Audio Engineering Society

Convention Paper

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Sound Field Reconstruction: An Improved Approach For Wave Field Synthesis

Mihailo Kolundžija[1], Christof Faller[1], and Martin Vetterli[12]

[1] *Audiovisual Communications Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

[2] *Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA*

Correspondence should be addressed to Mihailo Kolundžija (mihailo.kolundzija@epfl.ch)

ABSTRACT

Wave field synthesis (WFS) is a prevalent approach to multiple-loudspeaker sound reproduction for an extended listening area. Although powerful as a theoretical concept, its deployment is hampered by practical limitations due to diffraction, aliasing, and the effects of the listening room. Reconstructing the desired sound field in the listening area, accounting for the medium propagation characteristic, is another approach termed as sound field reconstruction (SFR). It is based on the essential band-limitedness of the sound field, which enables a continuous matching of the reconstructed and the desired sound field by their matching on a discrete set of points spaced below the Nyquist distance. We compare the two approaches in a common single-source free-field setup, and show that SFR provides improved sound field reproduction compared to WFS in a wide listening area around a defined reference line.

1. INTRODUCTION

The first spatial sound reproduction systems date back to the work of Blumlein [1] on stereo systems in the first half of the last century. The successful two-channel stereo principle—still used widely today—was extended to four-channel quadraphonic system [2] with the aim of providing full-circle spatial reproduction, but it was quickly abandoned due to its poor sound localization performance in the front half-plane. Surround systems using higher number of channels, such as 5.1 and 7.1, are based on the observation that accurate localization of the sound coming from the front is more important, and they use more loudspeaker in front of the listener for improved frontal localization. Additionally, loudspeakers on the side and behind the listening position are used for providing ambience and side/rear localization.

The downside of all previously mentioned surround sound systems comes from the fact that the sound field provides correct localization and surround experience only in a single "sweet spot" position, and the sound field outside of the "sweet spot" can not be controlled.

The problem of extending the listening area was addressed by two notable surround sound systems: Ambisonics (e.g., see [3, 4]) and Wave Field Synthesis (e.g., see [5, 6]). Both approaches are based on an attempt to reproduce a *desired sound field*—and consequently provide correct localization—in an extended listening area.

Ambisonics is based on the decomposition of the sound field into orthogonal functions, such as circular or spherical harmonics, in a single point. The idea is that matching the orthogonal components of the desired and the reproduced sound field makes them matched in an extended listening area. The reproduction setup is not fixed, and various loudspeaker configurations can be used.

Wave field synthesis systems, on the other hand, are based on the Huygens principle, which shows that a desired sound field in a closed listening area can be reproduced by a continuous distribution of secondary sources on a closed surface around that listening area. The reproduction setup can take a form of planar or linear loudspeaker arrays, or a combination thereof.

However, both Ambisonics and WFS have limited sound field reproduction capabilities that arise from discrepancies between their theoretical model and systems in practice. Namely, while the theory assumes continuous and infinite reproduction setups, practical systems are both finite and discrete, causing impairments due to diffraction and aliasing.

This paper presents a different approach, denoted as *sound field reconstruction* (SFR), that is based on the spatio-temporal spectral properties of the wave equation kernel—the plenacoustic function [7]—and in particular on the essential band-limitedness of the sound field that emanates from sources of limited temporal spectrum [7]. Similarly to the ambisonic systems, this approach does not assume a fixed reproduction setup, and it aims at matching the reproduced and the desired sound field inside the listening area. However, unlike the ambisonic systems, which aim at matching spherical harmonics of the desired and the reproduced sound field in a single point, this approach aims at matching the reproduced and the desired sound field on a grid of points covering the listening area, that satisfies the Nyquist sampling criterion relative to a desired frequency range for accurate reconstruction.

The paper is organized as follows. Section 2 briefly presents the essentials of wave field synthesis. Section 3 describes the proposed approach for sound field reproduction based on spatio-temporal spectral properties of the plenacoustic function. Section 4 compares the two spatial sound reproduction approaches at reproducing a sound field of a point source having different locations and temporal frequencies. Conclusions are presented in Section 5.

2. WAVE FIELD SYNTHESIS

Wave field synthesis (WFS) is a spatial sound reproduction technique which is based on the Huygens principle and its mathematical description expressed through Kirchoff-Helmholtz and Rayleigh integrals [5, 6].

Huygens principle expressed through the Kirchoff-Helmholtz integral states that for reproducing a sound field in a source-free listening area, it is sufficient to use continuous distributions of monopole and dipole secondary sources on a closed surface enclosing the listening area.

