



AUDIO ENGINEERING SOCIETY
Italian Section

ANNUAL MEETING 2005

Paper: 05005

Como, 3-5 November
Politecnico di MILANO

SPATIAL SOUND REPRODUCTION WITH WAVE FIELD SYNTHESIS

RUDOLF RABENSTEIN, SASCHA SPORS

Telecommunications Laboratory, University of Erlangen-Nuremberg, Erlangen, Germany
rabe@LNT.de

Stereophonic spatial sound reproduction systems with two, five, or more channels are designed for a certain listener position and work well in its vicinity, the so called sweet spot. For listeners at other positions, the quality of the spatial reproduction may be degraded. This contribution describes an advanced spatial reproduction technique called wave field synthesis. It is based on a physical description of acoustic wave propagation and uses loudspeaker array technology for sound field reproduction without the sweet spot limitation. After discussing the physical foundations, the main steps from the acoustic description to the determination of the loudspeaker signals are outlined. Finally an implementation of a wave field synthesis system with 48 channels is presented.

INTRODUCTION

Conventional systems for the reproduction of spatial audio are mainly based on intensity panning techniques. They adjust the contributions from the different loudspeaker channels in such a way that their superposition produces the required intensity levels at the listener's ears. Consequently, the reproduction quality is only guaranteed in the vicinity of the targeted listener position, the so called sweet spot.

Several novel audio reproduction techniques have been suggested to enlarge the preferable listening area. They can be roughly categorized into advanced panning techniques, Ambisonics systems, and wave field synthesis. Advanced panning techniques aim at enlarging the sweet spot by increasing the number of loudspeakers. An example is the vector base amplitude panning technique (VBAP) [1]. Ambisonics systems represent the sound field in an enclosure by an expansion into low order three-dimensional basis functions [2].

Wave field synthesis is based on a physical description of the propagation of acoustic waves. It uses loudspeaker array technology to correctly reproduce sound fields without the sweet spot limitation. The main applications of wave field synthesis are in the areas of entertainment and the performing arts but it may also be used for the creation of virtual room reverberation or of virtual noise fields.

Wave field synthesis techniques are formulated in terms of the acoustic wave equation and the description of its solutions by Green's functions. These foundations have been initially developed by the Technical University of Delft [3, 4, 5] and were later extended within the European project CARROUSO [6].

This contribution gives an overview on the foundations, the design and the implementation of wave field synthesis systems. More details can be found in [7, 8].

1 PHYSICAL FOUNDATIONS

1.1 Huygens' Principle

Wave field synthesis (WFS) is based on Huygens' principle. It states that any point of a wave front of a propagating wave at any instant conforms to the envelope of spherical waves emanating from every point on the wavefront at the prior instant. This principle can be used to synthesize acoustic wavefronts of an arbitrary shape. By placing the loudspeakers on an arbitrary fixed curve and by weighting and delaying the driving signals, an acoustic wavefront can be synthesized with a loudspeaker array. Figure 1 illustrates this principle.

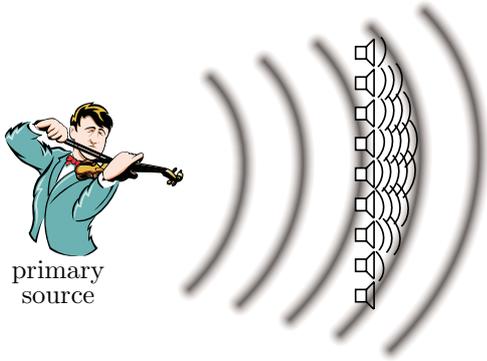


Figure 1: Synthesis of a wave front by a loudspeaker array with appropriately weighted and delayed driving signals.

