N >> 2: Multi-speaker Display Systems for Virtual Reality and Spatial Audio Projection

Perry R. Cook

Department of Computer Science (also Music)
Princeton University, Princeton, NJ, USA

[1]Georg Essl, [1]Georgos Tzanetakis, [2]Dan Trueman

[1]Department of Computer Science
[2]Department of Music
Princeton University, Princeton, NJ, USA

Abstract

This paper describes multi-speaker display systems for immersive auditory environments, collaborative projects, realistic acoustic modeling, and live musical performance. Two projects are described. The sound sub-system of the Princeton Display Wall project, and the NBody musical instrument body radiation response project. The Display Wall is an 18' x 8' rear-projection screen, illuminated by 8 high-resolution video projectors. Each projector is driven by a 4-way symmetric-multi-processor PC. The audio sub-system of this project involves 26 loudspeakers and server PCs to drive the speakers in real time from soundfile playback, audio effects applied to incoming audio streams, and parametric sound synthesis. The NBody project involves collecting and using directional impulse responses from a variety of stringed musical instruments. Various signal processing techniques were used to investigate, factor, store, and implement the collected impulse responses. A software workbench was created which allows virtual microphones to be placed around a virtual instrument, and then allows signals to be processed through the resulting derived transfer functions. Multi-speaker display devices and software programs were constructed which allow real-time application of of the filter functions to arbitrary sound sources. This paper also discusses the relation of spherical display systems to conventional systems in terms of spatial audio and sound-field reconstruction, with the conclusion that most conventional techniques can be used for spherical display systems as well.

1 Introduction and Motivation

The task of artifically creating a realistic auditory experience presents many challenges, bringing an abundance of mathematical and computational problems. Systems must provide flexible parametric models of the physical processes that create sounds, expressive controllers for manipulating sounds, and convincing display of sound sources at their correct locations in 3-dimensional space. The projects described in this paper focus on creating realistic auditory and gratifying musical experiences using multi-channel audio systems.

1.1 The Video Wall Project

The Princeton Video Wall is a large (18' wide x 8' high) rear-projection screen which supports research projects in a variety of areas. Sound for the Video Wall supports and enhances the basic visual functions of the project; collaboration, visualization, and immersion. Since the wall is large, there is the likelihood that certain visual display data will be out of the visual field of persons standing close to the wall. Sound can be used to draw directional auditory attention to an event, causing users of the wall to turn their heads toward the sound and thus bringing important visual information into their field of view. This is important for both collaboration and visualization projects. For collaborative projects, there is the possibility of different sub-groups working on different pieces of a project on different regions of the wall. For example, one of two groups might work on each half of the wall. Having enough speakers placed around the wall and room containing the wall allows for sounds to be displayed selectively within spatial regions. With additional signal processing, multi-speaker beam-forming and cancellation techniques can be employed to make the sound projection more focused then simply picking the speaker closest to the intended viewer/listener.

For visualization, data can be mapped to auditory as well as visual displays, allowing for enhanced multi-modal displays. By selecting the appropriate parameters and mappings of scientific data, processes, control flow, etc. , the human visual and auditory systems can be used to their best complimentary advantage.

Walk-through simulations which combine realistic sound with graphics are an active area of research. Simulators for training (flight cockpit, nuclear control rooms, battlefield, etc.) can benefit from realistic sounds, displayed with correct spatial placement [1]. Multi-site videoconferencing and collaboration can benefit from correctly spatialized display of voices and different geographical sites. Sound-field reconstruction for video-conferencing was recently found to give good perceptual results compared to binaural methods [2]. Other VR applications can benefit from having flexible, parametrically controlled, low-latency sound synthesis and processing. Many recent advances in sound synthesis by physical modeling can be exploited to make the immersive sound experience more realistic and compelling [3,4,5].

The goal of the sound sub-system of the Video Wall project is to obtain a working sound server solution which properly addresses the problems of synchronization and scalability while being flexible enough for a wide range of auditory display techniques. Applications should be able to seamlessly access the server and create their respective auditory events independently.