The Rayleigh I integral, on the other hand, states that under certain conditions (i.e., when the Green's function satisfies the Sommerfeld condition), reproduction of a sound field is possible with the use of a continuous distribution of monopole sources on an infinite plane that separates the reproduced primary sources from the listening area. Denoting by S the surface with the secondary monopole source distribution, the Rayleigh I integral gives the relation between the sound pressure $P(\mathbf{r}, \omega)$ in the listening area and the particle velocity vector's normal component $V_n(\mathbf{r}_S, \omega)$ on the surface S,

$$P(\mathbf{r}, \omega) = \rho_0 c \frac{jk}{2\pi} \int V_n(\mathbf{r}_S, \omega) \frac{e^{-jk|\mathbf{r}\text{-}\mathbf{r}_S|}}{|\mathbf{r}\text{-}\mathbf{r}_S|} \, \mathrm{d}S \,, \quad (1)$$

where ρ_0 is the density of air, c is the speed of sound, ω is the temporal frequency, and k the wave number given by $\frac{\omega}{c}$. Physically, the right side of (1) can

be interpreted as a sound field of a distribution of secondary point sources driven by a signal $Q(\mathbf{r}_S, \omega)$, given by

$$Q(\mathbf{r}_S, \omega) = \rho_0 c \frac{jk}{2\pi} V_n(\mathbf{r}_S, \omega)\,, \tag{2}$$

that is proportional to the component of the particle velocity vector that is normal to the surface S.

Derivation of the loudspeaker driving signals for practical WFS systems requires the following assumptions and simplifications:

- Primary sources, secondary sources, and listening area are situated in the same plane.
- The integral is approximated with a discrete sum.
- Planar source distribution is replaced with a linear source distribution in order to reduce the number of used secondary sources, which effectively collapses a linear distribution of *line sources* to a linear distribution of *point sources*.
- The secondary sources' driving signals are computed for approximately correct reconstruction on a reference listening line.

Fig. 1 shows a WFS setup with a linear loudspeaker array that is used for reconstructing the sound field of a primary monopole source on a reference listening line. After simplification with a line integral and discretization, the Rayleigh I integral becomes

$$P(\mathbf{r}, \omega) = \sum_i \tilde{Q}_i(\omega) \frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\,, \tag{3}$$

where $\mathbf{r}_i$ denotes the position of the i-th monopole secondary source (loudspeaker). The secondary source driving signals $\tilde{Q}_i(\omega)$ are given by [8]

$$\tilde{Q}_i(\omega) = \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{|x_l - x_i|}{|x_l - x_m|}} S(\omega) \frac{e^{-jk|\mathbf{r}_i - \mathbf{r}_m|}}{\sqrt{|\mathbf{r}_i - \mathbf{r}_m|}} \cos\theta_i \Delta y\,, \tag{4}$$

where x_l, x_i and x_m are the x coordinates of the reference listening line, the line containing the secondary sources, and the primary point source, respectively; $S(\omega)$ is the primary source's spectrum,

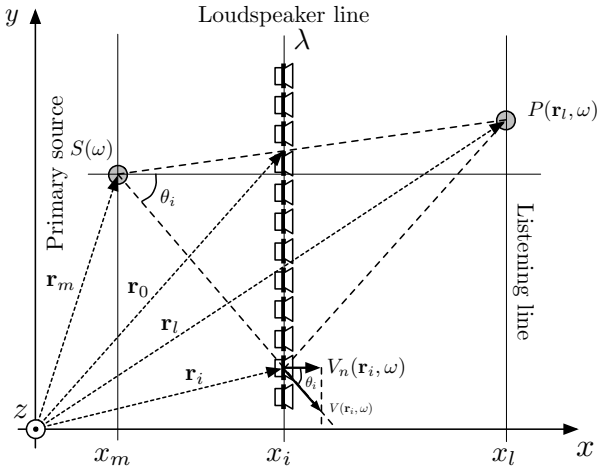


Fig. 1: Setup for the calculation of the line array loudspeaker signals for reconstructing the sound field of a primary monopole source on the listening line.

θ_i the angle between the axis x and the particle velocity vector $\mathbf{V}(\mathbf{r}_i, \omega)$ in the point $\mathbf{r}_i$, and Δy the spacing between secondary sources.

The main limitations of WFS can be summarized as follows:

- *Amplitude errors* that stem from approximating planar secondary source distribution with a linear one.
- *Truncation effects*, such as the reduction of the area of correct reproduction, resulting from approximating infinitely long secondary source distributions with finite apertures.
- *Aliasing artifacts*, resulting from a coarse spacing of secondary point sources. Namely, a correct reconstruction with practical WFS systems can be achieved only up to a maximum frequency f_m given by

$$f_m = \frac{c}{2\Delta y \sin\alpha_{max}}\,, \tag{5}$$

where c is the speed of sound, Δy the spacing between secondary sources, and $\alpha_{\max}$ the maximum incidence angle between the secondary source line and the particle velocity vector of the reproduced sound field.

3. SOUND FIELD RECONSTRUCTION

Sound field reconstruction (SFR) is a spatial sound reproduction approach which is based on the spectral properties of the plenacoustic function (i.e., the kernel of the wave equation) shown by Ajdler *et al.* [7], and more particularly, on the essential band-limitedness of the sound field that emanates from a band-limited sound source. This section provides a description of sampling and interpolation of the plenacoustic function, and shows how they can be used for sound field reconstruction with an arbitrary reproduction setup.

