High-Quality Acoustic Rendering with Wave Field Synthesis

S.Spors, H.Teutsch and R.Rabenstein

University of Erlangen-Nuremberg
Telecommunications Laboratory
Cauerstraße 7, 91054 Erlangen, Germany
Email: {spors,teutsch,rabe}@LNT.de

Abstract

The available technologies for the presentation of audiovisual scenes to large audiences show different degrees of maturity. While high quality physics based rendering of 3D scenes is found in many visual applications, the presentation of the accompanying audio content is based on much simpler technologies. Multi-channel cinema sound systems are only capable of delivering sound effects, but they do not faithfully reproduce an acoustic scene. New methods for acoustic rendering are required to provide physically correct recreation of audiovisual environments.

A new technology for the reproduction of sound fields is based on Huygen's principle. It utilizes a large number (tens or hundreds) of loudspeakers to recreate a 3D sound field based on the physical description of the original acoustic environment. This paper describes different aspects of this technology, presents applications pursued by the European project CARROUSO and our own contributions to the project including the development of algorithms for loudspeaker and listening room compensation and real-time software implementation of a rendering software on a PC.

1 Introduction

Recent advances in computer graphics and video coding have quickly penetrated the market. Techniques for simulation and visualization intended for the professional user only a few years ago can be now found in home applications like games for personal computers or gaming consoles. Streaming video in different levels of quality is readily available from countless numbers of web sites. The simultaneous immersion of hundreds of people into visual content is commonplace in large cinemas.

Somewhat lagging behind is the presentation of the accompanying audio content. Internet radio provides two channel stereo at best. For home entertainment, the 5.1 surround sound standard for DVD is well established. However, only a minority of home users has sufficient space available for a proper setup of the front and rear loudspeakers according to the ITU recommendation [1]. More than five broadband audio channels are found in movie theaters, where different multi-channel standards compete for the cinema market. Rather than true acoustic rendering of sound fields, these multichannel movie theater applications aim mostly at delivering various sound effects to a large audience.

In short, rendering of 3D scenes is commonplace in visual applications, but this is not at all the case for audio content. The closest one can get is within the so called sweet spot of the current 5.1 format, provided a good mix of the recorded material and a proper setup of the set of loudspeakers. Theater systems deliver stunning sound effects, but not a true 3D representation of an original or virtual sound field.

This paper discusses a new technology for conveying sonic spaces from the studio to a large number of remote listeners. Based on the principle of wave field synthesis, a true reproduction of a dynamically varying sound field is possible without being restricted to a sweet spot. Thus, also large audiences or moving listeners can benefit from this new multichannel approach. Furthermore, an implementation of this technology exploited in the European project CARROUSO [2] is presented.

We proceed as follows: Section 2 presents the theory of wave field synthesis, discusses two different rendering approaches and algorithms for listening room and loudspeaker compensation. Section 3 gives an overview of the CARROUSO project and shows details of the implementation, including model based rendering software.

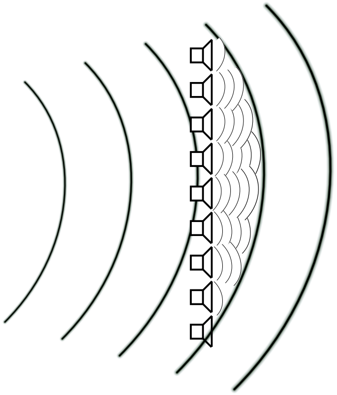


Figure 1: Basic principle of wave field synthesis

2 Wave Field Synthesis

The theory of wave field synthesis (WFS) has been initially developed at the Technical University of Delft over the past decade [3, 4, 5, 6, 7] and is now investigated further within the CARROUSO project. In contrast to other multi-channel approaches, it is based on fundamental acoustic principles. This section gives a broad overview of the theory as well as on methods for rendering and listening room and loudspeaker compensation. In the context of WFS, rendering denotes the production of an appropriate sound field according to the use of the term rendering for computer graphics.

