Combining Wave Field Synthesis and Multi-Viewer Stereo Displays

Jan P. Springer*
Bauhaus University Weimar

Christoph Sladeczek
Bauhaus University Weimar

Martin Scheffler
Bauhaus University Weimar

Jan Hochstrate
Bauhaus University Weimar

Frank Melchior
Fraunhofer IDMT Ilmenau

Bernd Fröhlich
Bauhaus University Weimar

ABSTRACT

We present our experiences of combining wave field synthesis audio with a projection-based multi-viewer stereo display. Wave field synthesis is able to simulate spatial sound sources of various kinds without the need for headphones or user tracking. Multi-viewer systems support multiple tracked users with individual perspectively correct stereoscopic images. The combination of both approaches allows the consistent display of virtual objects and spatial audio sources for multiple participants. First impressions with two application scenarios confirm that sound sources can be quite well located in space by each user. The system allows the creation of sound sources attached to virtual objects, which can be moved around in real-time without perceivable latency. In addition users appreciated the possibility of natural communication while they were exploring the audio-visual scenarios.

Keywords: Wave Field Synthesis, Multi-Viewer Stereo Displays, Immersive Virtual Environments

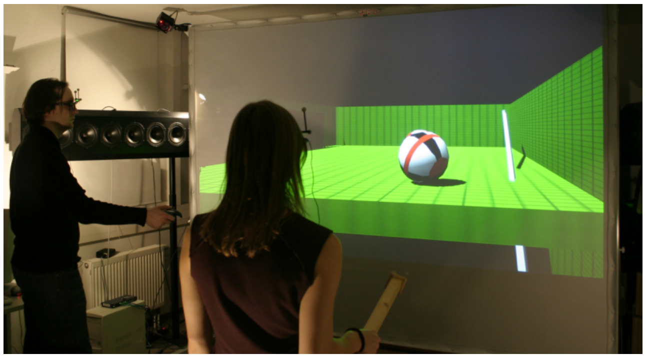

Figure 1: Two tracked users interacting with a billiard game. The installation is surrounded by a loudspeaker array to provide correct spatial sound to both users.

1 INTRODUCTION

Using spatial sound systems in combination with virtual environments (VE) has a long tradition. For projection-based environments the visual and the audio display support only a single tracked user with a spatially correct audio-visual feedback. Recently multi-viewer stereo displays (MVSD) as described in [3, 4] are becoming feasible for a small number of users. One of the questions is how spatial audio can be added to such a system. The head positions of all users are known to produce perspective correct images. Thus headphones could be used to supply them with the appropriate binaural signals. Unfortunately head phones limit the possibility for natural communication between the users, which is a much appreciated and used feature of projection-based systems.

Our idea is to combine wave field synthesis (WFS) with projection-based multi-viewer stereo displays. WFS delivers listener-position independent spatial audio reconstruction of high quality, which makes it the ideal choice for a consistent multi-user audio-visual environment. We have implemented such a system for two tracked users to evaluate the usability and the limitations of this approach.

The human auditory system uses inter-aural time and level differences to locate sound. Many of the existing methods for spatial audio reproduction are based on these effects and use psycho-acoustics. The most widespread methods are stereo and multi-channel surround reproduction. Stereo provides two channels and represents the minimal condition for spatial sound reconstruction. Extending this technique to multi-channel surround like 5.1 provides the listener with a more complete auditory experience. These techniques are mainly used in stationary environments. Users are generally constrained to the so called sweet spot, where the correct spatial impression will be achieved. If the listener leaves this area, the perception of spatial sound is immediately lost. Headphones are an easy solution to provide multiple users with correct spatial sound. However, communication between the participants in an application is seriously affected.

Since 2003 WFS is commercially available. A first application was its combination with traditional cinema projection techniques. For these 2D projection setups Melchior et al. [5] note that the image observed from different viewpoints combined with sound sources with true spatial depth generated by WFS may lead to perceptible discrepancies between the visible objects and their corresponding sound. Similar results were reported in [2]. This suggests that the combination of a WFS system with MVSD is beneficial since each viewer will have a perspective correct view onto the virtual scene. Initial tests for a single-viewer passive stereo setup showed encouraging results [6].

