A Spatial Auditory Display for the CyberStage

Gerhard Eckel

GMD - German National Research Center for Information Technology
Sankt Augustin, Germany

Abstract

The CyberStage is GMD's CAVE-like audio-visual projection system integrating a 4-side visual stereo display and an 8-channel spatial auditory display. A software-based sound server for the generation of auditory cues for interactive virtual environments has been developed for this display system in the context of a research project on integrated simulation of image and sound (ISIS). Hardware and software components of the auditory display and their integration in the CyberStage application development process are described. Four applications from different areas are discussed as examples.

1 Introduction

We report about our approach towards generating auditory cues for interactive virtual environments. This approach was motivated by the need to integrate a loudspeaker-based auditory display with a projection-based visual display to form GMD's projection system CyberStage [1], an improved version of the CAVE [2].

1.1 The CyberStage

The design of the CyberStage was aiming at providing for high degrees of immersion and presence for a wide spectrum of virtual environments including the following application areas:

- scientific visualization/sonification
- industrial data exploration
- product presentation and marketing
- product design and evaluation
- architectural planning and walk through
- exhibition design and evaluation
- media art installations
- promotional infotainment

The immersive features of the CyberStage are based on 4-side stereo image projection and 8-channel spatial sound projection both controlled by the position of the user's head followed by a tracking system. The sound projection is completed by vibration emitters built into the floor of the system and allowing for rendering low frequency signals perceivable through feet and legs. To create the illusion of presence in virtual spaces, the CyberStage system provides various interfaces and interaction metaphors to visually and acoustically respond to the users' actions. These interfaces allow for navigation in virtual spaces and manipulation of virtual objects. The software driving the CyberStage - the Avocado [3] application development toolkit - supports a wide range of output devices (e.g. computer monitor, CyberStage, Responsive Workbench [4] or Teleport [5]) and allows connecting these devices for distributed applications.

1.2 Integrated Simulation of Image and Sound

Since image and sound tend to mutually enhance each other by playing complementary roles in immersing the user in a virtual environment, they have to be regarded as equally important in the simulation. Therefore, the simulation process generating the visual and auditory stimuli has to take into account the visual and the auditory scene as well as their relationships at a time. To study the problems arising from this requirement, the GMD started a research project on the integrated simulation of image and sound (ISIS) in 1996. The first result of the ISIS project was the design of the hardware and software components of the CyberStage auditory display in 1996, which served as a reference system and testbed for ongoing research. In 1997, an improved version of the auditory display has been completed. Since then, the CyberStage has been used successfully by many of our industrial partners in research and marketing applications.

2 CyberStage Auditory Display

2.1 Requirements

The auditory display of the CyberStage is required to produce a sound field creating the illusion of sound reaching us from any direction and distance and carrying the signature of the acoustic environment in which it was propagated. The first aspect (localization) is important for navigation and orientation and hence contributes mainly to the illusion of presence. The second aspect (reverberation) conveys important information about the environment shared by the sound source and the observer (e.g. inside or outside space, room size) and therefore enhances essentially the degree of immersion. A third aspect concerns the radiation pattern of a sound source. Many sound sources do not emit sound omni-directionally but in a potentially complicated and frequency dependent pattern. Our auditory display is only required to simulate omni-directional and simple directed sound sources. For the CyberStage, it is sufficient to be able to distinguish if a directed sound source is pointed towards the observer or in the opposite direction. In an application all these aspects my vary dynamically since the observer and the sound sources may move with respect to each other and with respect to the spaces in which they are located.

In contrast to visual scene simulation, auditory scene simulation is always event driven. Sound is always the consequence of some kind of event happening. If the event in question is also part of the visual environment (e.g. two objects colliding), then it is required that the visual and the auditive cues are presented coherently, i.e. that the visual and auditive aspects of the event are simulated at the same location and at the same time (which doesn't mean that they are seen and heard at the same time, since sound needs much more time to propagate than light). Because of the complexity of real-time visual and auditive rendering processes and their control, this requirement is hard to meet.

Another important requirement concerns sound modeling and synthesis. If a sound has to be produced as a consequence of an event, then the characteristics of the sound have to follow those of the event. In the case of colliding objects, the material of both objects, the impact position, and the speed of the objects may determine how the collision sounds. Such flexibility can usually not be reached when using sampled sound material. This is why for a system like the CyberStage direct sound synthesis is required (besides sound sampling), which allows to model whole classes of sounds. Instances are synthesized when needed and according to a set of event-dependent parameters.

