Audio Engineering Society

Convention Paper

Presented at the 127th Convention
2009 October 9–12 New York NY, USA

Designing Practical Filters For Sound Field Reconstruction

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

Multichannel sound field reproduction techniques, such as Wave Field Synthesis (WFS) and Sound Field Reconstruction (SFR), define loudspeaker filters in the frequency domain. However, in order to use these techniques in practical systems, one needs to convert these frequency-domain characteristics to practical and efficient time-domain digital filters. Additional limitation of SFR comes from the fact that it uses a numerical matrix pseudoinversion procedure, where the obtained filters are sensitive to numerical errors when the system matrix has a high condition number. This paper describes physically-motivated modifications of the SFR approach that allow for mitigating conditioning problems and frequency-domain loudspeaker filter smoothing that allows for designing short time-domain filters while maintaining high sound field reproduction accuracy. It also provides comparisons of sound field reproduction accuracy of WFS and SFR using the obtained discrete-time filters.

1. INTRODUCTION

In the last decades, a number of multichannel sound field reproduction techniques for a wide listening area have been proposed. The most notable are Wave Field Synthesis (e.g., see [1,2]) and Ambisonics (e.g., see [3, 4]).

Recently, the present authors have proposed in [5] an approach termed Sound Field Reconstruction (SFR), that is based on transforming the problem of extended-area sound field reproduction into that of multiple-input multiple-output (MIMO) channel inversion and provides a numerical optimization procedure that minimizes the sound field reproduction error on a grid of points in the desired listening area. A similar MIMO-inversion-based technique was proposed by Corteel in [6] for room equalization for

Wave Field Synthesis systems. However, there are notable differences between the two approaches in the way the control point grid is chosen[1] and in the procedure used for the MIMO channel inversion.[2]

The analysis, and consequently the design of loudspeaker driving signals (including loudspeakers' filter design) of all the mentioned techniques is done in the frequency domain and so considers only steady-state system behavior. Moreover, as will be shown in Section 3, the frequency-domain filter responses in SFR are computed with a non-linear numerical matrix pseudo-inversion procedure, making the obtained filters sensitive to computational noise at low frequencies, where the problem's effective condition number tends to be high. Additionally, depending on the physical setup, filters for some of the loudspeakers can have low energy, making them additionally sensitive to numerical noise, which is reflected through irregular local oscillations around the mean in their frequency characteristic.

This paper deals with the design of practical SFR time-domain filters. It tackles the problems inherent to the SFR approach, but also to WFS and other techniques which define filters by their frequency characteristic. Namely, the SFR approach will be extended in order to allow for using physically-based (relative to the desired source's position) subsets of loudspeakers—a technique described by Corteel in [6]. On the other hand, filters specified in the frequency domain will be processed in order to make them more amenable to time-domain filter computation. In particular, magnitude and phase responses are smoothed so that they can be modelled by rational functions, and in the presented approach—as finite impulse response (FIR) filters. Additionally, the impulse responses of the filters thus obtained are pruned in order to reduce the implementation complexity while maintaining high reproduction accuracy.

A sound field reproduction system should be able to correctly reproduce impulsive, as well as stationary sound events with good accuracy. Therefore, in order to assess the performance of the WFS and SFR systems using the designed filters, this paper will show the time behavior (responses to impulsive sound events) and frequency behavior (responses to single-frequency tones) of a loudspeaker-array reproduction setup where the reproduction filters are obtained with the aforementioned WFS and SFR filter design procedures.

The paper is organized as follows. Section 2 briefly presents WFS and describes a procedure for designing discrete-time loudspeaker filters for WFS systems. WFS using filters obtained with the described procedure are later compared with the corresponding SFR system. Section 3 presents SFR and the design of practical, discrete-time loudspeaker filters for SFR systems. Section 4 presents simulation experiments which compare time- and frequency-domain reproduction accuracy of WFS and SFR. Conclusions are given in Section 5.

2. WAVE FIELD SYNTHESIS FILTER DESIGN

2.1. Wave Field Synthesis basics

Wave field synthesis (WFS) is a spatial sound reproduction technique which is based on Huygens' principle and its mathematical description expressed through Kirchoff-Helmholtz and Rayleigh integrals [1, 2].

The Rayleigh I integral states that[3] 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,

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

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

[1] SFR uses a grid of points spaced at the Nyquist distance corresponding to the maximum reproduced frequency, whereas in [6], control points are spaced at 100 mm.

[2] In [6], a regularized matrix pseudo-inversion described in [7] was used, whereas SFR uses a pseudo-inversion described in [8], which better addresses the problem of ill-conditioning at low frequencies.

[3] Provided the Green's function satisfies the Sommerfeld radiation condition [9].

be interpreted as a sound field of a distribution of secondary point sources driven with the signals

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

which are 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 is usually done for discrete and finite *linear* distribution of secondary sources (loudspeakers), making a simplifying assumption that primary sources, loudspeakers, and listening area are situated in the same plane. Additionally, since a correct reproduction with a line of monopole sources is not possible in a wide listening area, the loudspeakers' driving signals are computed for correct reproduction on a *reference line*. Fig. 1 shows a linear loudspeaker array that is used for reproducing the sound field of a primary monopole source on a reference listening line with WFS.

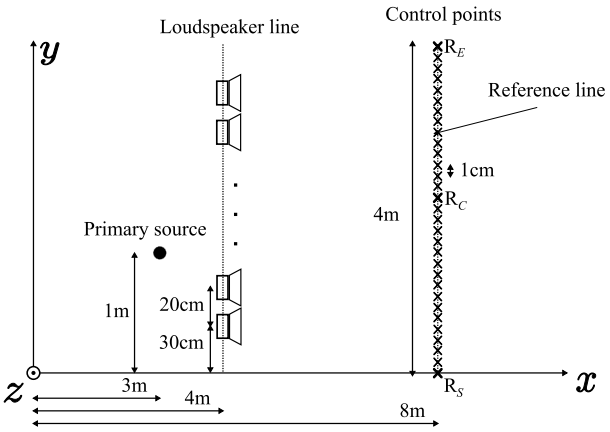


Fig. 1: A sound field reproduction setup using a line loudspeaker array of 18 loudspeakers spaced at $\Delta_l = 20$ cm. The loudspeaker array is located at $x_i = 4$ m, and extends from $y_i^0 = 30$ cm to $y_i^1 = 370$ cm. The primary source to be reproduced is a point source placed at $\mathrm{r}_m = (3\ m, 1\ m)$. The listening area of interest is located in front of the loudspeaker array, starting from $x_l = 8$ m, and extends from $y_l^0 = 0$ m and $y_l^1 = 4$ m. The reference line is located at $x_l = 8$ m and is covered by a grid of points spaced at $\Delta_c = 2$ cm, between $y_l^0 = 0$ m and $y_l^1 = 4$ m.

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 loudspeaker. The loudspeaker driving signals $\tilde{Q}_i(\omega)$ are given by [10]

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

where x_l, x_i and x_m are the x coordinates of the reference listening line, the line containing the loudspeakers, and the primary point source, respectively; $S(\omega)$ is the primary source's spectrum, 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 loudspeaker distribution with a linear one.
- *Truncation effects*, such as the reduction of the area of correct reproduction, resulting from approximating infinitely long loudspeaker distributions with finite apertures.
- *Aliasing artifacts*, resulting from a coarse spacing of loudspeakers. 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 loudspeakers, and $\alpha_{\max}$ the maximum incidence angle between the loudspeaker line and the particle velocity vector of the reproduced sound field.

2.2. Wave Field Synthesis improvements

2.2.1. Tapering on edges of the loudspeaker array to decrease diffraction

As already mentioned, WFS systems suffer from the truncation effects due to the use of finite-length

loudspeaker arrays. The truncation effectively reduces the sound field reproduction area to the inside of the angle subtended by the reproduced source and the edges of the loudspeaker array, inside of which diffraction waves interfere with the reproduced sound field [11].