The main contribution of this work is the successful combination of a WFS system with a projection-based multi-viewer system shown in figure 1. We describe our hardware and software setup in detail and report first experiences with two application prototypes. Due to the arrangement of the loudspeaker panels in a one-dimensional array, virtual sound sources are constrained to a two-dimensional plane spanned by the loudspeakers and the user's head. This is a severe limitation of this system. Nevertheless, there are a number of different scenarios, which may benefit from our approach since humans are less sensitive to vertical than to horizontal sound movements.

2 WAVE FIELD SYNTHESIS

In the late 1980s Berkhout proposed a fundamentally new concept for sound reproduction called Wave Field Synthesis [1]. In contrast to all existing methods, WFS is a volume solution based on wave theory. It generates an accurate representation of the original wave field for the entire listening space. The Kirchhoff-Helmholtz and Rayleigh representation theorem form the theoretical foundation of the concept. Basics of WFS, corresponding equations, and references can be found

---

*jan.springer@medien.uni-weimar.de

in [9].

The Kirchhoff-Helmholtz integral states that if the sound pressure and velocity on a surface S of a source-free volume V is known, the pressure at any point A inside V can be computed. The original sound source distribution outside V is called primary source distribution; the monopole and dipole sources on the surface of the volume V used to synthesize the wave field are called secondary source distribution. Since A can be anywhere within V enclosed by S, the wave field within that volume is completely determined by the Kirchhoff-Helmholtz integral, whereas the integral is identical to zero outside the enclosure. The Kirchhoff-Helmholtz integral can be simplified assuming a fixed surface geometry and an non-zero wave field outside the closed surface. The wave field inside the closed surface is correctly described by these solutions, known as the Rayleigh integrals. The derivation of the Rayleigh integrals can be found in [9] among others. For practical applications, the ideal setup of loudspeaker planes around the listening area could be reduced to a linear array. Therefore [7, 9] have proposed an approach, resulting in an expression for a so called $2\frac{1}{2}$D-operator. The operator can be generalized for a primary source behind and in front of the secondary sources. This makes it possible to synthesize a sound field in the listening plane by the use of a line distribution of secondary sources. Figure 2 shows the simulation of a wave field for a virtual source in front and behind a linear loudspeaker array using WFS.

The handling of aspects like spatial sampling and other effects occuring in practical implementations of WFS with a line distribution of secondary sources has been explored in several implementations and experiments. The experiences show that a superb audio reproduction system can be realized using the principles of WFS [9] and that it is possible to produce wave fronts of arbitrary shapes. A listener, who is inside the generated wave field, will have a correct spatial perception of the sound field at nearly any position – even while moving. The number of supported users in a given wave field is unlimited, neglecting sound reflection or dampening through other listeners.

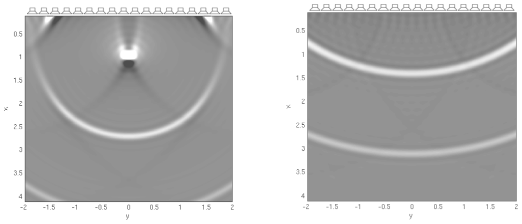

Figure 2: Wave field of a virtual sound source in front (left) and behind (right) the loudspeaker array.

3 MULTI-VIEWER STEREO DISPLAY

Our implementation of a multi-viewer stereo display system is based on the ideas described in [3, 4]. The system uses shuttering to separate the viewers and polarization to separate the stereo images per eye. In figure 3 a single state is shown with one user's shutters being open while the other user's shutters are closed. Liquid crystal (LC) shutters polarize light. Thus the light for the left and right eyes is polarized in orthogonal directions by simply rotating the left and right eye shutter and the corresponding shutters in front of the projectors against each other by 90 degrees.

The consumer level LCD projectors are calibrated to project to the same screen area. The left and right pairs for the users must be aligned pretty precisely, while the alignment between users is not as critical. Misaligned left and right projector pairs will have a strong impact on the users' stereo perception. Between the users a slight misalignment will hardly be noticeable. Since the shutters for the eyes and for the projectors are synchronized, there is no synchronization between the graphics cards necessary. The viewers in such a display system are provided with perspectively correct images for their individual view points. Thus users may reference virtual objects by just pointing at them with their fingers, which is not possible with conventional projection-based stereoscopic systems. This feature makes the interaction with the virtual scene as natural as using a white board in a group meeting.