3.1. Plenacoustic sampling and interpolation

The plenacoustic function $h_{\mathbf{r}'}(\mathbf{r}, t)$ is the kernel of the wave equation, or equivalently, the spatio-temporal impulse response of the acoustical medium from point $\mathbf{r}'$ to point $\mathbf{r}$. The sound field is obtained as a result of convolution between the plenacoustic function and the distribution of sound sources.

The work of Ajdler *et al.* [7] showed that the plenacoustic function for a point source at position $\mathbf{r}'$ and for a particular temporal frequency ω is essentially band-limited in spatial frequency ϕ. In other words, the changes of sound in space can not be arbitrarily fast, but are limited by its temporal frequency according to the relation

$$\phi \leq \omega / c \,, \tag{6}$$

where c is the speed of sound propagation.

Based on the observation given in (6), one can define a minimum spatial sampling frequency for a sound field of limited temporal bandwidth. If the maximum temporal frequency of the sources evoking the sound field is equal to ω_m, then the spatial sampling frequency of $\phi_s = 2\omega_m/c$ is sufficient for representing the sound field.

Additionally, the authors of [7] showed that even by sampling the sound field on a spatial segment of finite size (e.g., on a finite-length line segment), one can reconstruct the sound field inside that spatial segment with a high accuracy.

The result on the possibility of sampling a sound field has an implication that is useful in the context of SFR. Namely, SFR can be done with respect to a grid of points within the listening area, and the outcome will be a correct reconstruction in the entire listening area. This result is formalized with the following proposition:

Proposition 1. *If two functions $f(\mathbf{r}, t)$ and $h(\mathbf{r}, t)$, both band-limited with the maximum temporal frequency ω_m and the maximum spatial frequency ϕ_m ($\phi_m = \omega_m/c$), are identical on a grid satisfying Nyquist criterion with respect to ω_m and ϕ_m, then they are identical everywhere.*

Proof. The proof follows from the fact that a band-limited function (e.g., a sound field) is uniquely defined by its samples on a grid satisfying the Nyquist criterion. Since the functions $f(\mathbf{r}, t)$ and $h(\mathbf{r}, t)$ are both band-limited with the same spectral support, and have identical values on a sampling grid satisfying the Nyquist criterion, they must be identical everywhere. □

Based on this observation, it can be seen that a correct coverage of the listening area with control points and a reconstruction of the sound pressure signals in these control points is sufficient for a complete sound field reconstruction inside the listening area.

It should also be noted that even though Proposition 1 treats infinite grids, the result from [7] on the interpolation error in a limited spatial region ensures a small error in the limited-area reproduction case. This will be shown in Section 4.

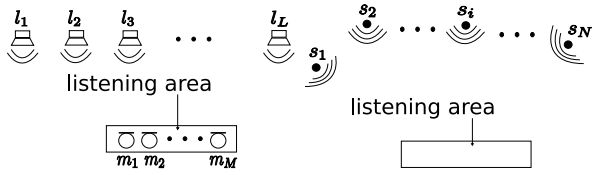
3.2. SFR using multichannel inversion

Controlling the sound pressure signal on a grid of discrete points with multiple loudspeakers is a multiple-input multiple-output (MIMO) channel inversion problem. Consider a sound reproduction setup that contains L loudspeakers and M control points placed in the listening area, as shown in Fig. 2(a). Furthermore, there is a desired acoustic scene that contains N sound sources. The goal of the MIMO channel inversion is the reproduction of the desired sound field in M control points using a given reproduction setup.

Positions of loudspeakers, control points, and the desired sources are known. The impulse responses $a_{ij}(t)$ of the sound propagation channel between the jth desired source and the ith control point are also known, through computation or measurements. The impulse responses $g_{ij}(t)$ of the sound propagation channel between the jth loudspeaker and the ith

control point are also known (can be computed or measured).

As previously mentioned, the goal of the multichannel inversion is reproducing the desired sound scene at the listening locations, i.e. determining the loudspeaker driving signals that evoke the same signals at the control points as the original sound scene.



(a) Reproduction setup (b) Desired sound scene

Fig. 2: Multichannel inversion problem overview.

Denote by $s_i(t)$, $x_j(t)$, and $y_k(t)$ the signals of the ith desired source, the output of the jth loudspeaker, and the sound pressure at the kth control point, respectively. Furthermore, denote by $d_l(t)$ the signal at the lth control point in the desired sound scene.

