Perceptual Properties of Data-based Wave Field Synthesis

Sascha Spors [1] and Hagen Wierstorf [2]

[1] Institute of Communications Engineering, Universität Rostock, Germany.

[2] Assesement of IP-based Applications, Technische Universität Berlin, Germany.

Email: Sascha.Spors@uni-rostock.de

Introduction

Wave Field Synthesis (WFS) is a loudspeaker-based auralization technique that aims at the physical synthesis of a sound field within an extended listening area. In order to synthesize the complex sound field of a reverberant enclosure, data-based rendering techniques are employed. Here the spatio-temporal structure of the sound field is captured by a microphone array. The microphone signals are then post-processed in order to compute the loudspeaker driving signals. The limited number of microphones and loudspeakers, equipment noise and sensor mismatch will impair the synthesized sound field in practice. This study investigates the technical and perceptual properties of data-based WFS.

Data-Based Wave Field Synthesis

The loudspeaker driving functions for WFS are given by the directional gradient of the captured sound field evaluated at the loudspeaker positions together with a sensible selection of active loudspeakers [1].

A propagating sound field can be represented as superposition of plane waves with respect to a chosen origin. The directional gradient of a plane wave is given by filtering its spectrum and a weight depending on the angle between the plane wave and the normal vector of the loudspeaker. The origin of the plane wave decomposition can be translated by extrapolating the plane waves. Based on this background it is obvious that a representation of the captured sound field with respect to plane waves is efficient for data-based WFS. The driving functions are computed by extrapolating the plane waves to the respective loudspeaker position, temporal filtering of their spectrum and spatial weighting [2, 1].

By considering the synthesis of a single plane wave and comparison to model-based WFS it becomes clear that data-based WFS can also be realized in an alternative fashion. Synthesizing plane waves with the incidence angles and spectra of the captured and decomposed sound field is equivalent to the procedure outlined above.

Modal Beamforming

Decomposing a captured sound field into plane waves can be understood as spatio-temporal filtering of the microphone signals. Related techniques are often termed as *beamforming* or *plane wave decomposition*. The distribution of microphones on the surface of a sphere is beneficial due to the invariance of the spherical aperture with respect to incidence direction of the captured sound field. Modal beamforming expands the captured sound field into spherical harmonics from which the plane wave decomposition can be computed easily [3]. Modal beamforming provides an elegant analysis technique for data-based WFS. However its practical implementation shows some fundamental limitations that have to be considered carefully. Sampling the sound field on the sphere with a limited number of microphones may lead to spatial aliasing. Furthermore, modal beamforming is an inverse problem which is sensitive to equipment noise, microphone mismatch and placement. The latter aspects are out of the scope of this paper.

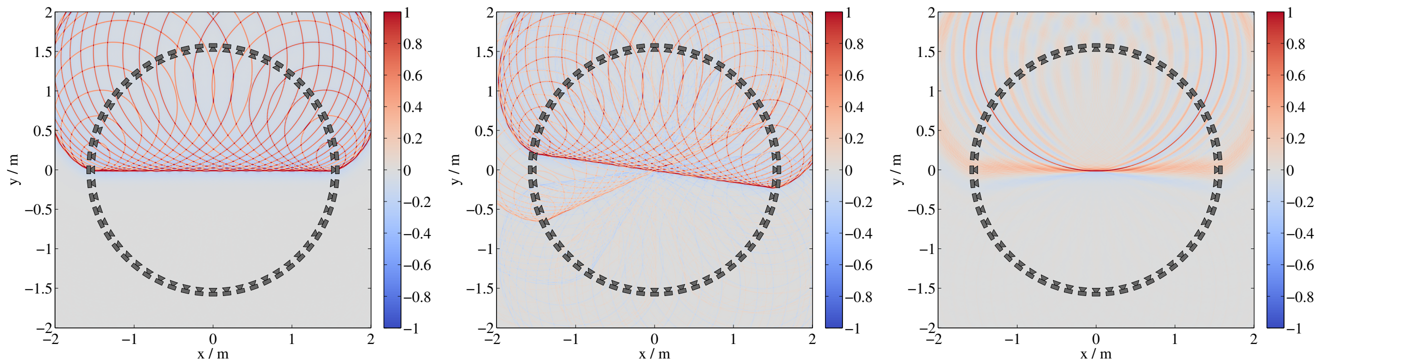
In order to understand the limitations imposed by spatial sampling, the concept of intrinsic dimensionality of a sound field [4] is briefly reviewed. A sound field can be approximated by a limited number of expansion coefficients with bounded error in a bounded region. For a spherical region of radius R, a suitable set of basis functions for expansion is given by the surface spherical harmonics of order N_{sph}. A monochromatic sound field can be approximated reasonably well by order $N_{\mathrm{sph}} = \lceil \frac{eR\omega}{2c} \rceil$, where e denotes the Euler number and c the speed of sound. The number of independent plane waves that can be extracted in this case are $N_{\mathrm{pw}} = 2N_{\mathrm{sph}} + 1$ in two-dimensional scenarios. Spatial sampling limits the intrinsic dimensionality of a sound field that can be captured by a limited number of microphones without prominent spatial aliasing.