One way of mitigating the effects of diffraction waves on the reproduced sound field is the use of a spatial window on the loudspeaker driving signals—a technique termed *tapering*. According to [11], the best diffraction energy smoothing is achieved through the use of a raised-cosine taper. An example of a raised-cosine taper with taper width of 40 cm, designed for the loudspeaker array from Fig. 1, is shown in Fig. 2. For a more detailed tapering description, refer to [11].

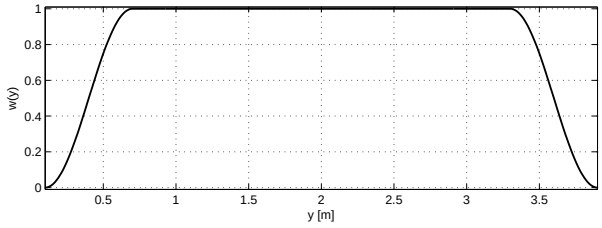


Fig. 2: Symmetric raised-cosine taper with taper width of 40 cm used for smearing the energy of diffraction waves in WFS reproduction area. This taper is later used for designing WFS filters.

2.2.2. Filter correction through power normalization on the reference line

From the WFS loudspeaker driving signals given in (4), it follows that WFS loudspeaker filters $\tilde{H}_i(\omega)$ are given by

$$\tilde{H}_i(\omega) = \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{|x_l - x_i|}{|x_l - x_m|}} \frac{e^{-jk|\mathbf{r}_i - \mathbf{r}_m|}}{\sqrt{|\mathbf{r}_i - \mathbf{r}_m|}} \cos\theta_i \Delta y \,, \tag{6}$$

and their magnitude frequency response is shown in Fig. 3(a).

WFS loudspeaker filters given by (6) are computed in reference to an infinite and continuous setup, where constructive interference of loudspeakers' sound fields takes place at all frequencies. However, as mentioned previously, above the aliasing frequency given by (5) the sound fields of single loudspeakers do not combine constructively, corrupting the resulting sound field with spurious patterns called aliasing artifacts. Additionally, due to the high-pass characteristic of the WFS loudspeaker filters (see Fig. 3(a)), the reproduced sound field's power surpasses that of the desired field at high frequencies.

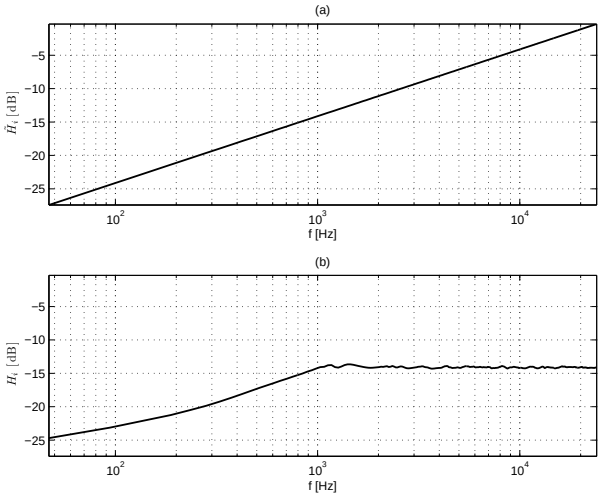


Fig. 3: Magnitude response of WFS loudspeaker filters, as given by (6), before (a) and after (b) normalizing power on control points on the reference line. Note that the response varies by a constant gain dependent on positions of the loudspeaker and the primary source.

A way of avoiding strong frequency response deviations in the reproduced sound field at high frequencies, while at the same time mitigating coloration at low frequencies due to tapering, is by normalizing loudspeaker filters' with a constant gain correction factor at each frequency ω. The normalization is done with the goal of matching the average power of the reproduced and the desired sound field on a grid of control points on the reference line, depicted in Fig. 1. In order to avoid aliasing in the process of power averaging, the control points are spaced densely enough to accommodate the highest reproduced frequencies [12].[4]

The WFS filter correction is done as follows. If $A_1(\omega), \ldots, A_M(\omega)$ and $Y_1(\omega), \ldots, Y_M(\omega)$ are respectively the amplitudes of the desired and the reproduced sound field at frequency ω in M control

[4] As shown in Fig. 1, control points are spaced at $\Delta_y = 1$ cm.

points, then each loudspeaker filter is corrected by

$$H_i(\omega) = c_f(\omega)\tilde{H}_i(\omega)\,, \tag{7}$$

where $c_f(\omega)$ is a correction factor given by

$$c_f(\omega) = \frac{\sqrt{\sum_{i=1}^{M} A_i^2(\omega)}}{\sqrt{\sum_{i=1}^{M} Y_i^2(\omega)}}\,. \tag{8}$$

The magnitude frequency response of WFS filters after the described correction are shown in Fig. 3(b).

2.3. Designing discrete-time filters for Wave Field Synthesis

Up to this point, the WFS loudspeaker driving signals—and consequently loudspeaker filters—have been treated solely in the frequency domain. This subsection deals with the problem of turning the frequency-domain specification of WFS filters, obtained by starting from the theoretical model in (6) and applying practical improvements such as tapering and gain correction, into a set of discrete-time filters.

In order to obtain a set of short finite impulse response (FIR) filters amenable to efficient implementation, one needs an appropriate filter model and frequency- and time-domain processing starting from the frequency-domain response $H_i(\omega)$. The used filter model, a procedure for delay extraction and compensation, and impulse response pruning used for obtaining these filters are described in the following.

The filter design procedure described here is done for the sound reproduction setup shown in Fig. 1, which corresponds to a real setup the authors are setting up for listening tests. The system's sampling frequency is $f_s = 48$ kHz. The filters' processed frequency responses are transformed to the discrete-time domain using discrete Fourier transform (DFT). The size of the used DFT, which defines the frequency grid on which the frequency response is computed, is $N_T = 1024$ samples. The used DFT size, as will be described in more detail later, was large enough to accommodate physically possible delays across the loudspeaker array without a risk of having circular shifts of the impulse responses, and at the same time bring the time-domain aliasing errors to a reasonably low level.

2.3.1. Discrete-time filter model

From the theoretical frequency response of WFS filters in (6) and its subsequent scaling through tapering and power normalization, one can see that filters $H_i(\omega)$ can be decomposed into a pure delay, defined by the complex exponential $e^{-jk|\mathbf{r}_i-\mathbf{r}_m|}$, and a filter $H_i^d(\omega)$ containing the remaining expressions, given by

$$H_i^d(\omega) = c_f(\omega)\sqrt{\frac{jk}{2\pi}}\sqrt{\frac{|x_l - x_i|}{|x_l - x_m|}}\frac{\cos\theta_i \Delta y}{\sqrt{|\mathbf{r}_i - \mathbf{r}_m|}}\,. \tag{9}$$

The impulse response of the filter $H_i^d(\omega)$ is non-causal. In order to make it causal, its impulse response is delayed and pruned to a causal FIR filter (as is shown later in this section).

In the time domain, WFS filters can thus modeled as

$$h_i[n] = \delta_{d_i}[n] * h_i^d[n]\,, \tag{10}$$

where the propagation delay d_i, expressed in samples, is equal to

$$d_i = \frac{|\mathbf{r}_i - \mathbf{r}_m|}{c} f_s\,, \tag{11}$$

and $h_i^d[n]$ is a delayed impulse response of the filter $H_i^d(\omega)$. Note that, due to filter realization convenience,[5] one could round the delay d_i to the closest smaller integer multiple of the sampling period, while moving the residual delay to the filter $H_i^d(\omega)$.

2.3.2. Delay extraction and compensation

The filters $H_i^d(\omega)$ are zero-phase and non-causal. However, as will be shown later, their amplitude decays at a high rate, making it possible to approximate them with a short FIR filter.

Obtaining the impulse response of filters $H_i^d(\omega)$ which are defined on a regular frequency grid is done by performing an inverse DFT. However, in order to avoid the wrap-around on the left DFT boundary of the impulse response due to non-causality, filter $H_i^d(\omega)$ is multiplied by a delay factor corresponding to a delay of $N_T/2$ samples

$$H_i^{dc}(\omega) = H_i^d(\omega)e^{-j\frac{\omega N_T}{2f_s}}\,. \tag{12}$$

[5] It is more convenient to have a delay integer number of samples long, as a fractional delay would require designing a fractional delay filter (e.g., see [13]) or rounding the delay to the closest integer.