2.2 Sound Projection

While the first generation CyberStage system (1996) used only 4 loudspeakers providing 2d localization, the second generation (1997) uses 8 loudspeakers in the classical cube configuration for full 3d localization. The upper four loudspeakers are mounted directly above the four corners of the cubical projection screen frame (3 x 3 x 3 m). The two lower speakers in the back are mounted on the floor at the lower corners of the projection frame. A special arrangement had to be made for the two lower speakers in the front, as they have to be positioned behind the projection screens in order not to occlude the visual projection. This is one of the most critical acoustical design issues of surround-view projection systems, which has not been solved in a satisfactory way yet. This is why we put much effort in compensating the frequency-dependent damping caused by the screen. Acoustic measurement have been undertaken to estimate the compensation filter responses. The acoustics of the room behind the front projection screen has been adapted to avoid reflections and sound paths other than through the front projection screen. All eight loudspeakers are oriented towards a point in the center of the projections system at average ear level. Additionally to the eight speakers there are four low frequency vibration emitters installed in the floor of the CyberStage. These are used to generate vibrations of up to around 100 Hz, which can be perceived directly through the body via feet and legs. Unlocalizable high-amplitude low-frequency components (e.g. collision shocks) are presented that way.

2.3 CyberStage Sound Server

The eight loudspeakers and the four vibration emitters are fed by the CyberStage Sound Server (CSS) designed in the context of the ISIS project. In order to meet the requirements described above, the CSS architecture has to be as open as possible. Therefore, we decided to build the server entirely in software and minimize the amount of external hardware needed. The CSS runs on any SGI machine and supports all kinds of audio interfaces including 8-channel ADAT compatible sound output devices, which connect via a fiber optic link to external D/A converters. The CSS automatically detects the hardware configuration of the machine it is running on and configures it for the best output configuration possible. The proper rendering modules (for 2, 4, or 8 channel output) are chosen accordingly. The vibration emitters are fed by an extra pair of analogue outputs (if available) and are passed though an external low-pass filter with a dynamic limiter before being sent to power amplifiers.

All rendering modules use intensity panning for directional encoding. Discrete surround encoding was chosen instead of Ambisonics [6] (i.e. B-format) encoding because it produces better localization results under the very particular acoustic conditions in the CyberStage (reflections from projection mirrors and screens).

The sound server is based on IRCAM's Max/FTS real-time sound processing system [7, 8] originally built for computer music applications. FTS is an extensible signal-processing kernel providing all necessary low-level modules to build sophisticated sound synthesis and processing applications. Max is a graphical programming environment used to interactively build FTS programs. Max allows controlling and monitoring the state of a signal-processing program running in FTS. The spatialization algorithms used by the CSS are partly based on IRCAM's Spatialisateur toolkit [9] developed in Max/FTS. The software built on top of this consists of parts realized in Max (synthesis control, resource management system, message parsing) and FTS extensions written in C (efficient spatialization modules, sound sample manager, custom synthesis algorithms, network communication). The CSS is not a closed application but an open toolkit adapted to a large class of applications. The application designer chooses among many templates provided by the server to solve standard problems.

3 CyberStage Application Design

CyberStage applications are built with the Avocado application development toolkit for distributed virtual environments [3] developed at GMD. Avocado uses the IRIX Performer library for rendering the four stereo images and the CSS for rendering the sound field. Visual rendering is performed on a SGI Onyx2 computer with 12 R10k processors and 4 Infinite Reality graphics pipes. Avocado handles the user interface devices, manages the scene graph, and communicates with CSS running on another R10k SGI machine over a dedicated Ethernet link using the UDP protocol for low-latency and low-jitter communication (if desired, the CSS can also run on one of the processors of the Onyx2). The Avocado scene graph organizes the visual and auditive elements of the scene and describes their behavior in time and space. Avocado provides a set of scene graph node classes, which define auditory scene elements. There are different nodes which describe the nature of the sound source (e.g. sample or sound model), the radiation pattern (e.g. directed or omni-directional), the mapping of event to synthesis parameters (e.g. intensity of impact to spectrum and amplitude), the rendering resolution (e.g. dynamic or static spatialization), and the characteristics of the acoustic environment (e.g. reverberation time and damping). These nodes communicate with the sound server to invoke and control the corresponding sound synthesis and rendering processes. The application designers have the freedom to use the existing classes and their counterparts in the sound server or develop new nodes along with new FTS synthesis and spatialization modules. This feature is essential in a research and experimentation context.