1.2 Kirchhoff-Helmholtz Integral

The mathematical foundation of this more illustrative description of WFS is given by the Kirchhoff-Helmholtz integral (1). It can be derived by using the wave equation and Green's integral theorem [9]

$$P(\omega, \mathbf{x}) = - \oint_{\partial V} \left(G(\omega, \mathbf{x}|\mathbf{x}') \frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}') - P(\omega, \mathbf{x}') \frac{\partial}{\partial \mathbf{n}} G(\omega, \mathbf{x}|\mathbf{x}') \right) d\mathbf{x}'. \quad (1)$$

Figure 2 illustrates the parameters: ∂V denotes the surface of an enclosed space V , $|\mathbf{x} - \mathbf{x}'|$ the vector from a surface point \mathbf{x}' to an arbitrary listener position \mathbf{x} within the volume V , and \mathbf{n} the surface normal. $P(\omega, \mathbf{x})$ and $P(\omega, \mathbf{x}')$ are the Fourier transforms of the sound pressure distribution within the volume V and on the surface ∂V , respectively. The Green's function $G(\omega, \mathbf{x}|\mathbf{x}')$ describes the propagation of sound waves within V .

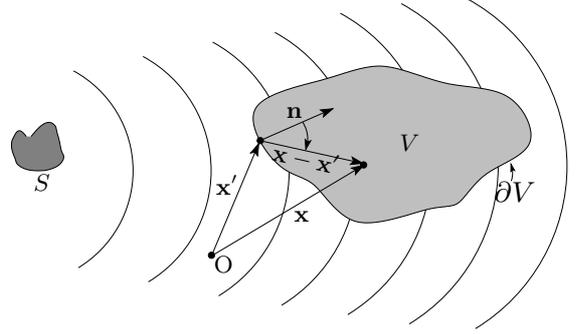


Figure 2: Reproduction of the spatial wave field emitted by the virtual source S inside the volume V . O denotes the origin of the coordinate system for the parameters of the Kirchhoff-Helmholtz integral (1).

The Kirchhoff-Helmholtz integral states that at any listening point within the source-free volume V the sound pressure $P(\mathbf{x}, \omega)$ can be calculated if both the sound pressure and its gradient are known on the surface enclosing the volume.

2 WAVE FIELD SYNTHESIS

2.1 Kirchhoff-Helmholtz Integral based Sound Reproduction

For the sound reproduction scenario according to Fig. 2 the Green's function $G(\omega, \mathbf{x}|\mathbf{x}')$ and its directional gradient $\frac{\partial}{\partial \mathbf{n}} G(\omega, \mathbf{x}|\mathbf{x}')$ can be understood as the field emitted by sources placed on ∂V . These sources are called *secondary sources*. The strength of these sources is determined by the sound pressure $P(\omega, \mathbf{x}')$ of the sound field which is emitted by the virtual source S and recorded at the surface ∂V and by its directional pressure gradient $\frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}')$.

Now the Kirchhoff-Helmholtz integral can be interpreted as follows: Imagine that a virtual source S causes a certain sound pressure field $P(\omega, \mathbf{x})$ inside the volume V and that the sound pressure $P(\omega, \mathbf{x}')$ and its directional gradient $\frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}')$ are known at the surface ∂V . Then the same sound pressure field $P(\omega, \mathbf{x})$ can be reproduced inside of V if appropriately chosen secondary sources are driven by the sound pressure and its directional pressure gradient at ∂V . This interpretation is the theoretical basis of WFS sound reproduction based on the Kirchhoff-Helmholtz integral (1).

It is not required to actually record a sound field at the surface ∂V in order to know the sound pressure and the directional pressure gradient. Suitable techniques allow to compute its values from microphone

recordings at other locations and from models for the acoustic wave propagation.

In the simplest case, the Green's function $G(\omega, \mathbf{x}|\mathbf{x}')$ is the field of a monopole point source distribution on the surface ∂V . Then the directional gradient of this Green's function is the field of a dipole source whose main axis lies in direction of the normal vector \mathbf{n} . Thus, the Kirchhoff-Helmholtz integral states, that the sound pressure inside the volume V can be controlled by a monopole and a dipole point source distribution on the surface ∂V enclosing the volume V .

This interpretation of the Kirchhoff-Helmholtz integral sketches a first draft of a technical system for spatial sound reproduction. In rough terms, such a system would consist of technical approximations of acoustical monopoles and dipoles by appropriate loudspeakers. These loudspeakers cover the surface of a suitably chosen volume around the possible listener positions. They are excited by appropriate driving functions to reproduce the desired sound field inside the volume.