Fig. 4 shows an example central part of an obtained impulse response of a filter $H_i^{dc}(\omega)$.

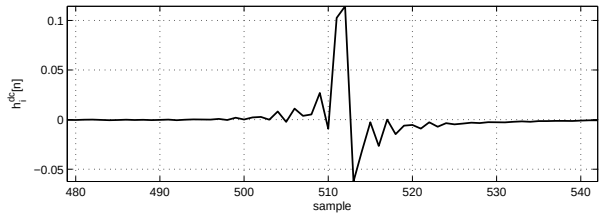


Fig. 4: Impulse response $h_i^{dc}[n]$ of the filter $H_i^d(\omega)$ (delayed version of $h_i^d[n]$) zoomed in around the half of the DFT length N_T.

2.3.3. Impulse response pruning

From Fig. 4, it can be seen that the impulse responses of filters $h_i^{dc}[n]$ decay from the main peak at a high rate. In order to illustrate this energy decay more clearly, Fig. 5 shows the relative amount of filter's energy outside of windows of increasing size centered at the half of the DFT length N_T, i.e., the position of the filter's main peak.[6]

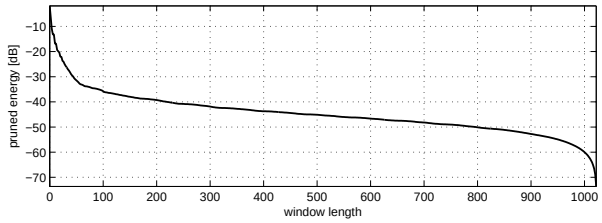


Fig. 5: Relative energy of the error of pruning the impulse responses $h_i^{dc}[n]$ with rectangular windows of different sizes centered at the filters' main peak (half of the DFT size N_T).

From Fig. 5, it is apparent that filters $h_i^{dc}[n]$ can be pruned to the size of 256 samples, as the pruning error at that filter length is below 40 dB. However, to allow for better reproduction accuracy and use the same filter length as SFR (Section 3) for comparison purposes, the filter length was chosen to be $N_F = 512$ samples.

Consequently, the impulse response of the filters $h_i^{dc}[n]$ are tapered by a zero-padded symmetric window with half-cosine edges given by

$$w[n] = \begin{cases} 0 & 0 \le n < 256 \\ \frac{1}{2} - \frac{1}{2}\cos\left(\frac{(n-255)\pi}{257}\right) & 256 \le n < 320 \\ 1 & 321 \le n < 704 \\ \frac{1}{2} + \frac{1}{2}\cos\left(\frac{(n-703)\pi}{257}\right) & 704 \le n < 768 \\ 0 & 768 \le n < 1024 \end{cases}, \tag{13}$$

which is shown in Fig. 6.

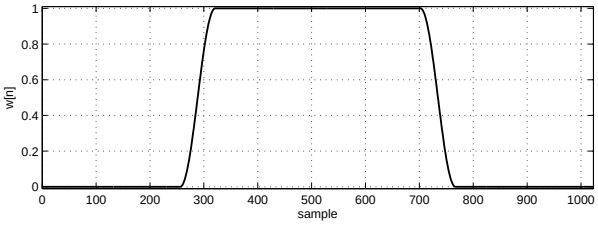


Fig. 6: Window used for pruning the loudspeaker filters to the length of 512 samples.

Finally, after tapering $h_i^{dc}[n]$, their $N_F = 512$ central non-zero samples define the filters $h_i^d[n]$. An example filter is shown in Fig. 7.

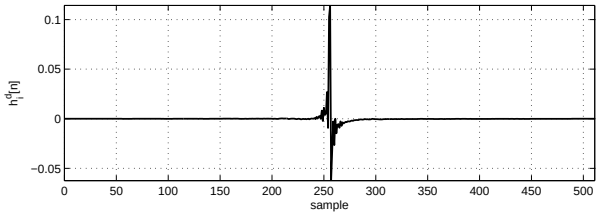


Fig. 7: Impulse response of filters $h_i^d[n]$.

3. SOUND FIELD RECONSTRUCTION FILTER DESIGN

3.1. Sound Field Reconstruction basics

Sound field reconstruction (SFR), described in [5], is a spatial sound field reproduction approach that is based on the spatio-temporal properties of the sound field and techniques for multiple-input multiple-output (MIMO) inverse channel filtering. In particular, the authors have shown in [5] that the problem of controlling the sound field that emanates from band-limited sources with maximum temporal frequency ω_m can be effectively stated as controlling

[6] The filter's energy outside of a window gives an indication of the error resulting from pruning the filter's impulse response to the support of that window.

the sound field on a grid of points spaced at or above the Nyquist spatial frequency $\phi_s = 2\omega_m/c$.

In the multiple-loudspeaker sound field reproduction, the goal is to reproduce a desired sound field of a certain maximum temporal frequency ω_m in a listening area of finite size. The above observation states that a correct sound field reproduction is possible if one is able to match the desired sound field on a grid of points inside the listening area which are spaced at the Nyquist spatial frequency ϕ_s. In other words, the problem of sound reproduction with multiple loudspeakers in a continuous listening area can be posed as a multiple-input multiple-output (MIMO) control problem.

3.1.1. Sound Field Reconstruction using MIMO channel inversion

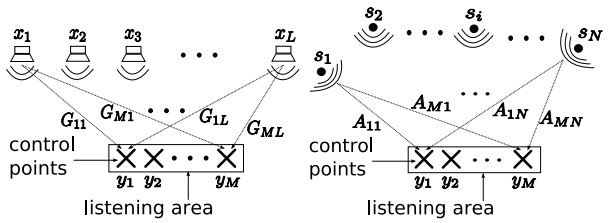
The problem of MIMO channel inversion in the context of sound field reproduction with an array of loudspeakers is illustrated in Fig. 8. The reproduction setup includes an array of L loudspeakers and a grid of M control points covering the listening area In addition, there is a desired acoustic scene that contains N sound sources that would evoke the desired sound field in the listening area.

Positions of loudspeakers, control points, and desired sources are known. The impulse responses $A_{ij}(\omega)$ and $G_{ik}(\omega)$ of the sound propagation channels between the jth desired source and the ith control point, and the kth loudspeaker and the ith control point, respectively, are also known through computation or measurements.

The goal of the MIMO channel inversion in the context of SFR is the reproduction of the desired sound scene in M control points, i.e., computation of the loudspeaker driving signals that evoke the same signals at the control points as the original sound scene.

Note that the problem of multichannel inversion can be represented as a superposition of N independent sub-problems, each involving a single desired source, as illustrated in Figure 1. The loudspeaker signals are then obtained by summing the contributions for each single-source sub-problem. In the following, the MIMO channel inversion analysis is therefore presented only for the first desired source.

Denote by $S_1(\omega)$, $X_j(\omega)$, and $Y_k(\omega)$ the signals of the first desired source, the output of the jth loudspeaker, and the sound pressure at the kth control point, respectively. Furthermore, denote by $D_l(\omega)$ the signal at the lth control point in the desired sound scene containing only the first desired source.



(a) Reproduction setup (b) Desired sound scene

Fig. 8: Multichannel inversion problem overview.