The signals $d_i(t)$ are determined by the effects of the sound propagation paths from the desired sources to the control points, and are described by the following convolution in the time domain:

$$\mathbf{d}(t) = \mathbf{A}(t) * \mathbf{s}(t)\,, \tag{7}$$

where

$$\begin{aligned}\mathbf{d}(t) &= [d_i(t)]_{M\times 1}\\ \mathbf{A}(t) &= [a_{ij}(t)]_{M\times N}\\ \mathbf{s}(t) &= [s_i(t)]_{N\times 1}\,.\end{aligned}$$

On the other hand, the signals reproduced at the control points are determined by the sound propagation paths' effects on the loudspeaker signals, and are given by

$$\mathbf{y}(t) = \mathbf{G}(t) * \mathbf{x}(t)\,, \tag{8}$$

where

$$\begin{aligned}\mathbf{y}(t) &= [y_i(t)]_{M\times 1}\\ \mathbf{G}(t) &= [g_{ij}(t)]_{M\times L}\\ \mathbf{x}(t) &= [x_i(t)]_{L\times 1}\,.\end{aligned}$$

The task of the multichannel inversion is computing the signals $x_j(t)$ using the desired signals $s_i(t)$, i.e.,

$$\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t)\,, \tag{9}$$

where

$$\mathbf{H}(t) = [h_{ij}(t)]_{L\times N}\,,$$

such that the signal vectors $\mathbf{d}(t)$ and $\mathbf{y}(t)$ are equal up to a constant delay Δ.

Multichannel inversion solution The problem of multichannel inversion can be represented as a superposition of N sub-problems, each involving a single desired source. By combining (7), (8), and (9), the following equalities are obtained for $i = 1,\ldots,N$:

$$\mathbf{G}(t) * \mathbf{h_i}(t) = \mathbf{a_i}(t-\Delta)\,, \tag{10}$$

where $\mathbf{h}_i(t) = [h_{1i}(t)\ldots h_{Li}(t)]^T$ and $\mathbf{a}_i(t) = [a_{1i}(t)\ldots a_{Mi}(t)]^T$ are the i-th columns of the matrices $\mathbf{H}(t)$ and $\mathbf{A}(t)$, respectively.

Taking the Fourier transform of both sides of (10) gives the polynomial matrix equality

$$\mathbf{G}(\omega)\mathbf{h}_i(\omega) = e^{-j\omega\Delta}\mathbf{a}_i(\omega)\,, \tag{11}$$

with

$$\begin{aligned}\mathbf{G}(\omega) &= [G_{ij}(\omega)]_{M\times L}\\ \mathbf{h}_i(\omega) &= [H_{ji}(\omega)]_{L\times 1}\\ \mathbf{a}_i(\omega) &= [A_{ji}(\omega)]_{M\times 1}\,.\end{aligned}$$

Having a general multichannel inversion solution for a given reproduction setup and a given set of control points requires that (11) have a solution for an arbitrary desired sound scene, i.e., for an arbitrary choice of sound propagation impulse responses $A_{ij}(\omega)$. The general solution exists if the sound propagation impulse response matrix $\mathbf{G}(\omega)$ has a right inverse polynomial matrix denoted as $\mathbf{G}^*(\omega)$.

Given the right inverse matrix $\mathbf{G}^*(\omega)$, the inversion filters at all loudspeakers can be determined for a set of sound propagation impulse responses $\mathbf{a}_i(\omega)$ from an arbitrary desired source:

$$\mathbf{h}_i(\omega) = e^{-j\omega\Delta}\mathbf{G}^*(\omega)\mathbf{a}_i(\omega)\,. \tag{12}$$

For the existence of the matrix $\mathbf{G}^*(\omega)$, it is necessary that the number of loudspeakers be greater than the number of control points ($L > M$) [9, 10]. Furthermore, a sufficient and necessary condition for the existence of $\mathbf{G}^*(\omega)$ is that the matrix $\mathbf{G}(\omega)$ be full-rank for every ω, or equivalently, its Smith canonical form should be equivalent to a constant matrix $[\mathbf{I}_M \; \mathbf{0}_{(L-M)\times M}]$, where $\mathbf{I}_M$ is an identity matrix of size $M \times M$, and $\mathbf{0}_{(L-M)\times M}$ is an all-zero matrix of size $(L-M) \times M$ [9, 11].

Approximate multichannel inversion The exact MIMO inversion conditions are too restrictive and impractical, since they require high number of loudspeakers (higher than the number of control points) and still do not provide any guarantee that the exact inversion is possible for a given loudspeaker setup containing the necessary number of loudspeakers.

A more practical approach to MIMO channel inversion involves using less loudspeakers than control points, and finding the loudspeaker driving signals by minimizing the reconstruction mean squared error (MSE) in the control points. It should also be noted that the MIMO channel inversion in SFR is done with respect to the control points spaced at the Nyquist distance, which guarantees that minimizing the MSE in a discrete set of points will result in minimizing the MSE in the entire listening area.

The solution to the MIMO inversion by MSE minimization is the pseudo-inverse of the transfer matrix $\mathbf{G}(\omega)$. However, the pseudo-inverse expression given by

$$\mathbf{G}^+(\omega) = \left(\mathbf{G}^H(\omega)\mathbf{G}(\omega)\right)^{-1}\mathbf{G}^H(\omega)\,, \tag{13}$$

where the matrix $\mathbf{G}^H(\omega)$ is the conjugate-transpose of the matrix $\mathbf{G}(\omega)$, is not correct if the matrix $\mathbf{G}(\omega)$ is rank-deficient, and the numerical solution of (13) gives filters with excessive gains beyond the practical limitations for common loudspeakers.