2.1 Theory

The basic principle of WFS is based on the Huygen's principle. Huygens stated 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. Of course, it is not very practical to position the acoustic sources on the wavefronts for synthesis. 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.

The mathematical foundation of this more illustrative description of WFS is given by the Kirchhoff-Helmholtz integral (1), which can be derived by using the wave equation and the Green's integral theorem [8]

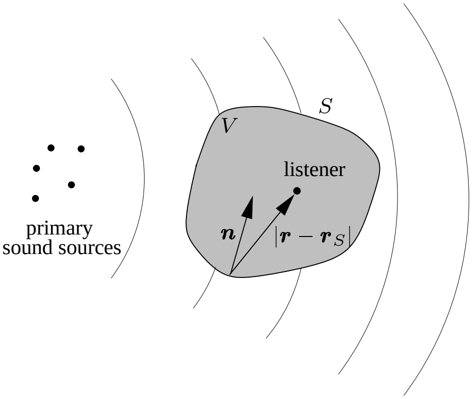


Figure 2: Definition of the parameters used for the Kirchhoff-Helmholtz integral.

$$P(\mathbf{r},\omega) = \frac{1}{4\pi}\oint_S \left[P(\mathbf{r}_S,\omega)\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right) - \frac{\partial P(\mathbf{r}_S,\omega)}{\partial \mathbf{n}}\frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|}\right] dS. \quad (1)$$

Figure 2 illustrates the parameters used. In (1), S denotes the surface of an enclosed space V, $|\mathbf{r}-\mathbf{r}_s|$ the vector from a surface point $\mathbf{r}_s$ to an arbitrary listener position $\mathbf{r}$ within the surface S, $P(\mathbf{r}_S,\omega)$ the Fourier transform of the pressure distribution on S, k the wave number and $\mathbf{n}$ the surface normal. The Kirchhoff-Helmholtz integral states that at any listening point within the source-free volume V the sound pressure $P(\mathbf{r},\omega)$ can be calculated if both the sound pressure and its gradient are known on the surface enclosing the volume. This can be used to synthesize a wave field within the surface S by setting the appropriate pressure distribution $P(\mathbf{r}_S,\omega)$, i.e. dipole sources and its gradient, i.e. monopole sources on the surface. This fact is used for WFS based sound reproduction. But there are several simplifications necessary to arrive at a realizable system:

1. Degeneration of the surface S to a plane between the listening area and the reproduction room
2. Degeneration of the surface S to a line
3. Spatial discretization

The first step is to degenerate the surface S to a plane between the listening area and the reproduction room. The wave field can be synthesized by

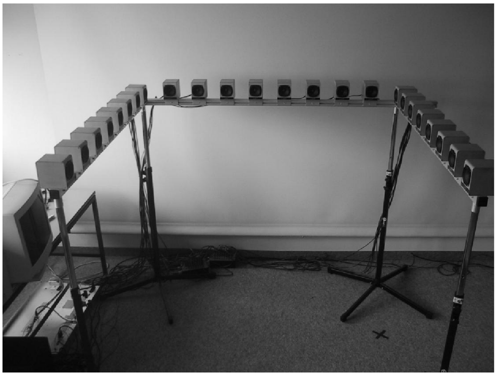


Figure 3: Typical setup of loudspeakers for WFS

either acoustic monopoles or dipoles alone. The Rayleigh I integral describes the mathematics for monopoles as follows

$$P(\mathbf{r},\omega) = \rho c \, \frac{jk}{2\pi} \int \left[v_n(\mathbf{r}_S,\omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_s|}}{|\mathbf{r}-\mathbf{r}_s|} \right] dS \quad (2)$$