The signals $D_i(\omega)$ are determined by the effects of the sound propagation paths from the desired source to the control points, and are described by the following product in the frequency domain:

$$\mathbf{D}(\omega) = \mathbf{A}(\omega)S_1(\omega)\,, \tag{14}$$

where

$$\mathbf{D}(\omega) = [D_i(\omega)]_{M\times 1}$$
$$\mathbf{A}(\omega) = [A_{i1}(\omega)]_{M\times 1}\ .$$

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

$$\mathbf{Y}(\omega) = \mathbf{G}(\omega)\mathbf{X}(\omega)\,, \tag{15}$$

where

$$\mathbf{Y}(\omega) = [Y_i(\omega)]_{M\times 1}$$
$$\mathbf{G}(\omega) = [G_{ij}(\omega)]_{M\times L}$$
$$\mathbf{X}(\omega) = [X_i(\omega)]_{L\times 1}\ .$$

The task of the multichannel inversion is computing the signals $X_j(\omega)$ using the desired signal $S_1(\omega)$, i.e.,

$$\mathbf{X}(\omega) = \mathbf{H}_1(\omega)S_1(\omega)\,, \tag{16}$$

where

$$\mathbf{H}_1(\omega) = [H_{i1}(\omega)]_{L\times 1}\ ,$$

such that the difference (error) between the vector $\mathbf{Y}(\omega)$ and vector $\mathbf{D}(\omega)$, corrected by a constant delay Δ accounting for the propagation time differences or the modeling delay, is minimized. The multichannel inversion problem is illustrated in Fig. 9.

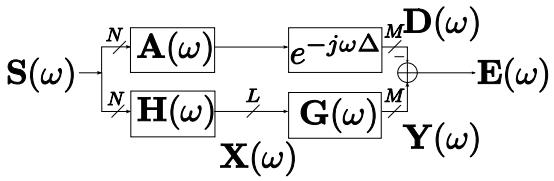


Fig. 9: Block diagram illustrating the MIMO channel inversion problem.

The solution which minimizes the error power, i.e. the mean squared error (MSE) solution, given by

$$\boldsymbol{H}_1(\omega) = e^{-j\omega\Delta}\boldsymbol{G}^{+}(\omega)\mathbf{A}_1(\omega)\,, \qquad (17)$$

uses a pseudo-inverse $\boldsymbol{G}^{+}(\omega)$ of the transfer matrix $\boldsymbol{G}(\omega)$. Thus, finding the pseudo-inverse of the matrix $\boldsymbol{G}(\omega)$ becomes central for the problem of MIMO channel inversion.

The classical full-rank pseudo-inverse expression given by

$$\mathbf{G}^{+}(\omega) = \left(\mathbf{G}^{H}(\omega)\mathbf{G}(\omega)\right)^{-1}\mathbf{G}^{H}(\omega)\,, \qquad (18)$$

where the matrix $\boldsymbol{G}^{H}(\omega)$ is the conjugate-transpose of the matrix $\boldsymbol{G}(\omega)$. However, it is not practical for the problem of loudspeaker filter design, as at low frequencies, where the condition number of the matrix $\boldsymbol{G}(\omega)$ is large (making it effectively low-rank), it gives filters with excessively large gains, beyond the practical limitations of common loudspeakers. The regularized pseudo-inversion used in [6, 7] is also of limited use, as it does not allow easy control of the trade-off between the reproduction accuracy and filters' gains.

Instead, SFR uses a pseudo-inversion based on the singular value decomposition (SVD) and pruning of the singular values that are below a defined threshold ϵ (e.g., see [8]). At high frequencies, where all singular values of the matrix $\boldsymbol{G}(\omega)$ are larger than the threshold, this procedure gives the result identical to (18). However, at low frequencies, it gives near-optimal solutions while keeping the loudspeaker filter gains within practical limitations. For a more detailed treatment of this MIMO channel inversion problem, see [5].

3.2. Sound Field Reconstruction improvements

3.2.1. Filter correction through power normalization on the reference line

Both WFS and SFR filter computation procedures give solutions above certain aliasing frequency that do not provide correct sound field reproduction accuracy.

Although coming from the same physical limitations as in the case of WFS, the problems of SFR at high frequencies can be explained from another perspective. Namely, at high frequencies, where the constructive interference of sound fields of different sources can not be achieved, the least mean squared error solution is biased towards highly attenuating all signals, such that the reconstruction error approaches the desired signal.[7]

SFR filter correction is done through normalization of the average power of the desired and the reproduced sound field on a grid of control points covering the reference line (see Fig. 1), which was explained in the case of WFS filters. Refer to Section 2.2.2 for details.

3.2.2. Loudspeaker subset selection

While it might seem beneficial to use all loudspeakers for reproduction with SFR, there are many cases where using only a subset of loudspeakers can give better reproduction provided the optimization is done for a finite listening area at a known location. This observation was also used by Corteel in [6], where he showed how based on the location of the primary (reproduced) source and the listening area, one can select a sub-array of loudspeakers which physically contribute the most to the sound field reproduction.

There is a plausible explanation for such a selection. Considering a case where an impulsive sound arrives from the location of a primary source, one expects that at all locations in the listening area the received sound is of similar duration and consequently without significant coloration. However, using all loudspeakers makes a combination of the impulse responses—due to different delays—more spread in

[7] This is a known phenomenon in Wiener filtering, where at low SNRs, the Wiener filter's response tends toward zero.

time and varying across different positions than in the case when only a subset of loudspeakers is used. Thus, the use of all loudspeakers causes not only temporal, but also spatial coloration.

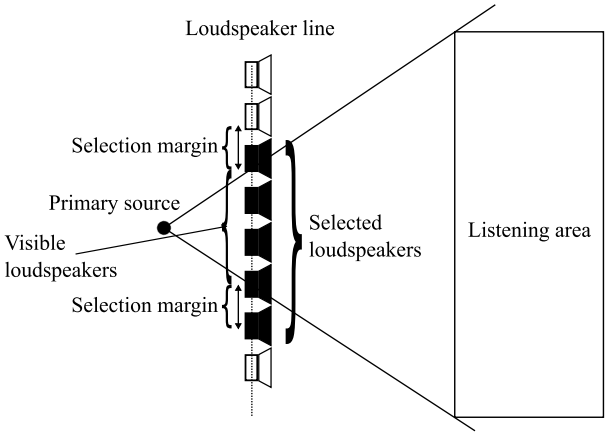


Fig. 10: Illustration of loudspeaker selection based on the primary source position. The visible loudspeakers are those which belong to the angle subtended by the listening area to the primary source. Loudspeakers within a selection margin of the visible loudspeakers are also selected.

The selection procedure, proposed by Corteel, considers only those loudspeakers that are visible—extended by a predefined selection margin—from the source when looking towards the listening area, as shown in Fig. 10. The rationale behind such a choice is twofold: firstly, it uses the loudspeakers whose contribution is largest when all loudspeakers are used, preserving most of the reconstruction accuracy and secondly, the used loudspeakers have lowest delay spreads due to propagation distance differences and their position relative to the sound wavefront.

Fig. 11 illustrates how SFR does not noticeably lose the reproduction accuracy when only a subset of six loudspeakers is used for reproducing a sinusoid at frequency $f = 500$ Hz. The six loudspeakers were selected as those within the angle subtended by the listening area to the primary source, extended on both sides by a selection margin of one inter-loudspeaker distance $\Delta_l = 20$ cm.

On the other hand, Fig. 12 shows that the same loudspeaker selection is not as effective with WFS,

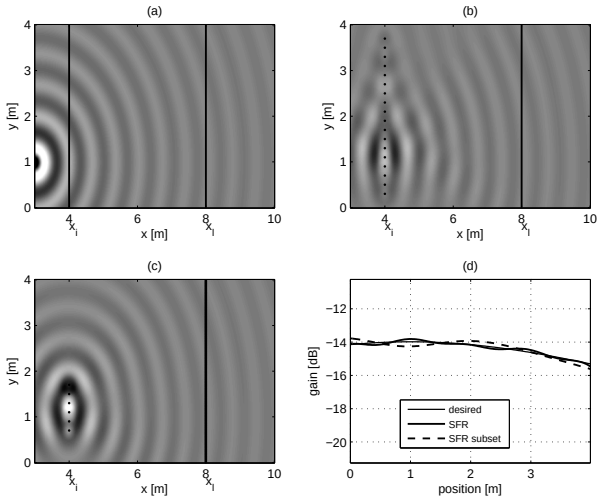


Fig. 11: Comparison of SFR with the entire loudspeaker array from Fig. 1 and SFR with only a selected 6-loudspeaker sub-array at reproducing a point source with frequency $f_1 = 500$ Hz located at $\boldsymbol{r}_m = (3\text{ m}, 1\text{ m})$: snapshot of the desired sound field (a), snapshot of sound fields reproduced with SFR using the entire array (b) and the sub-array (c), and amplitude of the three sound fields on the reference line (d).

since the reconstruction accuracy on the reference line aggravates outside of its center.