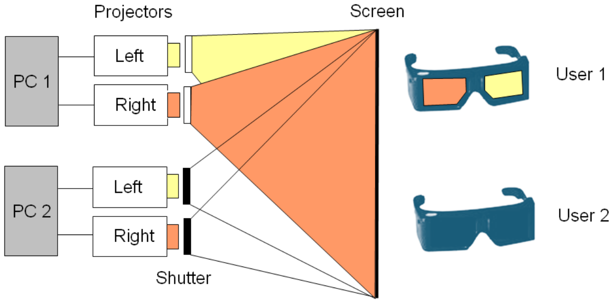

Figure 3: Multi-viewer stereo display functional principle.

4 IMPLEMENTATION

This section describes the actual hardware and software setup. The hardware setup supported stereoscopic viewing for two users and consisted of four speaker panels, a projection screen, four LCD projectors with LC-shutters, two LC shutter glasses, and an optical tracking system. Figure 4 gives an overview of the hardware components, which are explained in detail in section 4.1. Section 4.2 describes our software setup for driving the different hardware parts.

4.1 Hardware

The hardware setup of the WFS system consisted of four loudspeaker panels produced by IOSONO (http://www.iosono-sound.com/) and a driver PC. Each panel consists of eight loudspeakers and it is driven with a digital audio signal. The driver PC was installed with two RME Hammerfall DSP9636 multi-channel sound cards providing the audio signals to the panels. An additional PC with an RME Hammerfall DIGI9636 installed was used for the actual playing of sounds. The outputs of this sound player PC were directly connected to the inputs of the driver PC.

Four Hitachi PJ-TX100 LCD-projectors with mechanical lens shift were used with a front projection setup projecting onto a perforated screen of three by two meters in size. The mechanical lens shift of the projectors allowed an accurate alignment of the projectors. The projectors were mounted at roughly two meters height to allow a comfortable viewing and interaction distance to the screen. They were driven by a single PC supporting a PCIe-based dual graphics board installation using two nVidia Quadro FX4400 cards setup in dual-head mode.

Users were wearing shutter glasses and projectors were fitted with the appropriate shutter elements from disassembled glasses. The shutter system was driven by a programmable micro controller providing a 100Hz signal, which resulted in 50 images per eye per second. The projectors ran at 60Hz. No synchronization between the projectors or graphics cards was necessary. Viewer position and orientation as well as input device position and orientation were tracked using the ARTrack2 optical tracking system by A.R.T. A separate Windows PC was used to run the A.R.T tracking software. The machines were connected by a local 100MBit network.

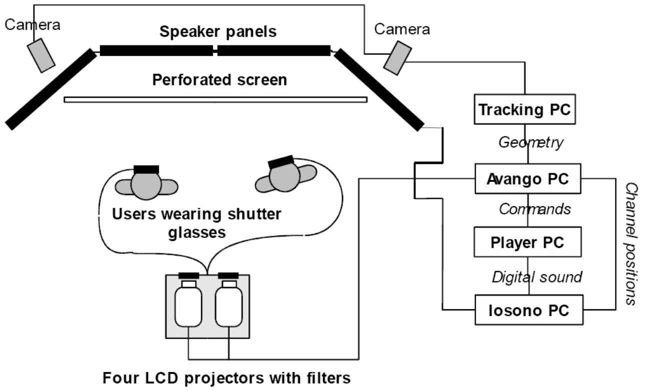

Figure 4: Hardware setup components.

4.2 Software

For controlling our various sound applications that were distributed over several PCs as described in section 4.1 the OpenSound Control protocol (OSC) [10] was deployed. OSC was developed to enable communication between computers, sound synthesizers, and other multimedia devices. OSC messages use an open-ended, dynamic, URL-style symbolic naming scheme to address the different functions provided by a device. In addition to the function address, each message can hold numeric and symbolic arguments and a high resolution time tag which can be used for synchronization.

The IOSONO sound renderer was used to generate the driving signals for the loudspeaker panels by processing the channels from the sound player PC. The software was running on a separate PC mentioned in section 4.1. The actual rendering parameters for each channel were provided by the VE application via the OSC protocol. For actually playing sounds we developed a sound server that is able to play sound sources on request via a custom OSC command scheme. The sound server uses one ALSA player per available hardware channel which redirects its output to a JACK low-latency audio server managing those channels. The ALSA players can be used to load, play, stop, etc. sounds via the above mentioned OSC command scheme.