where ρ denotes the static density of the air, c the speed of sound and v_n the particle velocity perpendicular to the surface. The Rayleigh II integral [7] applies for dipoles. For our applications it is sufficient to synthesize the wave field correctly in the horizontal ear plane of the listener. For this scenario the surface degenerates further to a line surrounding the listening area in the second step. As a first approximation, closed loudspeakers act as acoustic monopoles mounted on discrete positions. By spatial discretization and assumption of an approximately stationary phase in a third step, the two dimensional Rayleigh I integral (2) can be transformed into one dimension

$$P(\mathbf{r},\omega) = \sum_i \left[A_n(\omega) P(\mathbf{r}_S,\omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|} \right] \Delta x, \quad (3)$$

where $A_n(\omega)$ denotes a weighting factor and Δx denotes the distance between the loudspeakers. Using the above equation, WFS can be realized by mounting closed loudspeakers in a linear fashion (linear loudspeaker arrays) surrounding the listening area leveled with the listeners ears. Figure 3 shows a typical setup.

Up to now we assumed that no acoustic sources lie inside the volume V. The above presented theory can also be extended to the case that sources lie inside the volume V [5]. This allows to place acoustical sources between the listener and the loudspeakers within the reproduction area (focused sources). This is not possible with traditional stereo or 5.1 setups.

In practice, two effects limit the performance of real WFS systems:

1. *Spatial aliasing*
   The discretization of the Rayleigh integral results in spatial aliasing due to spatial sampling. The cut-off frequency is given by [7]

$$f_{\text{al}} = \frac{c}{2\Delta x \sin \alpha_{\max}}, \quad (4)$$

   where $\alpha_{\max}$ denotes the maximum angle of incidence of the synthesized wave field relative to the loudspeaker array. Assuming a loudspeaker spacing $\Delta x = 10\text{cm}$, the minimum spatial aliasing frequency is $f_{\text{al}} = 1700\,\text{Hz}$. Regarding the standard audio bandwidth of 20 kHz spatial aliasing seems to be a problem for practical WFS systems. Fortunately, the human auditory system is not very sensitive to these aliasing artifacts.
2. *Truncation effects*
   These effects are caused by wavefronts which propagate from the ends of the loudspeaker array. They can be understood as diffraction waves caused by the finite number of loudspeakers in practical implementations. Truncation effects can be minimized by filtering in the spatial domain (tapering) [7].

2.2 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. For a point source equation (3) becomes

$$P(\mathbf{r},\omega) = S(\omega) K \sqrt{\frac{jk}{2\pi}} \times \sum_i \left[\frac{e^{-jk|\mathbf{r}_i-\mathbf{r}_m|}}{|\mathbf{r}_i-\mathbf{r}_m|^3} \frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} \right] \Delta x; \quad (5)$$

where $S(\omega)$ denotes the spectrum of the source signal, K is a geometry dependent constant and $\mathbf{r}_m$

denotes the position of the point source. The loudspeaker driving signals Q_i can be derived from (5) as

$$Q_i(\omega) = S(\omega) K \sqrt{\frac{jk}{2\pi}}\ \frac{e^{-jk|\mathbf{r}_i - \mathbf{r}_m|}}{|\mathbf{r}_i - \mathbf{r}_m|^3} \Delta x, \quad (6)$$

By transforming this equation back into the time-domain and employing time discretization the loudspeaker driving signals can be computed from the source signals by delaying, weighting and filtering,

$$q_i[k] = a_n \ (\ h[k] * s[k]\ ) * \delta[k - \kappa], \quad (7)$$

where a_n and κ denote an appropriate weighting factor and delay respectively, $h[k]$ the inverse Fourier transform of $\sqrt{jk/2\pi}$. Multiple (point) 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, which is discussed in a subsequent section.