4 Examples of Application

We described a new and quickly evolving system, which already meets most of the requirements, stated above but which is still under development. Nevertheless, many applications have been realized with the CSS and the CyberStage already, four of which we will briefly describe now.

4.1 Vector Field Visualization and Sonification

A virtual environment for the exploration of vector field data from airflow simulations of a car air conditioning system has been developed for Daimler Benz. In a real-size model of the interior of a car, the spatially distributed static vector field of air velocity data is visualized by particles injected into the field, which then float along the streamlines. These particles may also leave traces following the streamlines. This conventional approach was refined by sonifying the velocity data with low-pass filtered noise portraying wind-like sounds. The velocity data was mapped on the noise amplitude and the cutoff frequency of the filter. For each particle or streamline, the additional noise sonification can be activated individually. This turned out to significantly improve the detection of velocity changes, which are much harder to detect from the visual movement. To complement the visual exploration, a virtual audio probe was developed which can be placed into the vector field to interactively sonify the velocity value at a certain point in the field. The probe is moved around in the virtual environment by means of a stylus-shaped position sensor. The real success of the combined sonification and visualization in this application came from this audio probe, which allows for a much more intuitive and efficient exploration of certain "hot spots" in the vector field than the particle and streamline visualization. We found that this application is a good example for the complementary role sonification can play in visualization applications.

4.2 Development of an Innovative Museum Guide

Together with the Kunstmuseum in Bonn, we develop a new type of audio guide. The project aims at creating augmented environments by seamlessly integrating virtual auditory spaces with existing physical spaces such as exhibition rooms. The main purpose of this type of audio augmented environments is to provide visitors with situated and individual acoustic information without confining their deliberate movements in space. This is achieved by making the virtual auditory environment react intelligently to the movements of the visitors. Wireless head tracking combined with binaural auditory display on wireless headphones is the technique used to supply the visitors with an advanced sense of acoustic immersion. Together with a model of the physical environment, the tracking data is used to analyze and interpret the movements of the visitors. Distance and orientation with respect to interesting objects in the physical environment are used to compute the acoustically mediated reactions. Localizable virtual sound sources can thus be attached to visible objects. Observing the movements of visitors over a longer period of time allows drawing conclusions regarding their preferences, which can then be taken into account in the composition of the acoustic information to be presented. Very elaborate and intuitive interaction metaphors may be developed with this approach due to the refined reaction possibilities of the virtual soundscape.

In the first phase of the project a test scenario is simulated in the CyberStage to experiment with different technical and artistic solutions, develop the necessary authoring software, and convince industrial partners. The museum spaces have been modeled and reproductions of the artwork have been scanned to reproduce a particular exhibition for which a virtual acoustic environment was composed. Once this phase is completed and the remaining technical problems are solved, the real installation will be installed in the museum. The CyberStage will then still be used as authoring tool and the already realized environments can be transferred to the museum installation, which will also use the CSS.

4.3 Promotion Show for the RAG Energy and Technology Group

In this project the CyberStage and the CSS served as platform to develop a highly impressive promotion show for a big German industrial company involved in mining, development of mining technology, construction and operation of power stations and development of ultra-light compound materials for aircraft construction (Ruhrkohle AG, Essen). Our portable CyberStage was installed on an international industrial fair in Hannover where more than thousand people visited the show. The task was to present the various facets of the company, to describe the research areas and introduce future projects in very short time (a few minutes) and with maximum impact on the visitors - they should be provided with a unique experience they would not forget so easily. The different scenarios where connected through a simple narrative structure. The most impressive part was a virtual visit of a coalmine - a very hostile and dangerous environment normally inaccessible to visitors. The scenario included fully functioning elevators, large conveyor belts, and rock grinding engines, which were very carefully modeled - both with respect to their visual and acoustic appearance. Most of the impact of such a scenario is achieved by a perfect combination of visual and sound design. The CyberStage and the CSS proved to be a very efficient tool to build such kind of applications.

4.4 Sound Spheres: A Virtual Sound Installation

Sound Spheres is a virtual sound installation exploring the basic features of the CyberStage audio-visual display system with the aim of shaping some of the yet unstructured vocabulary of musical expression and experience in cyberspace. The concentration on a few fundamental aspects of integrated audio-visual simulation was a conscious decision in the design of the installation and led to its abstract and minimalist character. In essence, Sound Spheres is about localization of moving sound and light sources. The role which direct and reflected sound and light play in the perception of space is explored in an experimental context.