Our VE application scenarios, which will be described in section 5 in more detail, were developed with Avango. Avango [8] (http://sourceforge.net/projects/avango/) is a framework for building stand-alone and distributed VE applications. It is based on OpenGL|Performer and provides a scripting layer connecting the interpreted language SCHEME to Avango. The scripting layer is used to develop applications using componentized objects. A field and field container concept similar to Inventor provides a complete SCHEME wrapper of the OpenGL|Performer scene graph API as well as of other VR specific functionality. The same mechanism is used to provide application specific functionality not programmed in the scripting language.

To support 3D audio within a VE application we developed a three layer software abstraction. Application programmers use the `avSoundSource` component to specify and control sound objects. `avSoundSource` is a specialization of `avDCS`, a transformation node. As such `avSoundSource`-nodes can be inserted in the scene graph. The field interface allows for specifying the actual sound file associated with the node as well as starting and stopping the sound file, adjusting the volume, and specifying one or more sound protocols. Sound protocols are modeled with `avSoundProtocol`. This abstract component defines an interface for controlling a certain sound library or protocol. We implemented specializations for the IOSONO OSC protocol as well as for OpenAL. The specializations provide additional interfaces for their configuration properties such as the location of the remote sound server or a local device to be used. Since most audio protocols assume absolute coordinate values `avSoundSource` provides its position as the absolute transformation to the scene graph's root node when sending updates via its associated sound protocol(s). `avSoundProtocol` instances can be shared between different instances of `avSoundSource` and register themselves at a sound service (`avSoundService`), which is responsible for sound channel virtualization. I.e. if an application creates more sound sources than there are sound channels available in hardware, then `avSoundService` will remap sound sources to free channels as soon as they become available. In case no hardware channels are available events from sound sources that are not assigned to any hardware channel are suppressed until one becomes available or the sound source stops sending updates.

An example scene graph is depicted in figure 5 where three sound sources have been inserted. As can be seen only two of the the three nodes use all available sound protocols while one sound source sends its updates to the WFS sound protocol only.

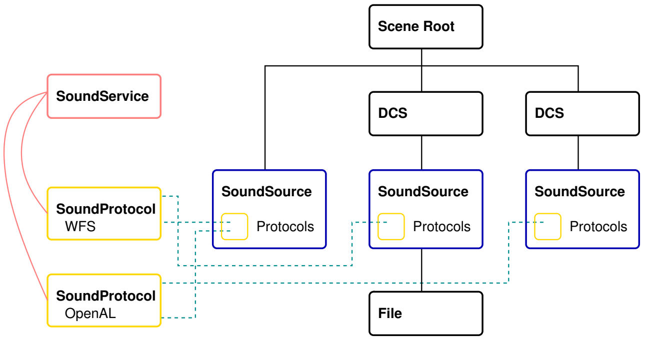

Figure 5: Sound components in the Avango scene graph.

5 APPLICATION SCENARIOS

Our application prototypes were designed to gather experiences with respect to the following issues: latency problems of the audio-visual system, effects of time differences between auditive and visual events, effects of vertical offsets between auditive and visual objects, and performance problems due to a large number of sound sources. We decided to implement these tests in two different application scenarios, described in sections 5.1 and 5.2, each demonstrating a different type of interaction.

5.1 The Billiard Scenario

To test responsiveness of the audio system and the overall realism of sound effects, we built a billiard game simulation, shown in figure 1. In the application, the users look into a room three meters in width and depth. Two billiard balls are placed randomly on the floor. A virtual billiard queue is controlled through a tracked stick. The balls roll around on the floor, bounce off the cushion, and collide with each other. For any of these events a sound emitting event is generated, i.e. the queue hitting a ball, the balls colliding, and a ball hitting a cushion. The billiard game simulation takes place on a plane. By adjusting the level of the floor, we were able to constrain the movement of the balls within the area of realistic sound projection. The application creates short sound sequences in rapid succession.

5.2 An Interactive Installation: A Forest Brook and Stones

The purpose of this application was to test the separation of different sound sources and to investigate slow moving sound sources. Users were standing in front of a forest landscape, as shown in figure 6, with a brook flowing from left to right. Using a hand held trackball, users were able to move a virtual 3D-cursor above the ground of the landscape. When pressing a button on the input device, a stone dropped from the position of the 3D-cursor to the ground. If the stone hit the water of the brook, a splashing sound was generated followed by a gurgling or murmuring sound as if the water was flowing around it. If the stone hit the forest ground, a short thudding sound was generated and the stone disappeared. Stones placed in the brook could be picked up by the users and moved around in a drag-and-drop kind of way. The cursor was constrained to move within a depth of about five meters behind the screen. Two ambient sound sources positioned further away suggested a forest setting by providing continuous bird songs as wells as forest sounds including an occasional wood-pecker working a tree.