2.3 Data Based Rendering

The loudspeaker driving signals $Q_i(\omega)$ for arbitrary wave fields can be computed according to equation (3) as follows

$$Q_i(\omega) = A_n(\omega) P(\mathbf{r}_s, \omega). \quad (8)$$

The pressure distribution $P(\mathbf{r}_s, \omega)$ contains the entire information of the sound field produced on the surface S by the sources. By transforming this equation back into the discrete time domain, the loudspeaker driving signals $\mathbf{q}[k]$ can be expressed as a convolution of measured or synthesized impulse responses with the source signals $\mathbf{s}[k]$

$$\mathbf{q}[k] = \mathbf{A}[k] * \mathbf{s}[k], \quad (9)$$

where $\mathbf{A}[k]$ denotes a matrix of impulse responses. 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 [9].

2.4 Loudspeaker and listening room compensation

The theory and rendering methods discussed so far do not take into account that the listening room may have its own acoustic characteristics. Reflections at the surfaces surrounding the reproduction unit interfere with the wave field produced by the loudspeaker array. However, since WFS allows full control over the wave field within the loudspeaker array (at least up to the aliasing frequency), it can be used to compensate for the reflections caused by the listening room as well.

The basic principle of all approaches to listening room compensation is relatively simple: record the reverberation characteristics of the listening room and derive suitable impulse responses for deconvolution of the room response through suitable loudspeaker signals. However, there are a number of pitfalls and practical limitations to consider. At first, the reverberation pattern of a room cannot be completely determined by a few impulse response measurements. Similar considerations as for data based rendering apply. Secondly, the derivation of deconvolution filters is subject to stability restrictions. Finally, the implementation of the correct deconvolution filters may require excessive computing power such that approximations are required.

A method for the inversion of the room acoustics based on microphone array measurements has been presented in [10]. However, it relies on a plane wave assumption and is restricted to linear loudspeaker arrays. To extend this method to arbitrary array geometries, the listening room characteristics have to be recorded with special microphone techniques [9] that capture the entire sound field information (sound pressure and velocity). The sound field caused by the room is then transformed into the frequency domain. The individual loudspeaker filters are then obtained using a multi-channel least squares estimation (LSE) algorithm with a suitable wave field as a desired wave field. This procedure is only applicable for frequencies below the spatial aliasing frequency of the loudspeaker array, because only here we have full control over the produced wave field. Above the aliasing frequency no destructive interference is possible to compensate for room reflections. Above the aliasing frequency individual loudspeaker equalization is used to equalize the frequency response of the loudspeakers. Another possibility is to use an energy control algorithm as presented in [11].

A subset of listening room compensation is the compensation of the loudspeaker characteristics, because here only the magnitude of the frequency response has to be considered. Of course, the measurements have to be performed in an anechoic environment for this case.

3 Implementation

In the course of the CARROUSO project, several WFS systems have been installed at various partner institutions. These systems differ with respect to their needs and research interests. This section gives a short introduction to the CARROUSO project and describes the WFS implementation at the Telecommunications Laboratory at the University of Erlangen-Nuremberg (LNT).

3.1 The CARROUSO project

CARROUSO is founded by the European Commission and consists of 10 partners. The acronym CARROUSO stands for **c**reating, **a**ssessing and **r**endering in **r**eal-time **o**f high-quality a**u**dio-vi**s**ual environments in an MPEG-4 c**o**ntext. The key objective of the CARROUSO project [12, 2] is to provide a new technology that realizes the transfer of a sound field, generated at a certain real or virtual space, to another usually remote located space. Full interactive control of relevant temporal, spatial and perceptual properties of the sound field is included, especially in combination with visual data. The CARROUSO system consists of the following main building blocks:

- *Recording*
  To allow a high degree of freedom for the sound engineer and for user interactivity, the recording side has to capture the dry source signals, source positions (e.g. [13]) and the room impulse responses separately. Dry source recording and source localization is performed by close-up microphones and microphone arrays. As mentioned earlier, the room response is captured by special microphone setups [9].
- *Coding*
  The source signal, source positions and side information (room impulse responses) are encoded using the MPEG-4 standard.
- *Transmission*
  The MPEG-4 stream is multiplexed (optionally with video) by an MPEG-4 server, transmitted and demultiplexed at the receiver side.
- *Decoding*
  The MPEG-4 stream is decoded on the receiver side and the extracted information (source signals, positions, side information) is presented to the renderer.
- *Auralization*
  Auralization is performed with WFS. The source signals are auralized at the correct position with the help of side information (position, impulse responses) by the renderer. Room and loudspeaker compensation is optionally performed at the receiver side.

