The VIEW auditory display subsystem allows audio cues responsive to both discrete events and continuous data changes to be designed and linked to arbitrary events and data flows in VIEW scenarios. Refer to the system overview diagram, Figure 3, for the following discussion.

Development of the initial binaural display capability based on MIDI (Musical Instrument Digital Interface) sound-synthesis technology began in 1987. More recently, true spatial cueing was added to the system with the integration of the Convolvotron. The auditory display subsystem, like most of the VIEW system, is currently implemented with a Hewlett Packard HP9000/835 computer. Two Ensoniq ESQ-M synthesizer modules handle the actual production of audio cues, supplemented with a Digital Equipment Corporation DECTalk speech synthesizer. MIDI protocol is used to communicate between the HP host and the ESQ synthesizers. A Hinton Instruments MIDIC interface converts 19.2 Kbps RS232 signals from the HP into 31.25 Kbps 5 mA. current-loop signals that are required by the MIDI standard.

Each ESQ synthesizer has two outputs. One ESQ's output pair is mapped directly into the VIEW system's left and right audio channels. Up to eight independent (polytimbral) voices, mixed to the stereo output, may be played through this synthesizer. The second ESQ's output pair is patched into the Convolvotron. As described above, this device is capable of synthesizing, in real time, an apparent three-dimensional location for up to four independent audio inputs. In this case, since only two channels of sound are available from the ESQ, only two of the Convolvotron's inputs are used. The simple stereo pair from the first ESQ, and the 3D-imaged output from the Convolvotron, are mixed, amplified, and sent to headphones integrated into the VIEW headset, or optionally, to room speakers.

The central software component of the auditory display is the cue driver. As in all VIEW applications, the software is written in C in a Unix environment. Without delving too deeply into its implementation details, it can be described as consisting of an event scheduler, a MIDI, speech, and Convolvotron event generator, and a VIEW/Auditory Display rendezvous mechanism. It also handles several housekeeping chores, such as loading cue files, initializing the MIDI interface, the synthesizers and the Convolvotron, downloading patch files to the synthesizers, and making sure all is quiet before a VIEW scenario exits.

Up to ten (monophonic) auditory "objects" may be displayed simultaneously. If a more complex, or polyphonic, sound is desired, several voices can be assigned to the same icon and the number of possible simultaneous objects is reduced accordingly. In general, the basic sound signature or identity of an individual object or "icon" derives from the particular ESQ patch assigned to it. Since this technology was developed for music synthesis, one can often think of a patch as having the attributes of a particular musical instrument. However, some "environmental" sounds analogous to sound effects, e.g., footsteps or explosions, are also possible.

Custom ESQ patches may be designed off-line using the front-panel capabilities of the synthesizer, a patch editor/librarian software package, or by selecting from collections of commercially-available pre-

programmed patches. Icons may take advantage of one or more of the controllable parameters made available by the ESQ. These include oscillator frequency, filter cutoff frequency, amplitude level, and stereo pan position. Any of these can be modulated in real time and associated with events or information flow in a VIEW display scenario. One of the advantages of the ESQ, and the main reason for its choice during the specification of this system, is that it allows access to these parameters through standard MIDI controllers; many synthesizers require the use of "system-exclusive" messages to achieve this level of control. Designing the cue driver around standard MIDI controllers makes it less system-dependent; as more synthesizers adopt this level of control (and this seems to be the trend), the auditory display may be readily adapted to them.

Because of the limited outputs of the ESQ (two per synthesizer), only two of the ten icons may be assigned specific locations via the Convolvotron at any one time. Alternatively, using all four ESQ outputs, up to four simultaneous icons could be independently localized. The current configuration was a compromise solution which traded localized cues for an increase in the number of possible icons. A future solution would be to adapt the system to a synthesizer which has independent outputs for each voice. Also, integrating a digital-sampling device would be useful for presenting the kinds of sounds that Gaver [23] advocates in his notion of "everyday listening." At the time we began developing the system, digital samplers tended to be expensive and allowed very little real time control over the acoustic parameters of the sounds. Since a major goal of the display was to allow continuous control over the icons' acoustic structure, we opted for a more standard and inexpensive synthesizer with a relatively well-developed MIDI implementation. [See [8] for a useful discussion of the pros and cons of various MIDI devices.]

Editing & Display Capabilities

Cues or icons are designed and refined with ACE, the Auditory Cue Editor. ACE is a stand-alone program, which makes it unnecessary to activate the entire VIEW system merely to work on auditory cues. Multi-level menus, interactive prompting, and extensive syntax checking aid the user in designing complex auditory cues with relative ease.