The work of Kirkeby *et al.* [12] addresses the problem of excessive loudspeaker filter gains of the pseudo-inverse approach to MIMO channel inversion by using a regularization term in the pseudo-inverse computation, which enables trading-off between the reconstruction accuracy and the effort (energy used for the MIMO channel inversion). This approach to MIMO channel inversion was later used by a number of authors (e.g., see [13, 14]).

Since the regularization of the pseudo-inverse does not allow for an easy control of the trade-off between the reconstruction precision and filters' gains, the SFR uses the pseudo-inversion based on the singular value decomposition (SVD) and pruning of the singular values which are below a defined threshold (e.g., see Golub *et al.* [15]). In particular, if

$$\mathbf{G}(\omega) = \mathbf{U}(\omega)\mathbf{\Sigma}(\omega)\mathbf{V}^H(\omega) \tag{14}$$

is the SVD of the matrix $\mathbf{G}(\omega)$, then the pseudo-inverse of the matrix $\mathbf{G}(\omega)$ is given by

$$\mathbf{G}^+(\omega) = \mathbf{V}(\omega)\mathbf{\Sigma}^+(\omega)\mathbf{U}^H(\omega)\,, \tag{15}$$

where the matrix $\mathbf{\Sigma}^+(\omega)$ is obtained from $\mathbf{\Sigma}(\omega)$ by first setting to zero the singular values whose absolute values are below a defined threshold ϵ, replacing the other singular values by their reciprocal, and taking the matrix transpose in the end [15].

Using (12), the loudspeaker driving signals are given by

$$\mathbf{h}_i(\omega) = e^{-j\omega\Delta}\mathbf{G}^+(\omega)\mathbf{a}_i(\omega)\,, \tag{16}$$

where Δ is a time-delay used to meet the physical constraints of the reproduction setup.

4. SIMULATIONS

In order to assess the performance of the SFR method, it was compared with WFS at reproducing a sound field stemming from a point source located at different positions:

- Point source far behind the loudspeaker array, such that the field in the listening area resembles a plane-wave sound field.
- Point source closely behind the loudspeaker array, such that the field in the listening area exhibits a clear spherical-wave behavior.

4.1. Simulation setup

The simulation of WFS and SFR used the same 6m-long line loudspeaker array containing five loudspeakers. The loudspeaker array was positioned at $x_i = 4$ m and the loudspeakers were spaced at a distance $\Delta y = 1.5$ m from each other, starting from

$y_0 = 0$ and extending to $y_1 = 6$ m. This coarsely-spaced loudspeaker array was chosen in order to enable plotting wave fields at frequencies where WFS systems are aliased while the wavelengths are long enough such that details are visible in the figures.

Both loudspeakers and primary (reproduced) sources were simulated as point sources, such that the propagation characteristic between the source and any point in space is given by

$$G(\omega) = \frac{e^{-j\omega r/c}}{r}, \qquad (17)$$

where r is the distance of the given point to the point source, and c is the speed of sound.

The used reproduction reference line was positioned at $x_l = 8$ m. It was four meters long, extending from $y_2 = 1$ m to $y_3 = 5$ m. Control points on the reference listening line were spaced at $\Delta_l = 17$ cm to allow a correct reproduction with SFR up to the frequency $f_m = 1$ kHz. It should be noted that the loudspeaker array length limits the effective length of the reproduction area, and that making the reference line slightly shorter than the loudspeaker array makes the SFR better conditioned.

The simulation experiments with SFR and WFS included comparing snapshots of the two reproduced sound fields in an extended listening area, comparing loudspeaker filters' gains used by the two approaches, and comparing the reconstruction error on two 6m-long control lines parallel to the loudspeaker array, positioned at $x_1 = 8$ m (the reference line) and $x_2 = 10$ m, and extending from $y_0 = 0$ to $y_1 = 6$ m.

In all the simulations, WFS loudspeaker driving signals were computed using (4), and SFR loudspeaker driving signals were computed using (16) and (9), with $\epsilon = 0.001$.

4.2. Point source far behind loudspeakers: non-aliased case

Far-away point sources are interesting since the sound field evoked by them resembles a plane wave in the listening area. In this simulation, the primary point source was placed at position $\mathbf{r}_m = (-20\text{m},\ 1\text{m})$, and the two approaches were compared for the primary source having the frequency $f_1 = 100$ Hz.

Fig. 3 shows snapshots of the desired sound field and sound fields reconstructed with WFS and SFR, and compares loudspeaker filter gains of the two approaches.