Figure 4 shows a typical WFS-based sound field recording and reproduction system. The blocks encoding, transmission and decoding are missing in this scheme. Figure 4 shows only those building blocks, which are directly related to the WFS system, i.e. recording and auralization. The recording room contains equipment for dry recording of the primary sources, their positions, and room impulse responses. For reference, the size and position of the listening room relative to the recording room is shown by a dashed line. The recorded signals are encoded, transmitted to the listing room, and decoded. These steps are not shown in Figure 4.

The WFS system in the listening room composes the original acoustic scene by positioning the dry source signals at their respective positions relative to the recording room. Again, the size and position of the recording room relative to the listening room is shown by a dashed line for reference. The virtual sources created by this process may lie outside the listing room. This creates the impression of an enlarged acoustic space on the respective side of the listener. Thus, the signals driving the loudspeakers of the array are synthesized from the source signals by convolution with the respective impulse response. As mentioned above, these impulse responses may not only consider source position and recording room acoustics, but also the characteristics of the listening room acoustics and the loudspeakers.

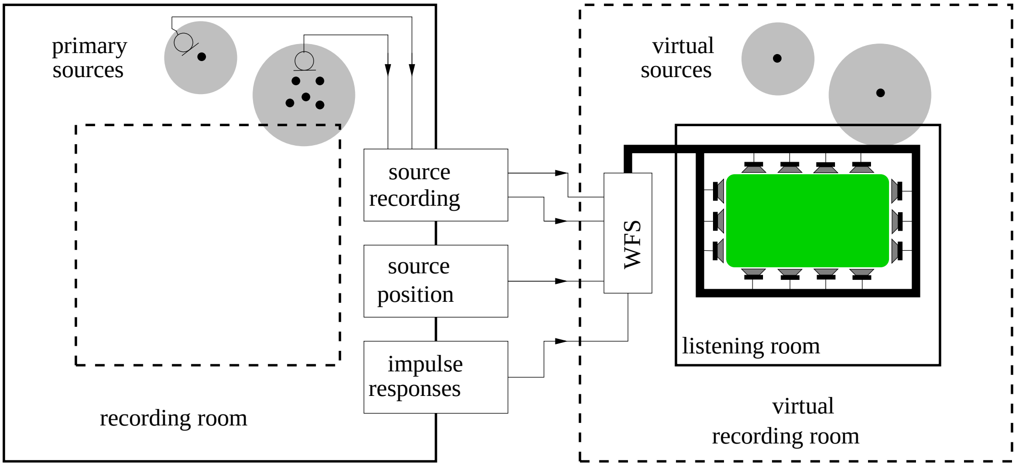


Figure 4: Block diagram of a typical wave field synthesis system

3.2 WFS setup at LNT

The WFS setup at our laboratory consist of 24 wideband loudspeakers and a subwoofer. Figure 3 shows the loudspeaker setup at our laboratory. The loudspeakers are driven by digital amplifiers that were developed within the project for this purpose. The digital amplifiers consist of DA-converters and power amplifiers and use optical digital signals (ADAT) as input. These signals are provided by a digital multi-channel soundcard in a PC. We developed software for model and data based rendering. The software was developed for the LINUX operating system and is running in real-time.