3.3. Designing discrete-time filters for Sound Field Reconstruction

As was the case in designing discrete-time WFS filters, the goal of the discrete-time SFR filter design is obtaining short FIR filters amenable to efficient implementation. In this case though, the appropriate modeling and frequency- and time-domain processing of these filters are even more critical, as they are obtained by a numerical procedure and have no theoretical model to start with.

In the following, discrete-time filter modeling, frequency-domain smoothing, delay extraction and compensation, and impulse response pruning are described, specifying a complete SFR filter design procedure.

The filter design is done in reference with the same sound reproduction setup shown in Fig. 1. The sampling frequency $f_s = 48000$ kHz and DFT length $N_T = 1024$ are the same as with the WFS filter design presented in Section 2.

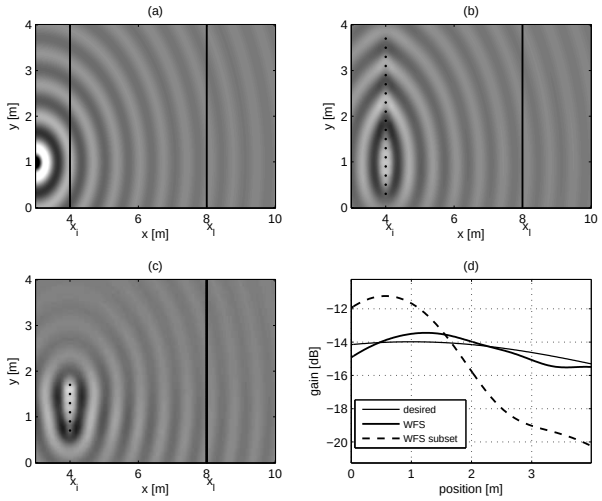


Fig. 12: Comparison of WFS with the entire loudspeaker array from Fig. 1 and WFS with only a selected 6-loudspeaker sub-array at reproducing a point source with frequency $f_1 = 500$ Hz located at $\boldsymbol{r}_m = (3\text{ m},\ 1\text{ m})$: snapshot of the desired sound field (a), snapshot of sound fields reproduced with WFS using the entire array (b) and the sub-array (c), and amplitude of the three sound fields on the reference line (d).

3.3.1. Discrete-time filter model

Similarly to the WFS discrete-time filter model, SFR discrete-time filters are modeled by a convolution of several delay filters and a short FIR filter. The filter model is given by

$$h_k[n] = \delta_{d_C}[n] * \delta_{d_k}[n] * \hat{h}_k[n]\,, \tag{19}$$

where d_C is a common delay determined by the distance between the primary source and the loudspeaker array line, $\delta_{d_k}[n]$ is the delay filter associated with the residual propagation delay d_k of the kth loudspeaker, and $\hat{h}_k[n]$ is a short FIR filter.

It should be clarified here why this filter model was chosen. The common delay filter $\delta_{d_C}[n]$, that is associated with propagation delay d_C from the primary source to the loudspeaker array is used in order to avoid delay estimation errors from the phase responses of SFR filters.[8] Delay d_C can be trivially determined from the positions of the primary source and the loudspeaker array by

$$d_C = r_x/c\,, \tag{20}$$

where r_x is the shortest distance between the two. It should be noted that in order to allow for the common delay extraction in the original frequency-domain filter design procedure, described in Section 3, the delay parameter Δ has been set to $-d_C$. Additionally, delay filters $\delta_{d_k}[n]$ of each loudspeaker are introduced in order to maximally shorten the impulse responses $\hat{h}_k[n]$. In particular, filters $\hat{h}_k[n]$ are aligned in time using delays d_k, which enables capturing maximum energy of the impulse response in a given number of samples.

3.3.2. Frequency response smoothing

The main problem in the frequency-domain design procedure described in Section 3.1 comes from the non-linear operation of pruning small singular values of the system matrix $G(\omega)$. The result of this non-linear operation is illustrated in Fig. 13, where one can clearly see sharp discontinuities in the loudspeaker filters' frequency responses exactly at frequencies where different singular values of the matrix $G(\omega)$ cross the predefined threshold value.

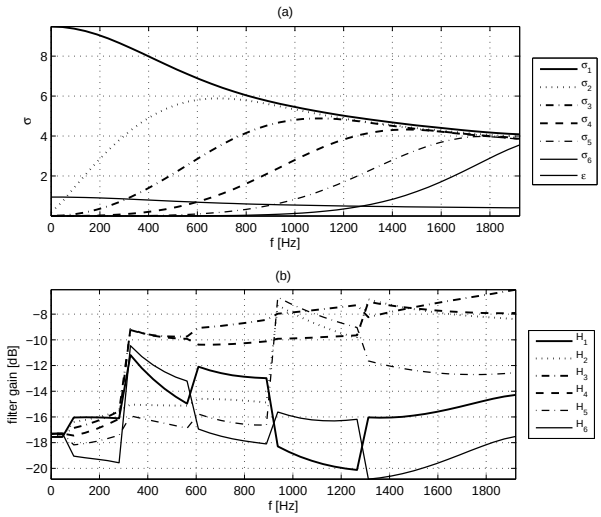


Fig. 13: Singular values of the loudspeaker propagation matrix $\boldsymbol{G}(\omega)$ at different frequencies ω (a) and magnitude response of SFR filters $H_k(\omega)$ obtained from the SFR frequency-domain filter calculation procedure (b).

[8] In cases when the primary source is far from the loudspeaker array and the grid on which the frequency response is discretized is not fine enough, the phase wrap-around by multiples of 2π is possible, giving a wrong estimated delay.

Discontinuities in the frequency response cause time-domain aliasing artifacts when the frequency response is sampled and transformed to the time domain, and require the use of long DFTs in order to keep the aliasing artifacts low. In order to mitigate the time-domain aliasing artifacts while allowing using shorter DFT transforms, a moving-average smoothing filter in the frequency domain is used prior to transforming the frequency response to the time domain.

The argument for such a step is the following. Away from the threshold-crossing point, where the discontinuities occur, the frequency characteristic is relatively smooth on both sides. If the threshold were set at a different level, the discontinuity would be moved to the lower or higher frequency, and the frequency response would stay smooth on both of its sides. Therefore, the moving-average filter effectively provides a smooth transition between two smooth characteristics, which does not excessively affect the resulting response, as both frequency characteristics—with and without the singular value at stake—provide high sound field reproduction accuracy.

Fig. 14 shows the smoothed frequency responses $\tilde{H}_k(\omega)$ of the SFR filters $H_k(\omega)$ obtained from the SFR frequency-domain filter calculation procedure. The smoothing in the frequency domain was done with a 11-tap zero-phase moving-average filter.

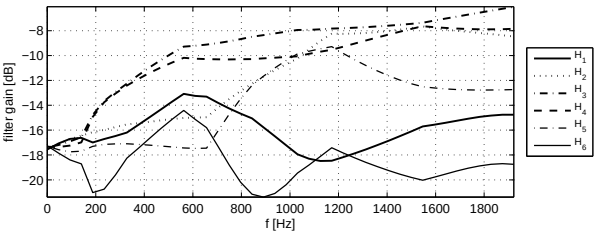


Fig. 14: Magnitude response of SFR filters $\tilde{H}_k(\omega)$ obtained by smoothing the frequency responses $H_k(\omega)$ with a 11-tap zero-phase moving-average filter.

3.3.3. Delay extraction and compensation

Each loudspeaker filter's frequency response, as obtained from the SFR filter computation procedure described in Section 3.1, has an intrinsic delay which depends on the relative positions of the primary source and that loudspeaker. The delay is apparent in the unwrapped phase characteristics, shown in Fig. 15, as the average slope which can be subtracted.