However, there remain a number of fundamental questions to be resolved on the way to a technical realization. These include the necessity of both monopole and dipole sources, the reduction to a two-dimensional loudspeaker configuration, the required density of loudspeakers on the surface and the determination of appropriate driving functions for the speakers.

2.2 Monopole and Dipole Sources

Technical approximations of acoustical monopoles and dipoles consist of loudspeakers with different types of enclosures. A restriction to only one type of sources would be of advantage for a technical realization. For example the exclusive use of monopole sources facilitates a technical solution with small loudspeakers in closed cabinets.

The use of true monopole and dipole sources in (1) recreates $P(\omega, \mathbf{x})$ for all positions inside of V but it would ideally cause zero sound pressure outside. Such a restriction is usually not required for spatial sound reproduction. However, as long as the reproduction is correct inside of V , almost arbitrary sound fields outside may be tolerated, as long as their reproduction volume is moderate. This situation suggests the following trade-off: Use one type of sound sources only and tolerate some sound pressure outside of V .

To realize this trade-off, a Green's function $G(\omega, \mathbf{x}|\mathbf{x}')$ is constructed with zero derivative

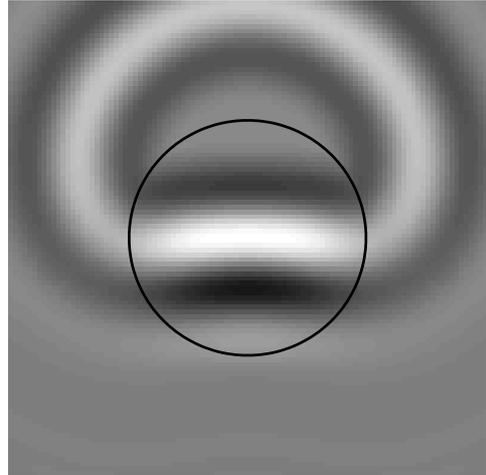


Figure 3: Reproduction of a plane wave with monopole sources. The black circle denotes a circular contour with a diameter of 3m.

$\frac{\partial}{\partial \mathbf{n}} G(\omega, \mathbf{x}|\mathbf{x}') = 0$ on the surface ∂V . Then the second term in (1) vanishes and the Kirchhoff-Helmholtz-Integral for $\mathbf{x} \in V$ reduces to

$$P(\omega, \mathbf{x}) = - \oint_{\partial V} G(\omega, \mathbf{x}|\mathbf{x}') \frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}') d\mathbf{x}' . \quad (2)$$

With a suitable choice of the Green's function this relation describes a distribution of monopoles on the surface ∂V .

Fig. 3 shows the simulation of a circular array reproducing a plane wave front travelling from the top to the bottom. The symmetry induced by requiring a zero normal derivative at the boundary (i.e. the circular array) is evident from the wave front outside the array. Each point source on the boundary contributes to a plane wave front inside the array and to another wave front outside.

2.3 Reduction to two dimensions

The volume V certainly has to be large enough to enclose at least a small audience or to give a single listener room to move within the sound field. Covering the whole surface with suitable sound sources appears to be a technological and economical challenge. Furthermore, it may not be required to reproduce the sound field within the entire volume. A correct reproduction in a horizontal plane at the level of the listeners' ears may be sufficient. Such a simplification requires to reduce the 3D problem to two spatial dimensions. Two steps are taken to convert the 3D volume description of (2) to a 2D surface description.

The first step consists of a change of the geometry of the problem. Since the volume V in (1) may have arbitrary shape, it can be specialized to a prism where the shape does not depend on one of the three spatial coordinates, say z . If furthermore the source terms $\frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}')$ in (2) do not depend on z , then the sound field inside the prism will not depend on z either. In other words the values of $P(\omega, \mathbf{x})$ are the same in every cut through the prism, i.e. for $\mathbf{x} = [x \ y \ z]$

$$P(\omega, [x \ y \ z]) = P(\omega, [x \ y \ z_0]) = P_{2D}(\omega, [x, y]) .$$

Here, $P_{2D}(\omega, x, y)$ describes the planar sound pressure distribution for a fixed value z_0 . A suitable choice for z_0 is the height of the listener's ears.

The sources on the surface of the prism can be modeled as line sources in the direction of z with varying root point (x, y) along the contour around the cut through the prism. At this point the problem is still three-dimensional, but with a specialized geometry.