ACE is composed of four basic sections, organized as independent screens, each with its own menu of commands.

On the Main Screen, cues may be created, deleted, loaded, saved, and named (an important function; all rendezvous between VIEW events and data are made through the names of cues as specified on the Main Screen). The Main Screen may also be used to specify certain basic parameters, e.g., synthesizer patch number and localization method (convolved or simple stereo). In addition, a "play" command allows quick, interactive audition of cues during their design.

The Sound Event Editing Screen allows the construction of the main body of a cue, which is in the form of a list of time-ordered MIDI, speech, or Convolvotron localization commands. While entire pieces of music could theoretically be entered here, note by painstaking note, this is usually not the case. Single notes, chords, or short sequences of notes and chords are the most common items entered on this screen. This is because very simple events (with carefully chosen and distinct timbres) are often all that is needed for basic cueing functions. Even the more complex auditory icons generally have fairly simple event lists, since most higher-level display capabilities result from linkage to continuous data streams and response to real-time changes of those data streams.

The Modulation Editing Screen enables the connection and scaling of incoming data streams to several different MIDI modulators as well as Convolvotron sound-source position coordinates. This makes it possible to have an arbitrary data value produced by the VIEW system displayed as a proportional deflection in a variety of auditory parameters, such as pitch, timbre, or apparent three-dimensional location. The data structure has provisions for incorporating nonlinear mappings between incoming data and an auditory parameter, a feature which can be very important in evaluating the perceptual consequences of a particular icon. Specifications of modulation behavior composed on this screen are given textual names, which enables rendezvous with the appropriate VIEW-generated data streams.

Patches, as mentioned above, are a very important part of an auditory icon; they define the basic sound that the synthesizer produces, and the manner in which it will react to incoming controls sent by the cue driver.

The Patch Maintenance Screen allows the uploading and downloading of individual patches and complete patch banks between the host system and the synthesizers. In this way, a particular set of sound programs can be directly associated with a set of auditory cue definitions.

Application in the Virtual Environment:
Telerobotic Control

The VIEW telerobotic scenario was designed to illustrate the capability of telepresence, i.e., the manipulation of objects or interaction with persons or objects remote from one's location, that a virtual environment makes possible. In this scenario, the visual and kinematic characteristics of a Puma robotic arm are modeled with a high degree of accuracy. The scenario participant may, upon donning the VIEW stereoscopic head-mounted display, align his or her arm with that of the lifesize model. The participant's arm and the modeled robotic arm may then be "coupled," which simply means that the robotic arm will move, to the extent of its kinematic capabilities, in correspondence to the movement of the participant's arm. During the coupled mode, an end-effector with a vise-like gripping apparatus may be opened and closed merely by opening and closing the hand.

This graphic computational model of a robotic arm is meant to test the efficacy of the intuitive mapping of control between machine and human counterpart. The success of this mapping is tested by assigning a simple task to be completed by the telerobotic participant. In the foreground of the scenario, a "circuit card" is plugged into a slot on a "task board." The participant is instructed to remove the circuit card and replace it with another one which is just off to the side of the task board. This entails coupling with the robotic arm, maneuvering the end-effector into position so that the two jaws surround the edge of the circuit card, closing the end-effector jaws around the card to grasp it, and pulling it out and away from the task board. Once this is completed, the replacement circuit card must be grasped in a similar manner, lined up exactly, and inserted into the slot.

With perfect telepresence, this task could be accomplished with little more difficulty than if one were using one's own hand and a real task board and circuit card. However, factors such as slower-than-ideal graphic refresh rates, lower-than-ideal contrast and focus in the VIEW display, etc., conspire to make the precision manipulation required somewhat difficult. In a situation such as this, auditory feedback can make an important difference, particularly with the current paucity of good haptic or force feedback display systems.

At the simplest level, auditory feedback is used to indicate the occurrence of discrete events in the scenario. For example, many commands and actions in VIEW are initiated by hand gesture. A VPL "Data Glove" reports finger positions to the host computer, which examines those positions for correspondence to any of several pre-defined gestures, such as "single-finger point," or "fist." When one of these gestures is detected by the host, a sound is made by the auditory display to indicate gesture recognition.

Other simple auditory cues fall into the category of "reality-mimicry," or sound effects. In the telerobotic scenario, bumping an end-effector into a "solid" object in the virtual (or teleoperated) world causes a "bump" sound to be produced. Since direct force feedback is not yet available in the VIEW system, this form of audio display is particularly critical, as it warns of a situation which could cause damage to a real-world robotic arm or to objects with which it is colliding. At a more mundane level, this sort of sound effect enhances the sense of presence; objects tend to make a sound when they collide in the real world, so it is reasonable to expect them to do so in a virtual environment.