1.2 The NBody Project

Musical instruments radiate sound in directional, frequency dependent spatial patterns. For some instruments such as brass, the patterns are fairly predictable from the known properties of horns. For other instruments such as woodwinds, the patterns are more complex due to a number of toneholes which can radiate sound, and the configurations of these tonehole radiation sources vary with different fingerings [6].

For stringed instruments, the radiators are wooden boxes whose shapes, materials, and techniques of construction vary greatly between families, and from instrument to instrument within a sub-family. Players of electric stringed instruments are aware of the benefits of solid bodied instruments with electronic pickups, such as increased sustain times, decreased problems with feedback when amplifying the instrument, and the ability to process the sound of the instrument without the natural sound being heard. However, performers using solid body electric stringed instruments often find that these instruments lack the "warmth" associated with acoustic instruments, and using loudspeakers to amplify the electronic instrument does not provide a satisfactory dispersion of sound in performance spaces.

In recent years, synthesis by physical modeling has become more possible and popular [3]. To synthesize stringed instrument sounds using physical modeling, models of the bodies of these instruments are required which are efficient, realistic, and parametrically controllable. The latter is important to composers and interactive performers wishing to exploit the flexibility of parametric body models, allowing for dynamic changes in the parameters to be used as compositional and performance gestures. Another application area is virtual reality and 3D sound, which has brought a need for data and algorithms for implementing the directional radiation properties of musical instruments, the human voice, and other sound sources [7].

The goal of the NBody project is to obtain a set of useable filters for implementing realistic spatial radiation patterns of a variety of stringed instruments for simulation and performance. The data will be made publicly available in both raw and processed forms, allowing researchers to use it for various purposes, including verification of theories about the radiation properties of instruments. Using the NBody data, application programs for physical modeling synthesis and multi-speaker display for real-time performance have been written. Multi-speaker display devices have been built which take direct advantage of the NBody data set.

2 Approach and Methods

2.1 The Video Wall Multi-Speaker Audio Subsystem

Sound spatialization displays have been created which use binaural headphones and signal processing techniques, or stereo speakers using transaural cross-talk cancellation. Such systems require that the users head be motionless, or head tracking must be employed (for a review refer to [1,8]). Audio display systems like this are unsuitable for an untethered, multi-person, large work space like the Video Wall. A few virtual environment projects, such as The Cave, use a number of speakers, typically on the order of 4 to 8 [9].

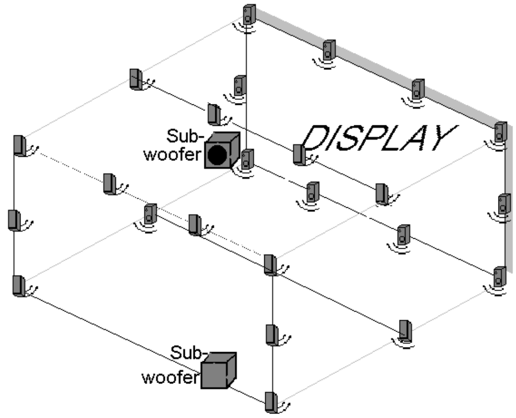
The Video Wall audio sub-system uses 26 loudspeakers, as shown in Figure 1. Some, but not excessive, acoustical treatment was done to the room, to cut down reflections from the speakers (and to the microphones). The room could have been made more acoustically dead, but one motivation of this project is to keep the budget and construction required for the sound sub-system in a range where other systems like it can be easily built or retrofitted to existing large visual display systems. Also, the room should be useable for video-only projects when the sound system is not desired, and rooms which are acoustically dead can prove to be unpleasant work environments.

2.1.1 Spatialization Research System

Widely known algorithms for spatial composition [10,11] have been implemented and tested on a small sub-system of 8 speakers in a cubic arrangement. It allows composers to specify arbitrary trajectories in the virtual sound space and an arbitrary order of reflection which are rendered using the image-source method [12]. The implementation scales well with the number of speakers and can be used for spatial compositions for the Video Wall audio setup.