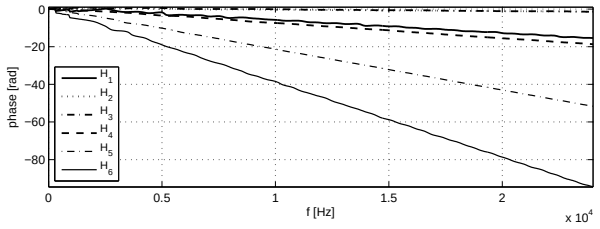


Fig. 15: Phase responses of the SFR filters $H_k(\omega)$ obtained from the SFR frequency-domain filter calculation procedure.

The delay d_k of each loudspeaker is estimated from its unwrapped phase response $\phi_k^u(f)$ using linear regression. In particular, if the unwrapped phase at evaluation frequencies $f_1, \ldots, f_{N_T}$ are stacked in a vector $\boldsymbol{\phi}_k^u = [\phi_k^u(f_1) \ \ldots \ \phi_k^u(f_{N_T})]^T$, then the delay d_k of the kth loudspeaker filter is given by

$$d_k = (\boldsymbol{f}^T \boldsymbol{f})^{-1} \boldsymbol{f}^T \boldsymbol{\phi}_k^u, \tag{21}$$

where the vector $\boldsymbol{f} = [f_1 \ \ldots \ f_{N_T}]^T$ contains the evaluation frequencies.

Note that subtracting the delay d_k from the filter of the kth loudspeaker would align the filter's main peak at delay zero, effectively making the filter zero-phase, and consequently non-causal. To avoid a circular shift when transforming the loudspeaker filters to the discrete-time domain using an inverse DFT, the delays are compensated by only $d_k - N_T f_s/2$ to make the filters' main peaks aligned in the middle of the DFT block, in the same way as is described for the WFS case in Section 2.3.2.

3.4. Impulse response pruning

After the smoothing and delay compensation in the frequency domain, an inverse DFT is applied to the loudspeaker filters' frequency responses in order to obtain their impulse responses. As mentioned previously, each filter's underlying delay has been subtracted in order that the main peak and maximum energy of each filter is concentrated at that delay. Fig. 16 shows the filters of two loudspeakers obtained in this way.

Looking at the impulse responses of the loudspeaker

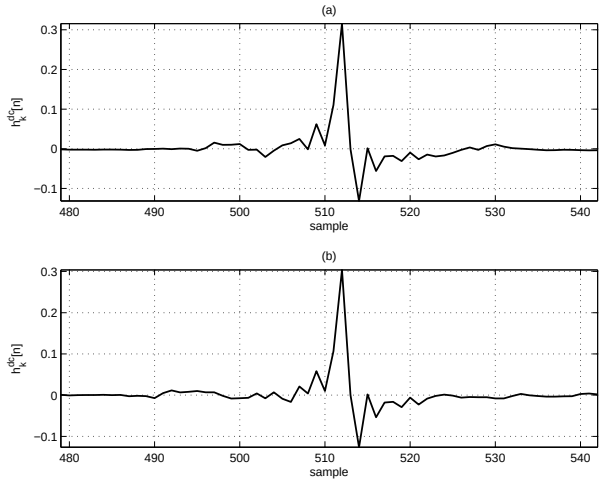


Fig. 16: Impulse responses $h_k^{dc}[n]$ of the delay-compensated filters $\tilde{H}_k(\omega)$ of the third ($k = 3$) loudspeaker (a) and the fourth ($k = 4$) loudspeaker (b) zoomed in around the half of the DFT length N_T.

filters,[9] it can be seen that they decay in time very quickly. In order to illustrate this energy decay more clearly, Fig. 17 shows the relative amount of filter's energy outside of windows of increasing size centered at the half of the DFT length N_T, i.e., the position of the filter's main peak, similarly to Fig. 5.

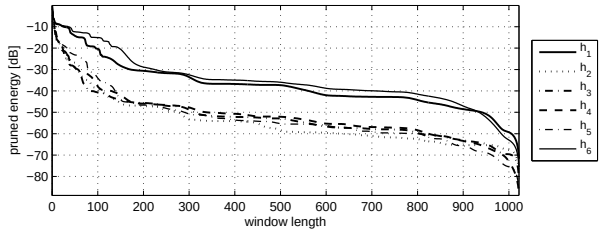


Fig. 17: Relative energy of the error of pruning the SFR impulse responses with rectangular windows of different sizes centered at the filters' main peak (half of the DFT size N_T).

From Fig. 17, it is apparent that the SFR filters can be pruned to the size of 512 samples in the interest of reducing complexity, as the pruning error at that filter length is below 40 dB.

[9] Fig. 16 shows impulse responses of only two loudspeaker filters for synoptical reasons. The impulse responses of other loudspeakers have a similar time-domain structure.

In the same way as was presented in Section 2.3.3, the impulse responses of loudspeaker filters were pruned to $N_F = 512$ samples using a zero-padded symmetric rectangular window with smooth half-raised-cosine edges given by (13) and shown in Fig. 6. For details, refer to Section 2.3.3.

4. SIMULATIONS AND EVALUATION

In order to compare the performance of the two sound field reproduction techniques—WFS and SFR—used with the filters designed by the procedures described in Section 2 and 3, respectively, a number of simulations were carried out using the sound field reproduction setup shown in Fig. 1. All the simulations were done using loudspeaker filters designed for reproducing a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$ (see Fig. 1).

The first three simulations were done with a goal of giving a visual insight into the frequency (using sinusoidal signals) and temporal (using trains of pulses) accuracy of the sound field reproduction with WFS and SFR.

The other simulations presented in this section are used for quantitatively assessing the reproduction artifacts in the time domain (through group delay errors) and frequency-domain (through coloration) in points on the reference line.

4.1. Sound field snapshot

4.1.1. Sinusoidal sources

The first simulation gives insight into the extended-area sound field reproduction accuracy of WFS and SFR for sinusoidal sources. It compares sound fields reproduced with WFS and SFR with the desired sound field for frequencies $f_1 = 500$ Hz and $f_2 = 1500$ Hz, as shown in Fig. 18 and Fig. 19, respectively.

From Fig. 18, it can be seen that the sound field of a sinusoidal point source at frequency $f_1 = 500$ Hz is reproduced with similar accuracy with WFS and SFR. However, in Fig. 19 it is apparent that WFS suffers from aliasing artifacts already at 1500 Hz, while SFR preserves the reproduction accuracy across the entire extended listening area.

4.1.2. Train of low-pass-filtered pulses

In order to assess the extended-area sound field reproduction accuracy of WFS and SFR in a range

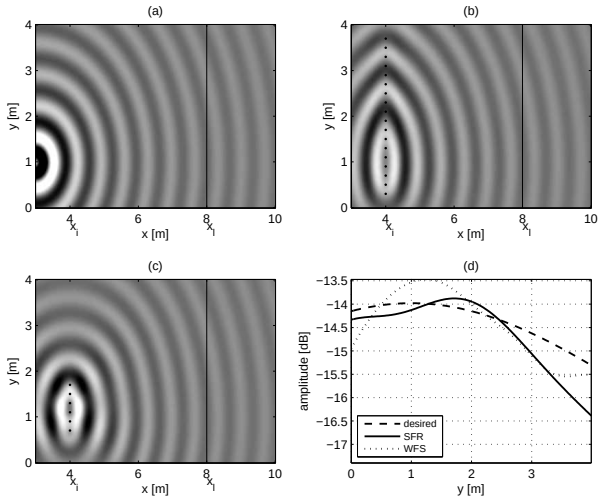


Fig. 18: Comparison of WFS and SFR in reproducing a point source with frequency $f_1 = 500$ Hz located at $\boldsymbol{r}_m = (3\text{ m},\ 1\text{ m})$: snapshot of the desired sound field (a), snapshot of sound fields reproduced with WFS (b) and SFR (c), and amplitude of the three sound fields on the reference line (d).