Audio feedback that supplements or replaces force feedback is not limited to mimicry of collision sounds or similar sound effects. Force can be represented as a continuum by changing one or more sound parameters in correspondence to the force's intensity. This type of display is utilized for a special circumstance in the telerobotic scenario. If the scenario participant attempts to force the replacement circuit card into the task board without orienting it correctly, a force-reflection display called "push-through" is initiated. It starts out as a soft, steady tone that gets louder, brighter (higher harmonics are let through the filter), and more frequency modulated the harder the participant pushes on the misaligned card. In this way, a potentially damaging increase in user input is signalled by an increasingly harsh and strident auditory warning.

Taking this idea one step further, not only force, but any arbitrary continua of data may be displayed. Perhaps the most successful use of auditory feedback in the telerobotic scenario comes into play while the participant attempts to guide the replacement circuit board into the target socket. As the board reaches a certain proximity to the socket, a cue initiates consisting of two sustained tones; the pitch of one of the tones is deflected with respect to the other by an amount proportional to the distance between the circuit card and its slot. As the card nears the slot, the two pitches come closer together (which is readily perceived due to the obvious decrease in the beat frequency produced by the increasingly adjacent pitches); at distance zero, the tones are in unison. The cue functions as an auditory "rangefinder," and greatly facilitates the proper positioning of the card in its slot.

Card orientation, which is also crucial to the completion of the replacement task, could be represented by some other continuous audio parameter, such as depth of frequency modulation. With careful selection and scaling of the modulated sound parameters, two continua (e.g., proximity and orientation) could be monitored simultaneously at a very intuitive level. Our work to date has not explored multiple-simultaneous displays of continuous data, but it is an intriguing area for further research.

Auditory Design Principles

The telerobotic scenario has served as an excellent test of the capabilities of the auditory display. While formal experimentation has yet to be done, it has also provided a rich environment for the discovery of certain basic guidelines for the design of auditory icons and the development of an auditory symbology.

Practical experience has shown that the most effective cues are simple cues. Long sequences or elaborate clusters of tones not only tend to clutter the auditory display, they can increase the load on cognitive processing and memory required to interpret the information, and in the long run, become downright annoying. Imagine a telephone that played "Three Blind Mice" every time a call came in. It would be only a short time before the exactly-repeating melody became maddening. The simple bell or digital chirp of a telephone manages to get attention without engaging the "music critic" part of one's cognition. Similarly, a "thud" sound suffices to signal a bump in a virtual world; it is not necessary to have a speech synthesizer say "You have bumped into something" at each and every collision. While these may be extreme examples, the basic principle holds; an auditory icon should be as simple as possible.
The need for simplicity is even more critical when several cues occur in close proximity to each other. For instance, in the telerobotic scenario, a gesture-recognition cue might be followed immediately by a sound that indicates movement of the jaws of the end-effector. If the jaws were then to close over the circuit board, a "board-grasped" cue would result. These three cues can occur in rapid succession, so they must be of short duration for the correct sequence of events to be properly represented.

This situation also points out the need for carefully choosing the sound signatures or patches which form the fundamental units of an auditory symbology. Patch design, including spectral content, amplitude and filter envelopes and various special effects, is the chief distinguishing feature of a simple icon. The best way to make an icon recognizable is to give it a distinctive sound. Much effort in the design of auditory icons is therefore concentrated in selecting or building an appropriate synthesizer patch.

As noted before, guidelines can be derived from the fields of psychoacoustics, music, and perceptual psychology. As illustrated in the proximity cue in the telerobot scenario, the close tuning of two pitches is a continuous parameter to which the human ear is very sensitive. However, the amplitude modulation or beat frequency which signals the change in proximity will only occur for a limited range of frequencies which must be considered when mapping the distance data to the difference in pitch. In developing a symbology, one can also take advantage of what one might think of as "natural" or metaphorical mappings even if a literal sound, such as a "bump," is not possible. For example, the "push-through" cue described above clearly signals an increasing violation of the allowable forward movement when inserting the task board at an incorrect orientation by a harsh sound which increasingly "violates" the ears. To minimize cognitive effort, it is important to build meaning by the relationships between icons as well. In the telerobot scenario, icons which provide feedback for related gestures have similar timbres that are distinguished by their temporal structure. For example, larger changes in pitch, at the level of short sequences, are used (much like the familiar two-chime doorbell). This particular type of icon has the virtue of being reversible like a short musical motive; a "grasp" gesture is represented by a high note followed by a slightly lower note. The complementary "release" gesture is the same two notes, only in reverse order.