To arrive at a model for a practical solution, a second step replaces the line sources by point sources placed at the root points of the line sources around the cut through the prism [4]. By this measure, the surface distribution of monopoles enclosing the volume V is converted to a contour distribution around a cut through a prism.

2.4 Spatial Discretization

The previous sections showed how the rather general statement of the Kirchhoff-Helmholtz integral can be narrowed down to a model for a spatial reproduction system. A hypothetical distribution of monopole and dipole sources on a 2D surface around the listener has been replaced by a distribution of monopoles on a 1D contour in a horizontal plane.

For a technical solution, this spatially continuous source distribution has to be replaced by an arrangement of a finite number of loudspeakers with a monopole-like source directivity. Experience with existing wave field synthesis implementations indicates that reasonable values for the loudspeaker spacing lie between 10 cm and 20 cm.

Fig. 4 shows a number of listeners surround by an array of loudspeaker cabinets which closely approximate monopole sources. The loudspeakers are mounted in the height of the listener's ears.

2.5 Driving signals

Once the source distribution is approximated by a sufficiently dense grid of loudspeakers, their driv-



Figure 4: Listeners surrounded by a wave field synthesis array.

ing signals have to be generated by signal processing hardware and digital-to-analog converters. To determine the loudspeaker driving signals, the nature of the desired wave field has to be taken into account.

Wave fields may be modeled by arrangements of different types of sources, e.g. monopoles and dipoles, and by plane waves. The determination of the driving signals from a model of the wave field is called *model based rendering*. On the other hand, a wave field can be recorded in a natural environment like a concert hall or a church. Obtaining the driving signals from a recorded wave field is called *data based rendering*.

2.5.1 Model based rendering

For model-based rendering, models for the sources are used to calculate the driving signals for the loudspeakers. Point sources and plane waves are the most common models used here.

The source signals $s(t)$ may be obtained by recording real sounds or by synthesis of virtual sounds. Then the well known analytic models for point sources or plane waves allow to calculate the value of the normal derivative in (2).

By transforming the result back into the time-domain, the driving signal $q_i(t)$ of the loudspeaker number i can be computed from the source signal $s(t)$ by delaying, weighting and filtering [3, 5],

$$q_i(t) = a_n (h(t) * s(t)) * \delta(t - \tau), \quad (3)$$

where a_n and κ denote an appropriate weighting factor and delay respectively, and $h(t)$ a low pass filter

which considers the effect of the normal derivative $\frac{\partial}{\partial \mathbf{n}} P(\omega, \mathbf{x}')$. $\delta(t)$ denotes the Dirac impulse function. Multiple sources can be synthesized by superimposing the loudspeaker signals from each source.

Plane waves and point sources can be used to simulate classical loudspeaker setups, like stereo and 5.1 setups. Thus WFS is backward compatible to existing sound reproduction systems and can even improve them by optimal loudspeaker positioning in small listening rooms and listening room compensation.

2.5.2 Data based rendering

The loudspeaker driving signals may also be determined from measurements of the room acoustics in an existing listening environment.

The impulse responses for auralization cannot be obtained the conventional way by simply measuring the response from a source to a listener position. In addition to the sound pressure also the particle velocity is required to extract the directional information. This information is necessary to take the direction of the traveling waves during auralization into account. These room impulse responses have to be recorded by special microphones and setups as shown in [10].

3 System Implementation

An implementation of a wave field synthesis system with a circular loudspeaker array is shown in Fig. 4. Here the listening area is a disc with a diameter of 3 m. A total of 48 two-way loudspeakers are mounted on the circumference of the circle with a spacing of about 20 cm. The analog driving signals are delivered by three 16-channel audio amplifiers with digital inputs shown in Fig. 5. The digital input signals are the result of the convolution (3) performed for each of the 48 channels. It is realized by fast convolution techniques in real-time on a personal computer. The system described here is located at the Telecommunications Laboratory (Multimedia Communications and Signal Processing) of the University of Erlangen-Nuremberg in Germany [11].

4 Conclusion

Wave field synthesis is a spatial audio reproduction technique which is based on the acoustic wave equation and the representation of its solutions by Green's functions. Starting from these physical foundations, it has been shown how to derive the driving signals for the loudspeaker array.