Although still under development, our model based rendering software already provides the following features:

- synthesis of point sources and plane waves
- synthesis of moving point sources with arbitrary trajectories
- interactive graphical user interface for loudspeaker and source setup
- room effects using a mirror image model
- source input from files or ADAT/SPDIF
- simulation of a 5.1 loudspeaker setup

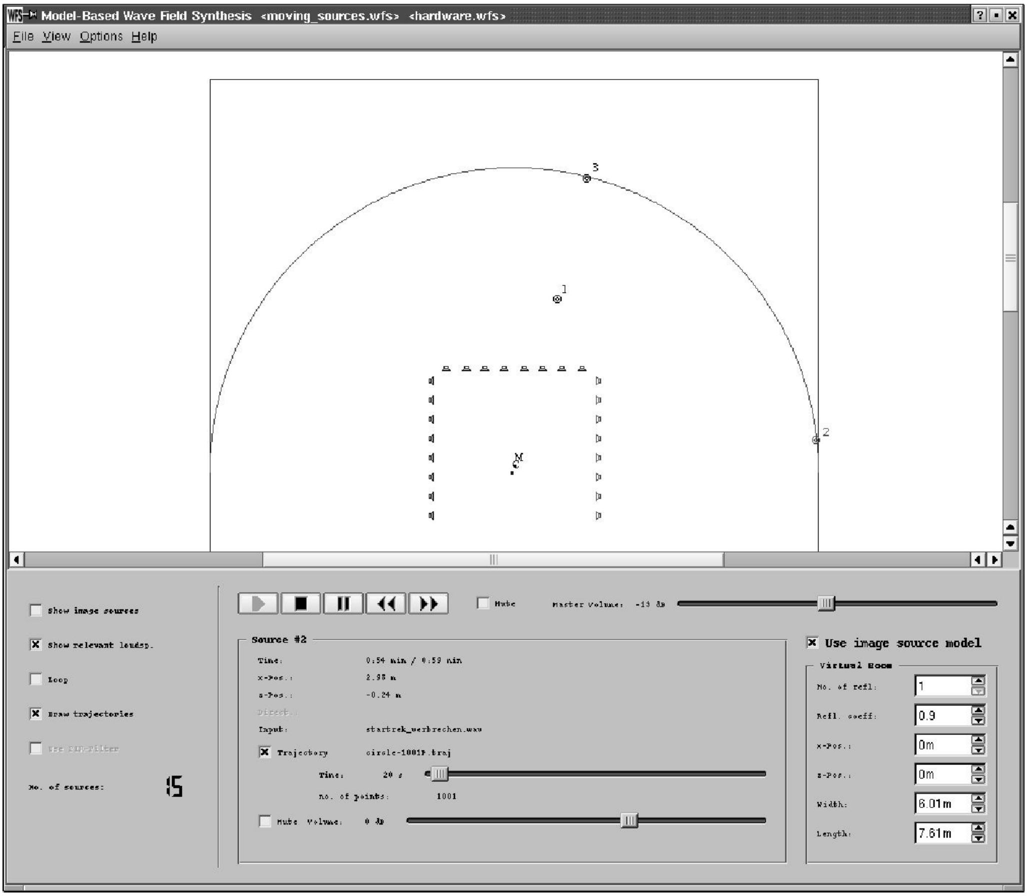
Figure 5 shows a snapshot of the graphical user interface of our model based real-time rendering software. The upper half of the application window shows the loudspeaker and source setup. Sources can be moved intuitively in real-time by clicking on the source and dragging using the computer mouse. One of the shown sources moves on a trajectory which is also displayed in the window. The lower half of the application window controls the synthesis and application parameters and the setup for the virtual room used for the mirror image model. All parameters can be changed in real-time during operation.

For data based rendering we utilize BruteFIR [14], a very fast real-time convolution software. Using a multiprocessor workstation, the computationally complex convolutions can be performed in real-time by our system.

Using the hard- and software described above the rendering features of our system include

- loudspeaker and listening room compensation,
- data based rendering of concert halls/churchs using actual on-site recorded impulse responses,
- model-based rendering of moving sources as described above,
- model-based rendering of virtual loudspeaker positions.