2.1.2 The ND Sound Server

The server software was written to run on commodity PC hardware, and allows for scalability in terms of the number of speakers and the number of PCs available for sound computation. To allow for the most flexibility for future unforseen sound applications, all sub-systems of the audio software: soundfile playback, effects processing of sound streams, and sound synthesis, were written in an integrated but extensible fashion. This will allow sources to be connected and mixed as arbitrarily as possible, and new components to be added as new applications arise. The software interface to the multi-channel sound cards provides sample accurate synchronization between channels. Our system has some similarities with NCSA's VSS system [13], though our system is able to control 24 independent channels from one PC instead of the 2-channel stereo of VSS. A merge of the best features of these two projects is being considered. A rough outline of the structure of the implemented server software can be seen in Figure 2. The sound servers are connected with PCs running applications on the video wall via a low-latency, high-bandwidth LAN although there is as yet no synchronization or latency guarantees over the network. The sound server plays and mixes sound according to commands received through a TCP/IP connection using a special-purpose protocol.



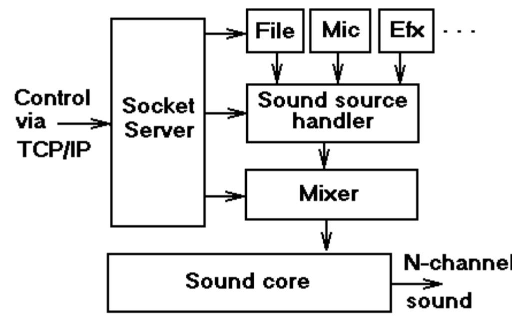


FIGURE 1. 26 Speaker array in Video Wall room. **FIGURE 2. Video Wall Sound Server Architecture.**

2.2 The NBody Musical Instrument Body Modeling Project

Impulse responses were collected for six stringed instruments, including three guitars, a mandolin, a violin, and a Hardanger (Norwegian folk) fiddle. An icosahedral (20 faces, 12 vertices) grid was constructed of ½" dowel rods, with a microphone mounted at each vertex. Figure 3 shows a photograph of the microphone array, with a researcher outside, a mandolin suspended inside the array, and the twelve microphone positions labeled . The total diameter of the sphere bounded by the microphone elements was approximately 4'. Twelve identical AKG C3000 cardioid microphones were positioned at the vertices of the icosahedron, pointing directly inward. The stringed instruments were excited using a calibrated force hammer. Data was collected both with the player holding the instrument, and with the instrument suspended without a player. A complete description of the data collection methods and NBody project can be found in [14].

In addition to impulse responses, samples of typical playing on each instrument were recorded using the 12-microphone array. To demonstrate the variety of sound radiation from an instrument in different directions in an anechoic condition, violin recordings are included with this paper. All example sounds were recorded simultaneously during a single performance. The first recording direction (Sound 1) is taken from microphone 2, pointed directly along the normal vector of the violin top plate. The second recording (Sound 2) is taken from microphone 3, to the player's left and below microphone 2. The third recording (Sound 3) is taken from microphone 8, directly behind and below the player and opposite microphone 2.

Two multi-speaker spherical display devices (nicknamed "the Boulder" and "the Bomb"), shown in Figures 4 and 5, were constructed. 12 speakers are arranged in an evenly-spaced array, facing outward. The Boulder and Bomb are essentially dodecahedral dual display devices for the icosahedral microphone data collection device shown in Figure 2. Any sound that was incident on a given microphone in the microphone array can be played back on the matching speaker in the speaker array, resulting in a fairly faithful reconstruction of the original spherical wavefront emitted by the instrument. Fast, multi-channel convolution has been implemented to allow any sound source, such as a solid body electric violin signal, to be filtered by the directional radiation impulse responses measured in the NBody data collection project.