When entering the installation, we are left in complete darkness and silence. With a virtual torch, we start to explore our obscure situation. While scanning the surroundings with the light beam, the scene slowly appears. By pointing the torch in different directions, we become aware of a big striped sphere enclosing us as well as several smaller rotating spheres slowly moving along circular paths. Then we discover how to activate these small spheres by inflating them with a virtual pump attached to the torch and operated by a button on the torch. The more we inflate a sphere, the longer it keeps on emitting percussive sounds and light flashes in regular rhythmical patterns. The omni-directional light flashes show us the entire scene for brief instants and the reflections on the shiny spheres help us to locate the light sources. The percussive sounds emitted synchronously with the light flashes excite the virtual room acoustics and provide us with a sense of distance and direction. The more spheres are active at a time, the denser the complex light and sound patterns will be. But pointing the virtual pump at a sphere for too long a time will lead to its explosion accompanied by a violent detonation noise and flash (see Figure 1). While freely floating in the space of weightlessness between the flashing and sounding spheres, we experience an ever-changing rhythmical tissue of spatialized sound and light. The perception of this space remains fragmentary and obscure due to the transient nature of the auditory and visual evidence leaving plenty of room for our imagination.

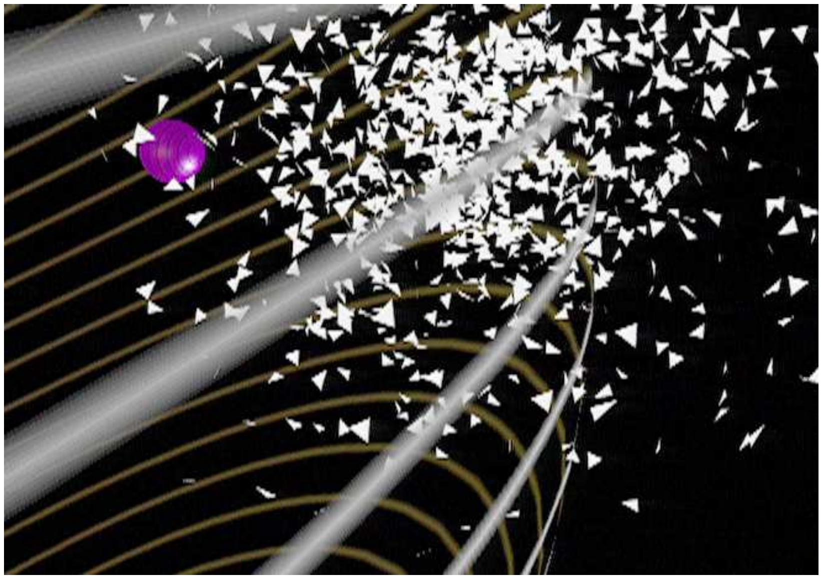

Figure 1: An Exploding Sound Sphere in the CyberStage

5 References

1. Eckel G, Göbel M, Hasenbrink F, Heiden W, Lechner U, Tramberend H, Wesche G, Wind J. Benches and Caves. In: Bullinger HJ, Riedel O (eds) Proc. 1st Int. Immersive Projection Technology Workshop. Springer-Verlag, London, 1997

2. Cruz-Neira C. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. Computer Graphics Proc., Annual Conference Series, 1993, pp 135-142

3. Dai P, Eckel G, Göbel M, Hasenbrink F, Lalioti V, Lechner U, Strassner J, Tramberend H, Wesche G. Virtual Spaces: VR Projection System Technologies and Applications. Tutorial Notes. Eurographics '97, Budapest, 1997, 75 pages

4. Krüger W, Fröhlich B. The Responsive Workbench. IEEE Computer Graphics and Applications. May, 1994

5. Breiteneder C, Gibbs S, Arapis C. TELEPORT - An Augmented Reality Teleconferencing Environment. Proc. 3rd Eurographics Workshop on Virtual Environments Coexistence & Collaboration, Monte Carlo, Monaco, 1996

6. Malham DG, Myatt A. 3-D Sound Spatialization using Ambisonics Techniques. Computer Music Journal 1995; 19/4

7. Puckette M. Combining event and signal processing in the Max graphical programming environment. Computer Music Journal 1991; 15/3

8. Dechelle F, DeCecco M. The IRCAM Real-Time Platform and Applications. Proc. of the 1995 International Computer Music Conference. International Computer Music Association. San Francisco, 1995

9. Jot JM, Warusfel O. A Real-Time Spatial Sound Processor for Music and Virtual Reality Applications. In: Proc. of the1995 International Computer Music Conference. International Computer Music Association, San Francisco, 1995