Figure 6: Users interacting with an "interactive installation" type application.

5.3 Results

The application scenarios were shown to a wider audience at an open house event in July 2005. Most users did not notice discrepancies between the vertical positions of sounds and their visual counterparts. Almost all visitors took the sound for granted.

Studies by de Bruijn and Boone show that sound localization accuracy in the vertical plane is not very high. Discrepancies between object positions in the video image and the audio source will not be noticed by the human auditory system within a deviation of 22 degrees in the vertical plane [2]. Users were able to approach the screen up to a minimum distance of one meter. This restriction resulted from the constrained space where the system was installed which limited the tracking system and WFS loudspeaker array positions. Because of the limited vertical sound position discrimination and the minimum horizontal viewer distance we were able to use an area of 40 cm above and below the speaker panels for virtual object positions while still accurately rendering the associated sound for a virtual object in the listening plane of the WFS system. The speaker panels where horizontally centered behind the screen.

The billiard simulation was used for verifying that audio and image events were synchronized. We tested the application successfully with up to eight interactive sound objects (balls) without noticing latency problems. Sound localization was better demonstrated with the interactive installation showing the brook. Users were able to point at the (virtual) origin of the different water sounds, even with their eyes closed and the referenced location was roughly equal for the two users.

Surprisingly little system problems were encountered. Latency problems exhibited by early versions of the billiard simulation were reduced to an acceptable level by tuning buffer sizes and using real-time functionality provided by the operating system.

6 CONCLUSION

We have presented the combination of a multi-viewer stereo display with a wave field synthesis audio system to create a consistent audio-visual multi-user environment. Initial tests with basic application scenarios revealed the potential of this approach for co-located collaborative as well as edutainment related VR applications.

This work is a first step in the direction of a deeper understanding of the interrelationship of multi-viewer stereo displays and wave field synthesis. We plan to conduct formal user studies to identify the relevant parameters of audio-visual objects and the limitations of our approach.

ACKNOWLEDGMENTS

We thank Robert Dünger, Moritz Fehr, Tom Freitag, Markus Hartleb, and Michael Strauss for their help.

REFERENCES

[1] A. J. Berkhout. A Holographic Approach to Acoustic Control. *Journal of the Audio Engineering Society*, 36:977–995, 1988.

[2] W. P. J. de Bruijn and M. M. Boone. Application of Wave Field Synthesis in Life-Size Videoconferencing. In *114th Convention of the AES*, pages 1–17, Amsterdam, Netherlands, March 2003. Audio Engineering Society (AES).

[3] B. Fröhlich, J. Hoffmann, K. Klüger, and J. Hochstrate. Implementing Multi-Viewer Time-Sequential Stereo Displays Based on Shuttered LCD Projectors. In *Proceedings of 8. International Immersive Projection Technology Workshop*, May 2004.

[4] B. Fröhlich, J. Hochstrate, J. Hoffmann, K. Klüger, R. Blach, M. Bues, and O. Stefani. Implementing Multi-Viewer Stereo Displays. In *WSCG 2005*, pages 139–146, January 2005.

[5] F. Melchior, S. Brix, Th. Roeder, Th. Sporer, and B. Klehs. Wave Field Synthesis in Combination with 2D Video Projection. In *24th AES Conference*, Banff, Canada, June 2003.

[6] F. Melchior, D. de Vries, and S. Brix. Zur Kombination von Wellenfeldsynthese mit monoskopischer und stereoskopischer Bildwiedergabe. In *DAGA'05*, 2005.

[7] E. W. Start. *Direct Sound Enhancement by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.

[8] H. Tramberend. Avocado: A Distributed Virtual Reality Framework. In *Proceedings IEEE Virtual Reality '99 Conference*, pages 14–21. IEEE Computer Society, March 1999.

[9] E. N. G. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.

[10] M. Wright, A. Freed, and A. Momeni. OpenSound Control: State of the Art 2003. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, pages 153–159, Montreal, Canada, 2003.