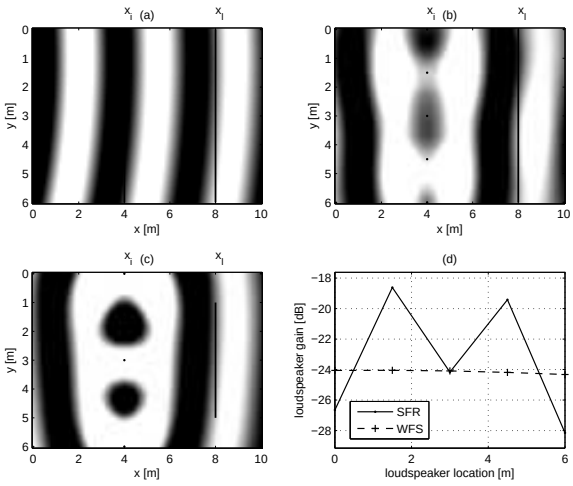


Fig. 3: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (-20\text{m},\ 1\text{m})$, with frequency $f_1 = 100$ Hz: snapshot of the desired sound field (a), snapshot of the sound field obtained with WFS (b), and snapshot of the sound field obtained with SFR (c); loudspeaker filter gains used by WFS and SFR (d).

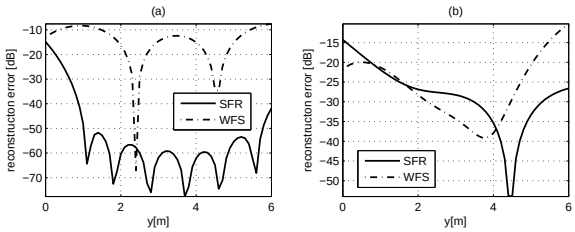


Fig. 4: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (-20\text{m},\ 1\text{m})$, with frequency $f_1 = 100$ Hz: reconstruction error normalized by the desired field's amplitude on the control line at $x_1 = 8$ m (a), and on the control line at $x_2 = 10$ m (b).

Fig. 4 compares reconstruction errors of the two approaches, normalized by the desired field's amplitudes, on the two previously mentioned control lines ($x_1 = 8$ m and $x_2 = 10$ m).

From Fig. 3, it can be seen that in the listening area around the reference line, the reproduced sound fields of both approaches closely resemble the desired

sound field. The sound field obtained with SFR is slightly closer to the desired field, but at the expense of having to use higher (by around 6 dB) loudspeaker gains at two central loudspeakers. It should be noted that at the reference line, as observed from Fig. 4, SFR is significantly more precise than WFS.

4.3. Point source closely behind loudspeakers: non-aliased case

Sources closely behind the loudspeaker array evoke a sound field that exhibits more variation in the wavefront's shape and amplitude decay within the listening area. In this simulation, the primary point source was placed at position $\mathbf{r}_m = (2\text{m},\ 1\text{m})$, and the two approaches were compared for the primary source of the frequency $f_1 = 100$ Hz.

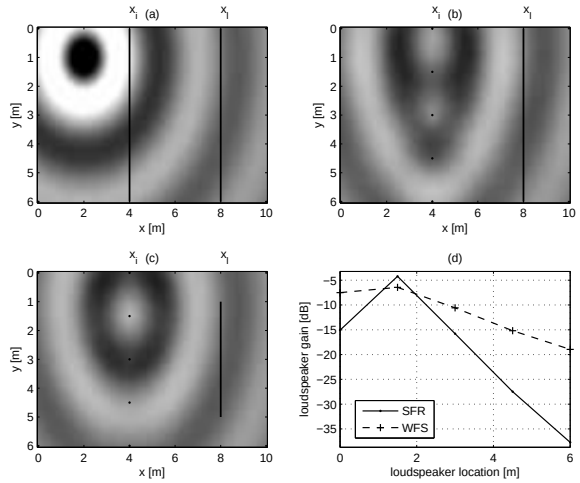


Fig. 5: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (2\text{m},\ 1\text{m})$, with frequency $f_1 = 100$ Hz: snapshot of the desired sound field (a), snapshot of the sound field obtained with WFS (b), and snapshot of the sound field obtained with SFR (c); loudspeaker filter gains used by WFS and SFR (d).

As previously, Fig. 5 shows snapshots of the desired sound field and sound fields reconstructed with WFS and SFR, and compares loudspeaker filter gains of the two approaches. Fig. 6 compares reconstruction errors of the two approaches, normalized by the desired field's amplitudes, on the two control lines ($x_1 = 8$ m and $x_2 = 10$ m).

From Fig. 5, it can be seen that in the listening area around the reference line, the reproduced sound

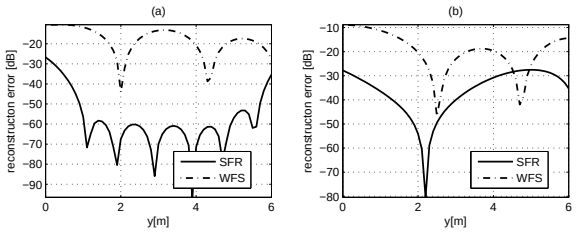


Fig. 6: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (2\text{m},\ 1\text{m})$, with frequency $f_1 = 100$ Hz: reconstruction error normalized by the desired field's amplitude on the control line at $x_1 = 8$ m (a), and on the control line at $x_2 = 10$ m (b).

fields of both approaches closely resemble the desired sound field. The sound field obtained with SFR is again closer to the desired field (this time more notably) than the field reconstructed with WFS, which could be seen more clearly from Fig. 4. Except for one loudspeaker with a gain higher by less than 3 dB than the corresponding WFS loudspeaker, the loudspeakers in SFR use significantly lower gains.