3 Relation of Spherical Displays to Conventional Spatial Displays

Conventional spatial displays are speakers surrounding the listeners and roughly pointing towards them. A vast body of techniques for these conventional spatial displays (like the Video Wall speaker array) has been developed. Spherical displays are far less common and there are few known techniques. In general, conventional spatial displays like the Video Wall are not directed towards reconstructing the binaural channels for the audience because both creation of those signals and tracking of a large number of targets present immense technical problems. Instead such displays, if used for spatial display, are typically approached by simulation or discrete approximation of sound-fields [15,16,17].

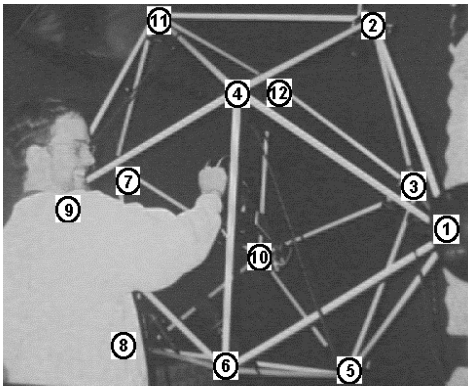


FIGURE 3. Microphone array structure.

FIGURE 4. “The Boulder”

FIGURE 5. “The Bomb”

For the conventional spatial display, the generated sound-field can be thought of as coming from sources spaced around the listener. One can say, that the generated sound-field is locally confined (or simply local) and that the placement of reproduced sources is global. A spherical display, on the other hand, simulates the radiation of one (or a few) local sound sources confined in the space described by the sphere. The radiated sound-field is, however, not inherently confined and hence can be simply viewed as being the global sound-field as generated by the local sources. Hence these two display methods are, in terms of sound-field reconstruction methods, complementary. In both cases, the speakers represent a discrete sampling of the sound-field, but the direction of radiation is inverted. From this it seems like techniques for improving the sampling errors made due to the spacing between the speakers, like panning techniques, can be directly applied to the spherical display. In this sense the distinction between conventional displays and spherical displays can be seen as the first having the speakers pointing inwards (global sources) while the latter having them pointing outwards (local sources), independent of the actual shape.

When talking about sound-field reconstruction from recorded sounds, one can look at the microphones discretely sampling the sound space, and the speakers reconstructing this discrete sampling by propagating the recorded sound in the direction of the incoming wave. As can be seen in Figure 3, this is also the method for measuring the radiation patterns of the musical instruments. First-order sound-field data, as typically recorded for Ambisonics[16], can therefore easily be played on an spherical-display, given that the right directionality is preserved.

4 Current Project Status and Future Directions

The initial Video Wall sound system uses multiple 8-channel soundcards, which just this year have begun to appear in the soundcard market, mounted in a single Pentium-class computer. The server software will be continually expanded to add spatialization, more effects processing, and more sound synthesis algorithms. As a research challenge, the computational limits of the single-processor/multi-card machine can be exceeded quite easily. The system will be re-architected to allow for multiple machines to compute and stream sound, and possibly for multiple machines to direct sound to the loudspeakers. Beam-forming for directional projection of sound will be added to the software. Beam-forming by speakers is a dual problem to the usual beam-forming from microphone arrays [18] and has close relation to the reconstruction of sound-fields [15]. We are investigating using the results from microphone array research in developing directional audio capabilities for the Display Wall project. Many microphones (on the order of 12) will be added to the system in the future. Research will be conducted on using the microphones to track an occupant as he/she moves about the room while speaking. The microphones will be used to implement a multi-site videoconferencing system which displays the voices of the participants in the correct locations around the screen. Finally, use of the microphones for active noise cancellation and improved directional speaker projection will be investigated [18,19].

The NBody project will continue forward, emphasizing analysis and parameterization of the measured impulse responses. It is planned to collect more impulse responses, for double bass violin and violincello. More efficient and parametric algorithms for implementation, modification, and spatial interpolation of the impulse responses will be investigated. The Bomb and Boulder have been played in live performance a number of times, and lessons learned from those performances can inform the construction of future multi-speaker performance systems.

5 Acknowledgements

The authors would like to thank Intel, Arial Systems Corporation and Interval Research for financial support. G. Essl would like to acknowledge financial support from the Austrian Federal Ministry of Science and Traffic.