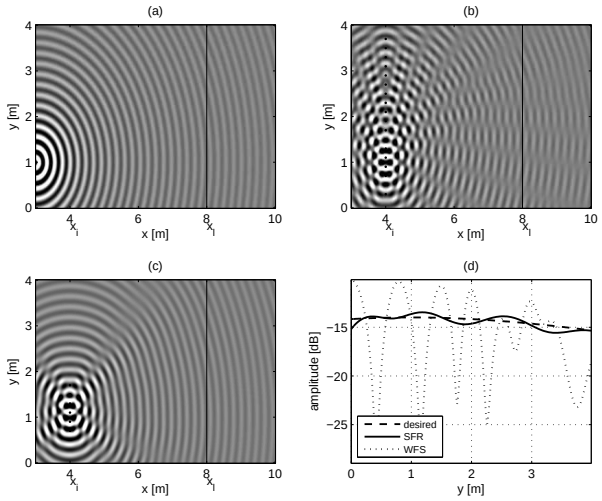


Fig. 19: Comparison of WFS and SFR in reproducing a point source with frequency $f_1 = 1500$ Hz located at $\boldsymbol{r}_m = (3\text{ m},\ 1\text{ m})$: snapshot of the desired sound field (a), snapshot of sound fields reproduced with WFS (b) and SFR (c), and amplitude of the three sound fields on the reference line (d).

of frequencies, the two approaches were compared at reproducing a sound field of a point source located at $\boldsymbol{r}_m = (3\text{ m},\ 1\text{ m})$, that emits a train of low-pass-filtered pulses $p(t)$ with a cut-off frequency $f_c = 3$ kHz, shown in Fig. 20. The pulse train period is $T_p = 4$ ms. The simulation results are shown in Fig. 21.

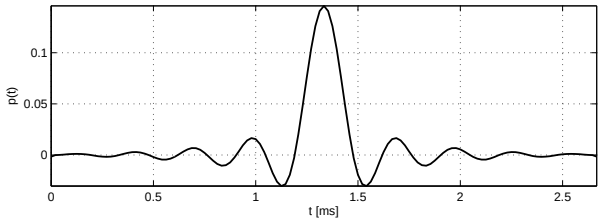


Fig. 20: A low-pass pulse with a cut-off frequency $f_c = 3$ kHz used for constructing a pulse train.

From Fig. 21, one can see that the sound field reproduced using WFS contains notable "tail wavefronts" between "the main wavefronts" conveying the emitted pulses. On the other hand, the sound field reproduced with SFR is closer to the desired one, which is apparent both visually and from the amplitudes on the reference line. The aforementioned tail wavefronts result from a non-constructive wavefront interference, they arrive later and cause impulse response time spreading and consequently coloration, as will be seen later.

4.2. Reference line impulse response analysis

4.2.1. Impulse response

Aside from the frequency-domain reproduction accuracy, it is also important that a sound field reproduction system provides accurate temporal reproduction. Namely, impulsive sounds from a desired source should be reproduced with a minimum modification of the temporal structure, thereby preserving the duration and limiting coloration.

In order to assess the temporal properties of sound field reproduction with WFS and SFR, the systems' impulse responses in three control points, $R_S(8\ m,\ 0)$, $R_C(8\ m,\ 2\ m)$, and $R_E(8\ m,\ 4\ m)$ (see Fig. 1) are shown in Fig. 22, 23, and 24.

In Figs. 22, 23, and 24 it can be seen that on the reference listening line, the impulse responses of both WFS and SFR are composed of a number of short

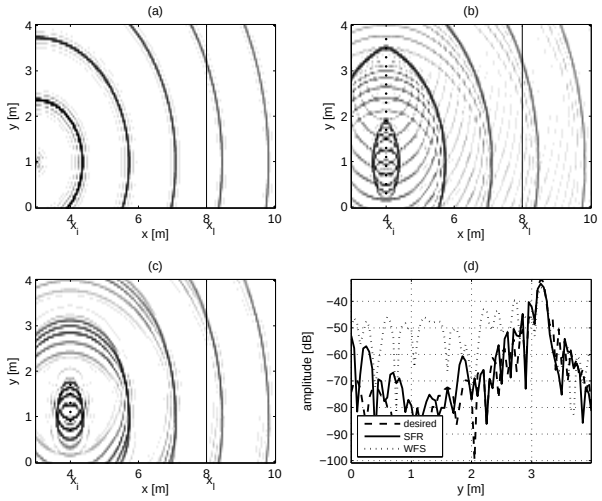


Fig. 21: Comparison of WFS and SFR in reproducing a point source located at $\boldsymbol{r}_m = (3\text{ m},\ 1\text{ m})$ emitting a train of low-pass pulses with a period of $T_p = 4$ ms: snapshot of the desired sound field (a), snapshot of sound fields reproduced with WFS (b) and SFR (c), and amplitude of the three sound fields on the reference line (d).

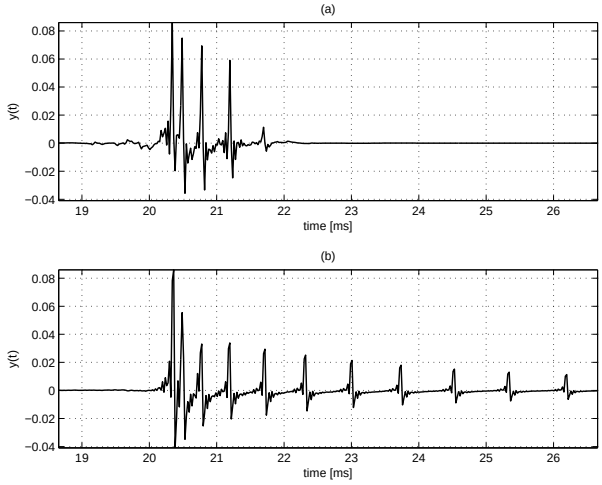


Fig. 22: Impulse responses of SFR (a) and WFS (b) in the control point $R_S(8\ m,\ 0)$ relative to a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$.

pulses differently distributed in time. The impulse responses of SFR are notably shorter than those of WFS, which is an expected result of the loudspeaker selection step described earlier in Section 3.2.2. As a consequence, SFR causes less coloration (also shown later) and reproduces impulsive sounds more accurately. It should also be noted that the impulse responses of SFR would be similar even if all loudspeakers were used, since the "side loudspeakers" (those not selected with loudspeaker subset selection) would have notably lower magnitude.

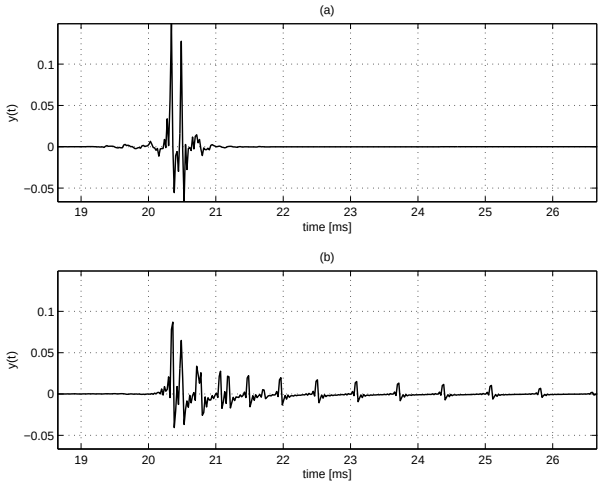


Fig. 23: Impulse responses of SFR (a) and WFS (b) in the control point $R_C(8\ m,\ 2\ m)$ relative to a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$.