17

As much as possible, this relationship between sound and meaning should remain consistent throughout a display system. Thus, in VIEW, the cues which provide feedback for the various gestures remain the same across the different types of display scenarios that have been developed.

Careful consideration of the possible interactions between icons will be particularly important when auditory cues must be presented simultaneously as in the combination of orientation and proximity cueing described above. Principles of acoustic perceptual organization, such as the Gestalt principles elaborated by Bregman [6], will provide important guidelines. For example, different acoustic objects may be defined by different auditory streams. Streaming is determined by such features as frequency separation, timbre, rate or tempo, spatial location, and "common fate" or the tendency of spectral components to be grouped according to similar frequency or temporal patterns.

Other VIEW Scenarios

Other examples of display scenarios which have been implemented in the VIEW system include the Extra-Vehicular Activity (EVA) Visor and a Computational Fluid Dynamics (CFD) data visualization. The EVA Visor is a concept for a helmet-mounted, three-dimensional dataspace which can be accessed by an astronaut during repair or inspection activities while outside the space station. In the scenario, several types of display windows can be used, including life-support system status, a "cuff checklist" of tasks to be completed, repair schematics, and a three-dimensional "map cube," which represents the entire EVA scenario as a miniature, manipulable cube, with which the astronaut can establish his or her position and orientation with respect to the other vehicles or objects in the scenario. Currently, auditory cues for the EVA Visor include a set of gesture recognition cues identical to those in the telerobot scenario (in keeping with the principle of transference of a learned auditory vocabulary between virtual worlds, when possible). Warning cues signal situations which might endanger the astronaut's safety, such as impending depletion of life support resources. A special sound effect cue indicates when MMU (the rocket-backpack vehicle which enables the astronaut to maneuver during EVA) thrusters are firing. Finally, activation of the various windows listed above is heralded by corresponding audio signals. Another cue, not yet implemented, which would be useful in the EVA display, is an orientation beacon which allows the astronaut to continuously monitor the location of the space station by means of a localized auditory icon to minimize disorientation in the absence of visual and gravitational referents.

The CFD data display visualizes the fuel-flow around the LOX (liquid oxygen) post of the main shuttle engine. Features include the ability to "fly through" the data, viewing it from different viewpoints including inside the fluid flow, and "grabbing" and scaling the data up or down to examine its finer or coarser features. Although not yet implemented, a potentially useful auditory visualization cue might be to "attach" auditory icons to one or more particles in the flow, and thus follow their progress as they interact with the structures of the shuttle engine.

In developing the VIEW auditory display, we have attempted to provide a flexible and general-purpose system which takes advantage of our knowledge of perceptual abilities as much as possible. The hardware architecture and software is designed to be applicable to a wide variety of display configurations and to allow a consistent approach to the design of auditory symbologies based on knowledge gleaned from music, psychoacoustics, and perceptual psychology.

References

[1] Begault, D.R. & Wenzel, E.M. (1990) Techniques and applications for binaural sound manipulation in man-machine interfaces. NASA Technical Memorandum No. TM102279, In Press.

[2] Blattner, M.M., Sumikawa, D.A., & Greenberg, R.M. (1989) Earcons and icons: Their structure and common design principles. Hum.-Comp. Interact., 4, 11-44.

[3] Blauert, J. (1983) Spatial Hearing. The MIT Press: Cambridge, MA.

[4] Bly, S. (1982) Sound and computer information presentation. Unpublished doctoral thesis (UCRL-53282) Lawrence Livermore National Laboratory and University of California, Davis, CA.

[5] Brooks, F.P. (1988) Grasping reality through illusion -- Interactive graphics serving science. Proc. CHI'88, ACM Conf. Hum. Fac. Comp. Sys., Washington, D.C., 1-11.

[6] Bregman, A. (1981) Asking the "what for" question in auditory perception. In Kubovy & Pomerantz (Eds.), Perceptual Organization,Lawrence Erlbaum Associates: Hillsdale, NJ.

[7] Brown, M., Newsome, S., & Glinert, E. (1989) An experiment into the use of auditory cues to reduce visual workload. Proceedings of CHI'89, ACM Conference on Human Factors in Computing Systems, 339-346.

[8] Buxton, W., Gaver, W., & Bly, S. (1989) The use of non-speech audio at the interface. Tutorial #10, CHI'89, ACM Press: New York.