Results

As discussed above, modal beamforming using spherical microphone arrays is subject to limitations. The same holds for WFS where the finite number of loudspeakers poses fundamental constraints. In this study we only consider the limited spatial resolution N_{sph} during capture and the limited number L of loudspeakers during synthesis. We furthermore limit our investigations to the capture and synthesis of one broadband plane wave with incidence angle $\phi_{\mathrm{pw}} = 270^{\circ}$ using a 2.5-dimensional circular loudspeaker array with $L = 56$ loudspeakers. The plane wave decomposition using a spatially continuous spherical microphone array for a incident plane wave is given by [3]. Based on the intrinsic dimensionality of the setup, the following conditions have been chosen

N_{sph}	750	28	10	5	1	28
N_{pw}	$2 \cdot N_{\mathrm{sph}} + 1$					1501
L	56					

The first condition is in accordance with the order N_{sph} required for fullband synthesis with an upper frequency of 20 kHz. Bandlimited interpolation has been used to



(a) $N_{\text{sph}} = 750$, $N_{\text{pw}} = 1501$ **(b)** $N_{\text{sph}} = 5$, $N_{\text{pw}} = 11$ **(c)** $N_{\text{sph}} = 28$, $N_{\text{pw}} = 1501$ (Interpolated)

Figure 1: Data-based synthesis of a broadband plane wave with incidence angle $\phi_{\text{pw}} = 270°$ captured by a continuous spherical microphone array of order N_{sph}.

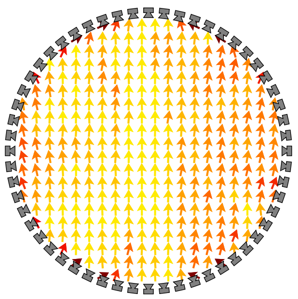


Figure 2: Estimated localization (arrows) for $N_{\text{sph}} = 5$, $N_{\text{pw}} = 11$. The colors denote the deviation from $\phi_{\text{pw}} = 270°$.

compute the plane wave decomposition in the last condition. Figure 1a shows the synthesized sound field for $N_{\text{sph}} = 750$. One plane wave with the desired incidence angle is synthesized. The typical artifacts of WFS due to the limited number of loudspeakers are clearly visible. The synthesized sound field for $N_{\text{sph}} = 5$, a typical order of practical microphone arrays, is shown in Figure 1b. The incident plane wave is decomposed into a small number $N_{\text{pw}} = 11$ of plane waves which do not include the incidence angle. While this may be reasonable close to the center, off-center positions may suffer from the multiple wave fronts incident from multiple directions. Figure 1c shows the sound field for $N_{\text{sph}} = 28$ when $N_{\text{pw}} = 1501$ plane waves have been extracted by means of bandlimited interpolation. The resulting sound field resembles strong similarities with near-field compensated higher-order Ambisonics (HOA) synthesizing a plane wave with the desired incidence angle.

From the simulations the following can be concluded for localization. For data-based synthesis with very high orders N_{sph} localization performance equal or very close to model-based WFS can be expected. For the case of bandlimited interpolation localization performance will be similar to HOA. The localization performance of model-based WFS and HOA has been reported in the literature. For a limited order, localization performance is expected to decrease. In this study the binaural model [5] has been applied to estimate the localization throughout the listening area for $N_{\text{sph}} = 28, 10, 5, 1$. Figure 2 shows the estimated perceived direction for $N_{\text{sph}} = 5$. It can been observed that the localization is quite accurate throughout the listening area. For orders $N_{\text{sph}} < 5$ localization performance dropped considerably.

Besides localization preliminary results have been derived for timbral properties. The frequency response in the center is close to or even better than model-based synthesis of a plane wave with WFS. However, for off-center positions the frequency response is significantly deteriorated for low orders. As a result, coloration is likely to be perceived for off-center positions and low orders.

This work was conducted within the research group '*Simulation and Evaluation of Acoustical Environments (SEACEN)*' supported by grant FOR 1557, Deutsche Forschungsgemeinschaft (DFG).

Reproducible Research

The numerical simulations in this paper are based on the Sound Field Synthesis Toolbox [6]. The scripts used for this paper and listening examples are available at `http://spatialaudio.net/perception-data-based-wfs/`.

References

[1] Sascha Spors. *Active Listening Room Compensation for Spatial Sound Reproduction Systems*. PhD thesis, University of Erlangen-Nuremberg, 2006.

[2] E. Hulsebos, D. de Vries, and E. Bourdillat. Improved microphone array configurations for auralization of sound fields by Wave Field Synthesis. *Journal of the Audio Engineering Society (AES)*, 50(10), Oct. 2002.

[3] B. Rafaely. Phase-mode versus delay-and-sum spherical microphone array processing. *IEEE Signal Processing Letters*, 12(10):713–716, 2005.

[4] R.A. Kennedy, P. Sadeghi, T.D. Abhayapala, and H.M. Jones. Intrinsic limits of dimensionality and richness in random multipath fields. *IEEE Transactions on Signal Processing*, 55(6):2542–2556, 2007.

[5] Hagen Wierstorf, Alexander Raake, and Sascha Spors. Binaural assessment of multi-channel reproduction. In Jens Blauert, editor, *The Technology of Binaural Listening*, pages 255–278. Springer, 2013.

[6] Hagen Wierstorf and Sascha Spors. Sound field synthesis toolbox. In *132nd Convention of the Audio Engineering Society*, April 2012.