6 References

[1] Wenzel, E. M., "Spatial Sound and Sonification," In Kramer, G. (ed.), "Auditory Display," Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume XVIII. Addison-Wesley, 1994, pp. 127-150.

[2] Evans, M. J., Tew, A. I., Angus, J. A. S., "Spatial Audio Teleconferencing – Which Way is Better?," In: Proceedings of the International Conference on Auditory Display, Palo Alto, 1997, pp. 29-37.

[3] Various authors, Computer Music Journal Special Issues on Physical Modeling, 16:4 & 17:1, 1992 & 1993.

[4] Huopaniemi, J., Karjalainen, M.,Välimäki, V., Huotilainen, T., "Virtual Instruments in Virtual Rooms – A Real-Time Binaural Room Simulation Environment for Physical Models of Musical Instruments," In: Proceedings of the International Computer Music Conference (ICMC), 1994, pp. 455-462.

[5] Cook, P. R., "Physically Informed Sonic Modeling (PhISM): Synthesis of Percussion Sounds," Computer Music Journal, 21:3, 1997, pp. 38-49.

[6] Caussé, R., Bresciani, J., and Warusfel, O., "Radiation of musical instruments and control of reproduction with loudspeakers," In: Proceedings of the International Symposium on Musical Acoustics, Tokyo, 1992.

[7] Hiipakka, J., Hänninen, R., Ilmonen, T., et al, "Virtual orchestra performance," In: Visual Proceedings of SIGGRAPH, 81, 1997, p. 81.

[8] Gardner, W. G., "3-D Audio Using Loudspeakers," Ph.D. thesis, Massachusetts Institute of Technology, 1997. Chapter 2.

[9] Pape, D., Carolina Cruz-Neira, C., and Czernuszenko, M., "The Cave User's Guide, version 2.5.6," http://www.evl.uic.edu/pape/CAVE/prog/CAVEGuide.2.5.6.html, 1996.

[10] Chowning, J. M., "The Simulation of Moving Sound Sources," AES Preprint No. 723 (M-3), Presented at the 38th Convention, 1970, pp. 1-9.

[11] Moore, F. R., "A General Method for Spatial Processing of Sound," Computer Music Journal, 7:3, 1983, pp. 559-568.

[12] Allen, J. B., Berkley, D. A., "Image Model for Efficiently Modelling Small-Room Acoustics," J. Acoust. Soc. Am. 65, 1979, pp. 943-950.

[13] Goudeseune, C., "Learning to use the Vanilla Sound Server. An Introduction to VSS 3.1," http://cage.ncsa.uiuc.edu/adg/VSS/doc/vss3.0usersguide.html, May 1998.

[14] Cook, P. and Trueman, D., "A Database of Measured Musical Instrument Body Radiation Impulse Responses, and Computer Applications for Exploring and Utilizing the Measured Filter Functions," In: Proceedings of the International Symposium on Musical Acoustics, Leavenworth, Washington, 1998, pp. 303-308.

[15] Berkhout, A. J., de Vries, D., Vogel, P., "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. 93:5, 1993, pp. 2764-2778.

[16] Gerzon, M. A., "Ambisonics in Multichannel Broadcasting and Video," J. Audio Eng. Soc., 33:11, 1985, pp. 859-871.

[17] Pulkki, V., "Virtual Sound Source Positioning Using Vector Based Amplitude Panning," J. Audio Eng. Soc., 45:6, 1997, pp. 456-466.

[18] Flanagan, J. L., Johnston, J. D., Zahn, R., Elko, G. W., "Computer-steered microphone arrays for sound transduction in large rooms," J. Acoust. Soc. Am. 78:5, 1985, pp. 1508-1518.

[19] Casey, M. A., Gardner, W. G., Basu, S., "Vision Steered Beam-Forming and Transaural Rendering for the Artificial Life Interactive Video Environment (ALIVE)," AES Preprint No. 4052 (B-5), Presented at the 99th Convention, 1995, pp. 1-23.