Figure 5: Three 16-channel audio amplifiers for the array in Fig. 4 mounted in a 19 inch rack.

The derivation is valid for rather general geometries and sizes of loudspeaker arrays. Furthermore, no assumption on the position of the listener is required. Then the reproduced sound field is physically correct within the limitations imposed by spatial discretization effects. The computation of the loudspeaker driving signals is conceptually simple and is performed by a multichannel convolution.

However, the practical realization of wave field synthesis has some pitfalls, which can be avoided by further signal processing techniques. These pertain the simplified monopole model of the loudspeakers and the acoustical reflections of the loudspeakers within the listening room.

So far it has been assumed that acoustical monopoles can be approximated well by small loudspeakers with closed enclosures. If required, this approximation can be improved with digital compensation of non-ideal loudspeaker properties [12]. The second pitfall consists of the reflections of the loudspeaker array signals in the listening room. They may degrade the performance level predicted from theory. Countermeasures are passive or active cancellation of these reflections. Especially, active cancellation seems promising by using the loudspeaker arrays for reproduction also for the cancellation of room reflections [8].

References

- [1] V. Pulkki, “Compensating displacement of amplitude-panned virtual sources,” in *Proc. of the AES 22nd Int. Conference*. 2002, pp. 186–195, Audio Engineering Society.
- [2] M.A. Gerzon, “Ambisonics in multichannel broadcasting and video,” *Journal of the Acoustic Society of America*, vol. 33, no. 11, pp. 859–871, Nov. 1985.
- [3] A.J. Berkhout, “A holographic approach to acoustic control,” *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, December 1988.
- [4] J.-J. Sonke, D. de Vries, and J. Labeeuw, “Variable acoustics by wave field synthesis: A closer look at amplitude effects,” in *104th AES Convention*, Amsterdam, Netherlands, May 1998, Audio Engineering Society (AES).
- [5] D. de Vries, E.W. Start, and V.G. Valstar, “The Wave Field Synthesis concept applied to sound reinforcement: Restrictions and solutions,” in *96th AES Convention*, Amsterdam, Netherlands, February 1994, Audio Engineering Society (AES).
- [6] S. Brix, T. Sporer, and J. Plogsties, “CARROUSO - An European approach to 3D-audio,” in *110th AES Convention*. Audio Engineering Society (AES), May 2001.
- [7] S. Spors, H. Teutsch, A. Kuntz, and R. Rabenstein, “Sound field synthesis,” in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y.Huang and J.Benesty, Eds. Kluwer Academic Publishers, 2004.
- [8] S. Spors, H. Buchner, and R. Rabenstein, “Adaptive listening room compensation for spatial audio systems,” in *European Signal Processing Conference (EUSIPCO)*, 2004.
- [9] A.J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” *Journal of the Acoustic Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.
- [10] E. Hulsebos, D. de Vries, and E. Bourdillat, “Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis,” in *110th AES Convention*, Amsterdam, Netherlands, May 2001, Audio Engineering Society (AES).
- [11] “Multimedia Communications and Signal Processing at the University of Erlangen-Nuremberg,” <http://www.lnt.de/LMS>.
- [12] S. Spors, D. Seuberth, and R. Rabenstein, “Multiexciter panel compensation for wave field synthesis,” in *31. Deutsche Jahrestagung fuer Akustik*, 2005.



Rudolf Rabenstein received the degrees “Diplom-Ingenieur” and “Doktor-Ingenieur” in electrical engineering from the University of Erlangen-Nuremberg, in 1981 and 1991, respectively, as well as the “Habilitation” in signal processing in 1996. He worked with the Telecommunications Laboratory of this university from 1981 to 1987 and since 1991. From 1998 to 1991, he was with the physics department of the University of Siegen, Germany. His research interests are in the fields of multi-dimensional systems theory and simulation, multimedia signal processing, and computer music.



Sascha Spors studied electrical engineering at the University of Erlangen-Nuremberg and received the degree “Diplom-Ingenieur” in 2001. He is working as research assistant at Telecommunications Laboratory since then. His research areas include multichannel sound reproduction and the active compensation of reflections emerging from the listening room.