4.4. Point source far away behind loudspeakers: aliased case

Apart from the reproduction of low-frequency point sources, it was interesting to compare WFS and SFR at higher frequencies, since it is known that WFS experiences aliasing distortions when reproducing sound sources with frequencies higher than the maximum frequency given by (5). Since the simulated loudspeaker setup was intentionally made sparse, WFS becomes aliased at frequencies slightly higher than 100 Hz. This simulation experiment was thus made with a point source located far behind the loudspeaker array (in order to produce a sound field which resembles a plane wave) at position $\mathbf{r}_m = (-20\text{m},\ 3\text{m})$ and with the frequency $f_2 = 300$ Hz.

As with previous experiments, Fig. 7 shows snapshots of the desired sound field and sound fields reconstructed with WFS and SFR, and compares loudspeaker filter gains of the two approaches. Fig. 8 compares reconstruction errors of the two approaches, normalized by the desired field's amplitudes, on the two control lines ($x_1 = 8$ m and $x_2 = 10$ m).

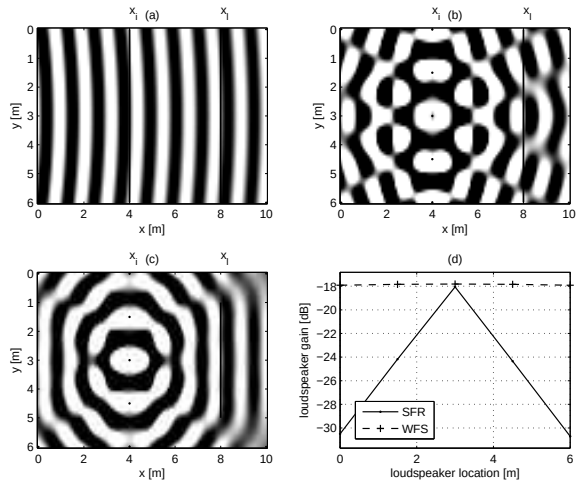


Fig. 7: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (-20\text{m},\ 3\text{m})$, with frequency $f_2 = 300$ Hz: snapshot of the desired sound field (a), snapshot of the sound field obtained with WFS (b), and snapshot of the sound field obtained with SFR (c); loudspeaker filter gains used by WFS and SFR (d).

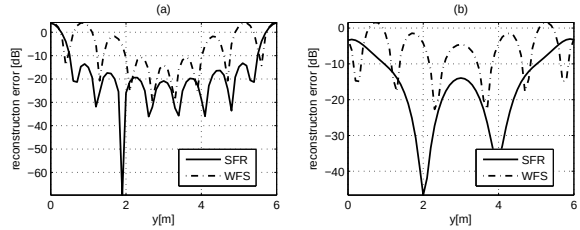


Fig. 8: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (-20\text{m},\ 3\text{m})$, with frequency $f_2 = 300$ Hz: reconstruction error normalized by the desired field's amplitude on the control line at $x_1 = 8$ m (a), and on the control line at $x_2 = 10$ m (b).

The simulations shown in Fig. 7 imply that the sound field reconstructed with WFS has significant aliasing artifacts in the entire listening area, whereas the sound field reconstructed with SFR closely resembles the desired sound field in the major part of the listening area around the reference line, with artifacts occurring only on boundaries of the listening area. The reconstruction error, shown in Fig. 8, additionally confirms the observations from Fig. 7. It should also be added that SFR achieves the superior sound field reconstruction by using *lower* loudspeaker gains.

4.5. Point source closely behind loudspeakers: aliased case

As with a far-away point source, this simulation compared the two reproduction approaches for a point source having a frequency for which WFS is aliased with the used loudspeaker array. The position of the primary point source in this simulation was $\mathbf{r}_m = (2\text{m},\ 1\text{m})$, and its frequency was $f_3 = 600$ Hz.

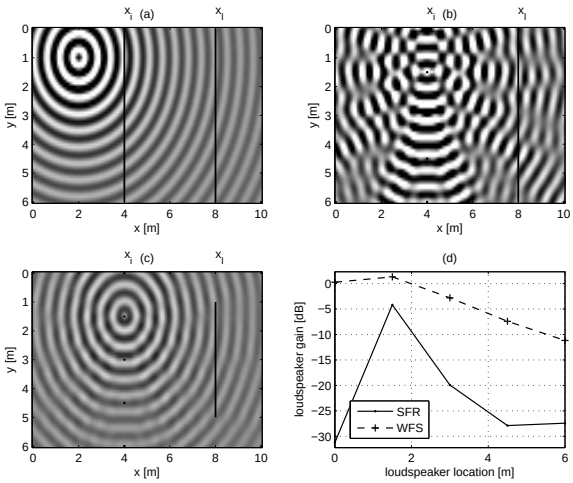


Fig. 9: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (2\text{m},\ 1\text{m})$, with frequency $f_3 = 600$ Hz: snapshot of the desired sound field (a), snapshot of the sound field obtained with WFS (b), and snapshot of the sound field obtained with SFR (c); loudspeaker filter gains used by WFS and SFR (d).