The last feature is not only included for backward compatibility with available two-channel stereo and 5.1 recordings. It also allows to place the two or five virtual speakers at positions outside the listening room. This creates proper two- or five-channel reproduction in small listing spaces not compatible with the recommendations for correct loudspeaker placement, e.g. [1]. Current research at our lab


Model-Based Wave Field Synthesis <moving_sources.wfs> <hardware.wfs>
File View Options Help
3
1
2
M

☐ Show image sources
☒ Show relevant loudsp.
☐ Loop
☒ Draw trajectories
☐ Use FIR-Filter
No. of sources: 5

☐ Mute Master Volume: -13 dB

Source #2
Time: 0:54 min / 0:59 min
x-Pos.: 2.98 m
z-Pos.: -0.24 m
Direct.:
Input: startrek_verbrechen.wav
☒ Trajectory circle-1001P.traj
Time: 20 s
no. of points: 1001
☐ Mute Volume: 0 dB

☒ Use image source model
Virtual Room
No. of refl.: 1
Refl. coeff.: 0.9
x-Pos.: 0m
z-Pos.: 0m
Width: 6.01m
Length: 7.61m

Figure 5: Screenshot of the model based rendering application

is focussed on loudspeaker and room compensation, recording techniques with microphone arrays, source localization and distributed mode loudspeakers (DMLs) for WFS.

4 Conclusions

This paper presented a novel approach to reproduce spatial audio based on the Huygen’s principle. Utilizing a large number of loudspeakers it is possible to recreate a given or prerecorded 3D sound field physically correct. As a result, no sweet spot limits the choice of the listening positions within the listening area. Listeners can move around and turn their head without losing the correct acoustical impression. The WFS approach is not only limited to large installations, even small setups provide great performance improvements over conventional multi-channel setups.

Model-based acoustic rendering can be used to complement visual object-based modeling and rendering with acoustic information to satisfy both the acoustic and visual senses of humans with high quality representations of real or virtual scenes. Data-based acoustic rendering provides high-quality auralization of recorded or simulated acoustic spaces. Using this approach allows to auralize the complete impression of the recorded room. Data visualization can be improved by supplementing or adding information through spatial acoustic representation of the data. Although WFS in general is a quite computationally complex task it can be performed with PC based hardware. Possible applications for WFS include cinema, exhibitions, events, auralization of simulated or predicted acoustics and typical living room applications such as high quality auralization of concerts, simulation of 5.1 setups and computer games.

References

[1] ITU, "Recommendation ITU-R BS.1116-1," 1994-1997.

[2] "The CARROUSO project," http://emt.iis.fhg.de/projects/carrouso.

[3] A.J. Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, December 1988.

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

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

[6] P. Vogel, *Application of Wave Field Synthesis in Room Acoustics*, Ph.D. thesis, Delft University of Technology, 1993.

[7] 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).

[8] 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.

[9] 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).

[10] U. Horbach, A. Karamustafaoglu, R. Rabenstein, G. Runze, and P. Steffen, "Numerical simulation of wave fields created by loudspeaker arrays," in *107th AES Convention*, New York, USA, September 1999, Audio Engineering Society (AES).

[11] E. Corteel, U. Horbach, and R.S. Pellegrini, "Multichannel inverse filtering of multiexciter distributed mode loudspeakers for wave field synthesis," in *112th AES Convention*, Munich, Germany, May 2002, Audio Engineering Society (AES).

[12] S. Brix, T. Sporer, and J. Plogsties, "CARROUSO - An European approach to 3D-audio," in *110th AES Convention*. Audio Engineering Society (AES), May 2001.

[13] S. Spors, R.Rabenstein, and N.Strobel, "A multi-sensor object localization system," *In Vision, Modelling and Visualization (VMV)*, pp. 19–26, 2001.

[14] A. Torger, "BruteFIR - an open-source general-purpose audio convolver," http://www.ludd.luth.se/~torger/brutefir.html.