[9] Calhoun, G.L., Valencia, G., & Furness, T.A. III (1987) Three-dimensional auditory cue simulation for crew station design/evaluation. Proc. Hum. Fac. Soc., 31, 1398-1402.

[10] Cherry, E.C. (1953) Some experiments on the recognition of speech with one and two ears. J. Acoust. Soc. Am., 22, 61-62.

[11] Colquhoun, W.P. (1975) Evaluation of auditory, visual, and dual-mode displays for prolonged sonar monitoring in repeated sessions. Hum. Fac., 17, 425-437.

[12] Deatherage, B.H. (1972) Auditory and other sensory forms of information presentation. In H.P. Van Cott & R.G. Kincade (Eds.), Human Engineering Guide to Equipment Design, (rev. ed.), Washington, DC: U.S. Government Printing Office, 123-160.

[13] Doll, T.J., Gerth, J.M., Engelman, W.R. & Folds, D.J. (1986) Development of simulated directional audio for cockpit applications. USAF Report No. AAMRL-TR-86-014.

[14] Durlach, N.I. & Pang, X.D. (1986) Interaural magnification. J. Acoust. Soc. Am., 80, 1849-1850.

[15] Edwards, A.D.N. (1989) Soundtrack: An auditory interface for blind users. Hum. Comp. Interact., 4, 45-66.

[16] Egan, J.P., Carterette, E.C., & Thwing, E.J. (1954) Some factors affecting multichannel listening. J. Acoust. Soc. Am., 26, 774-782.

[17] Fisher, S.S. (1986) Telepresence master glove controller for dexterous robotic end-effectors. Advances in Intelligent Robotics Systems, D.P. Casasent (Ed.), Proc. SPIE, 726.

[18] Fisher, S.S., McGreevy, M.W., Humphries, J., & Robinett, W. (1986) Virtual environment display system. ACM Workshop on Interactive 3D Graphics, Chapel Hill, NC.

[19] Fisher, S.S., Wenzel, E.M., Coler, C. & McGreevy, M.W. (1988) Virtual interface environment workstations. Proc. Hum. Fac. Soc., 32, 91-95.

[20] Foley, J.D. (1987) Interfaces for advanced computing. Sci. Amer., 257, 126-135.

[21] Furness, T.A. (1986) The super cockpit and its human factors challenges. Proc. Hum. Fac. Soc., 1986 (1), 48-52.

[22] Garner, W.R. (1949) Auditory signals. In A Survay Report on Human Factors in Undersea Warfare, Washington, D.C.: National Research Council, 201-217.

[23] Gaver, W. (1986) Auditory icons: Using sound in computer interfaces. Hum.-Comp. Interact., 2, 167-177.

[24] Genuit, K. (1986) A description of the human outer ear transfer function by elements of communication theory. Proc. 12th ICA (Toronto),Paper B6-8.

[25] Loomis, J.M., Hebert, C., & Cicinelli, J.G. (1990) Active localization of virtual sound sources. Submitted to J. Acoust. Soc. Am.

[26] O'Leary, A. & Rhodes, G. (1984) Cross-modal effects on visual and auditory object perception. Perc. & Psychophys., 35, 565-569.

[27] Patterson, R.R. (1982) Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority Paper No. 82017, London.

[28] Smith, S., Bergeron, R.D., & Grinstein, G.G. (1990) Stereophonic and surface sound generation for exploratory data analysis. Proceedings of CHI'90, ACM Conference on Human Factors in Computing Systems, 125-132.

[29] Sorkin, R.D., Wightman, F.L., Kistler, D.J., & Elvers, G.C. (1989) An exploratory study of the use of movement-correlated cues in an auditory heads-up display. Hum. Fact., 31, 161-166.

[30] Sutherland, I.E. (1968) Head-mounted three-dimensional display. Proc. Fall Joint Comp. Conf., 33, 757-764.

[31] Warren, D.H., Welch, R., & McCarthy, T.J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. Perc. & Psychophys., 30, 557-564.

[32] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988a) Development of a three-dimensional auditory display system. SIGCHI Bulletin,20, 52-57.

[33] Wenzel, E.M., Wightman, F.L., & Foster, S.H. (1988b) A virtual display system for conveying three-dimensional acoustic information. Proc. Hum. Fac. Soc., 32, 86-90.

[34] Wenzel, E.M., Wightman, F.L., Kistler, D.J., & Foster, S.H. (1988c) Acoustic origins of individual differences in sound localization behavior. J. Acoust. Soc. Amer., 84, S79.

[35] Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis. J. Acoust. Soc. Amer., 85, 858-867.

[36] Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation. J. Acoust. Soc. Amer., 85, 868-878