Fig. 9 shows snapshots of the desired sound field and sound fields reconstructed with WFS and SFR, and compares loudspeaker filter gains of the two approaches. Fig. 10 compares reconstruction errors of the two approaches, normalized by the desired field's amplitudes, on the two control lines ($x_1 = 8$ m and $x_2 = 10$ m).

Fig. 9 implies that the sound field reconstructed with WFS has high aliasing artifacts, whereas the sound field reconstructed with SFR closely resembles the desired field in the listening area around the reference line. The significant reconstruction accuracy

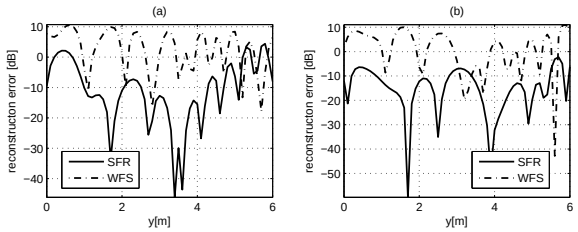


Fig. 10: Comparison of WFS and SFR at reproducing a point source located at $\mathbf{r}_m = (2\text{m}, 1\text{m})$, with frequency $f_3 = 600$ Hz: reconstruction error normalized by the desired field's amplitude on the control line at $x_1 = 8$ m (a) and on the control line at $x_2 = 10$ m (b).

advantage of SFR compared to WFS is confirmed by the reconstruction error plots in Fig. 10. Additionally, as in the previous experiment, SFR achieves better reconstruction while using *lower* loudspeaker gains.

5. CONCLUSIONS

This paper has presented a novel multiple-loudspeaker spatial sound reproduction approach denoted as sound field reconstruction (SFR), which is based on the spatio-temporal spectral properties of the plenacoustic function, in particular the plenacoustic sampling, and on numerical multichannel inversion.

SFR was compared with wave field synthesis (WFS), which is a prevalent approach to multiple-loudspeaker sound reproduction for an extended listening area. The comparison included simulating far- and near-field sound source reproduction at frequencies where WFS works correctly and at frequencies where it is known to experience aliasing artifacts.

The simulation experiments have shown that SFR has slightly better sound field reconstruction capabilities than WFS at frequencies where WFS works correctly (before aliasing starts becoming an issue). In addition, at higher frequencies, where sound field reproduced with WFS features strong aliasing distortions, SFR maintains its reproduction accuracy, even if it uses lower loudspeaker gains.

6. REFERENCES

[1] A. Blumlein, "Improvements in and relating to sound transmission, sound recording and sound reproduction systems," *British Patent Specification 394325*, 1931, Reprinted in *Stereophonic Techniques*, Aud. Eng. Soc., New York, 1986.

[2] P. S. Gaskel and P. A. Ratliff, "Quadraphony: developments in matrix H decoding," Tech. Rep., BBC Research Department, February 1977.

[3] M. A. Gerzon, "Periphony: Width-Height Sound Reproduction," *J. Aud. Eng. Soc.*, vol. 21, no. 1, pp. 2–10, January 1973.

[4] M. A. Gerzon, "Practical periphony: The reproduction of full-sphere sound," in *Preprint 65th Conv. Aud. Eng. Soc.*, Feb. 1980.

[5] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[6] A. J. Berkhout, D. de Vries, and P. Vogel, "Wave front synthesis: a new direction in electroacoustics," in *Preprint 93th Conv. Aud. Eng. Soc.*, Oct. 1992.

[7] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "The Plenacoustic Function and its Sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, 2006.

[8] C. Faller, "Signal processing for speech, audio, and acoustics," 2006, Course Notes, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[9] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 2, pp. 145–152, Feb 1988.

[10] Paul Flikkema, "An algebraic theory of 3d sound synthesis with loudspeakers," in *AES 22nd Int'l Conference: Virtual, Synthetic, and Entertainment Audio*, June 2002.

[11] N. Bose and S. Mitra, "Generalized inverse of polynomial matrices," *Automatic Control,*

IEEE Transactions on, vol. 23, no. 3, pp. 491–493, Jun 1978.

[12] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 189–194, Mar 1998.

[13] S. Spors, A. Kuntz, and R. Rabenstein, "Listening room compensation for wave field synthesis," in *ICME '03: Proceedings of the 2003 International Conference on Multimedia and Expo*, Washington, DC, USA, 2003, pp. 725–728, IEEE Computer Society.

[14] M. Guillaume, Y. Grenier, and G. Richard, "Iterative algorithms for multichannel equalization in sound reproduction systems," *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 3, March 2005.

[15] G. Golub and W. Kahan, "Calculating the singular values and pseudo-inverse of a matrix," *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, vol. 2, no. 2, pp. 205–224, 1965.