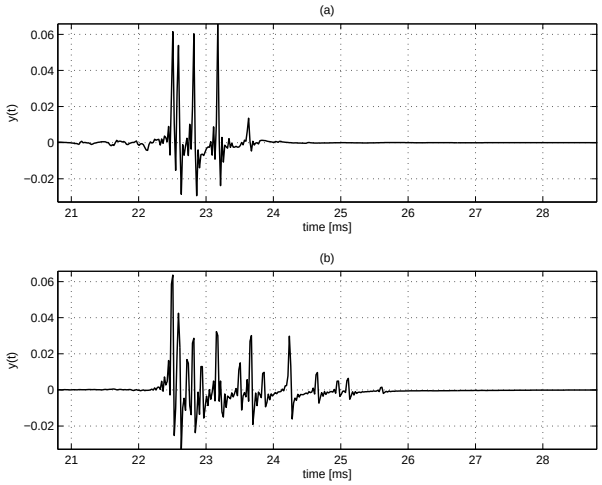


Fig. 24: Impulse responses of SFR (a) and WFS (b) in the control point $R_E(8\ m,\ 4\ m)$ relative to a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$.

4.2.2. Group delay

Fig. 25 shows group delay errors of the impulse re-

sponses in the control points R_S, R_C, and R_E. The group delay error is given by

$$e_\tau(f) = \tau_g(f) - \tau_g^d(f)\,, \tag{22}$$

where $\tau_g(f)$ is the group delay of the reproduction impulse response and $\tau_g^d(f)$ is the group delay of the desired impulse response.

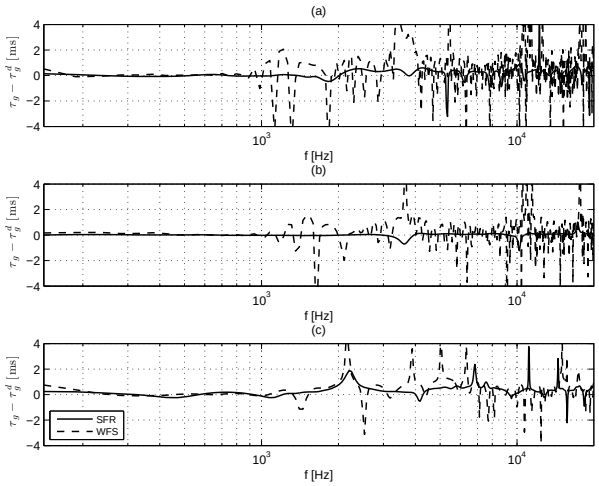


Fig. 25: Group delay errors $e_\tau(f) = \tau_g(f) - \tau_g^d(f)$ of SFR and WFS impulse responses in the control points $R_S(8\ m,\ 0)$ (a), $R_C(8\ m,\ 2\ m)$ (b), and $R_E(8\ m,\ 4\ m)$ (c) relative to a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$.

Considering the findings of Flanagan *et al.* [14][10] that the group delay discrimination threshold for click-like signals presented over loudspeakers is around 2 ms, the plots in Fig. 25 suggest that SFR responses have noticeable group delay errors only at a small number of high frequencies (above 5 kHz) in the points R_S and R_E. On the other hand, WFS responses have perceivable group delay errors in all three control pints on a larger set of frequencies, starting with frequencies below 2 kHz.

4.2.3. Frequency response

Fig. 26 shows magnitude frequency responses of SFR and WFS normalized by the corresponding desired magnitude responses in the points R_S, R_C, and R_E (see Fig. 1):

$$Y_n(f) = \frac{Y(f)}{Y_d(f)}\,, \tag{23}$$

[10] See also [15].

where $Y(f)$ is the magnitude response in a point and $Y_d(f)$ is the desired magnitude response in that point.

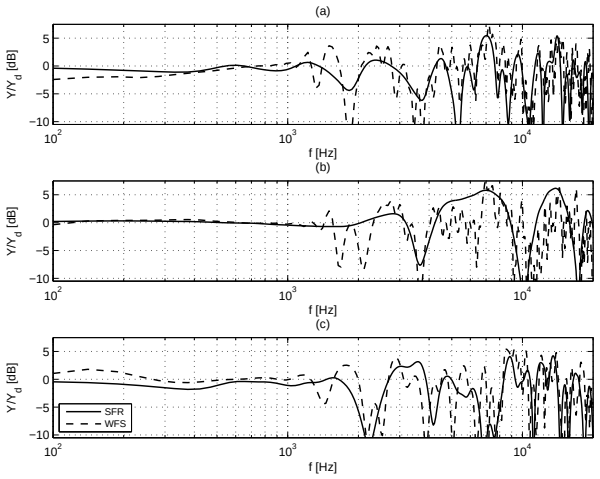


Fig. 26: Magnitude frequency responses of SFR and WFS in the control points $R_S(8\ m,\ 0)$ (a), $R_C(8\ m,\ 2\ m)$ (b), and $R_E(8\ m,\ 4\ m)$ (c) relative to a point source located at $\boldsymbol{r}_m = (3\ m,\ 1\ m)$.

From Figs. 26(a) and 26(c) it can be seen that while WFS causes small coloration at low frequencies up to 1 kHz at both ends of the reference line, SFR is much closer to the desired response in this frequency range in all three control points.

In the central point R_C, WFS starts having notable coloration artifacts (over 3 dB) above around 1.5 kHz, and above slightly lower frequencies in the points R_S and R_E. SFR, on the other hand, causes notable coloration in the point R_C only above 3 kHz, while in the two end point R_S and R_E the coloration becomes significant just below 2 kHz.

Additionally, Fig. 27 shows the relative magnitude response error

$$E_r(f) = \frac{Y(f) - Y_d(f)}{Y_d(f)}$$

of WFS and SFR averaged along the reference line.

4.3. Discussion

Considering the previous group delay analysis, it can be said that SFR provides superior time-domain performance on the reference line. Not only does SFR have less perceptible group delay errors ("post-echo"

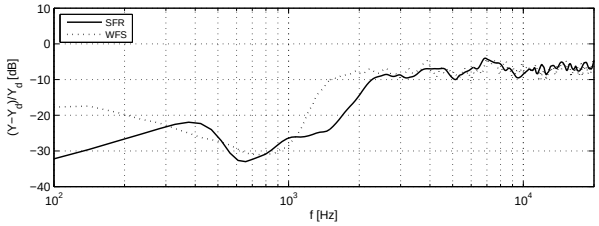


Fig. 27: Relative magnitude response error $E_r(f) = \frac{Y(f)-Y_d(f)}{Y_d(f)}$ averaged along the control points on the reference line (see Fig. 1).

artifacts) in every part of the listening line, but it also provides perceptually correct group delay response up to above 5 kHz, as opposed to 1.5 kHz by WFS.

From the frequency analysis of the impulse responses on the reference line, two observation can be made. Firstly, it is apparent that the low frequency response of SFR exhibits virtually no coloration up to almost 2 kHz in all parts of the reference line; on the other hand, WFS suffers from slight coloration at low frequencies in the points away from the reference line center. Secondly, it can be seen that in all tested points, more noticeable coloration starts at a higher frequency with SFR compared to WFS. Quantitatively, assuming that the aliasing occurs when the average relative magnitude response error reaches −10 dB, one can say that SFR suffers from aliasing artifacts above about 2.1 kHz, whereas WFS exhibits notable errors due to aliasing above 1.4 kHz (see Fig. 27).

It should be emphasized that even if the simulations presented in this section were done only for a single primary source location relatively closely behind the loudspeaker array, their results hold more generally for sources which are not far from the loudspeaker array (up to around the distance equal to the length of the loudspeaker array). For sources further away from the loudspeaker array—i.e., for plane wave reproduction—SFR performs only slightly better than WFS, and further improvements are a topic of current investigations.

5. CONCLUSIONS

This paper has presented procedures for designing practical, discrete-time finite impulse response (FIR) loudspeaker filters for two multichannel sound field reproduction techniques—Wave Field Synthesis (WFS) and Sound Field Reconstruction (SFR)—as both are frequency-based and provide frequency-domain descriptions of loudspeaker filters.

WFS filters obtained from a theoretical analysis [2] have practical limitations due to diffraction and aliasing. These problems were addressed through tapering the loudspeaker array edges [11] and power normalization on the reference line in order to avoid serious reproduction errors, especially at high frequencies. By modeling the loudspeaker filters through a superposition of a delay and an FIR filter, a computationally efficient set of discrete-time WFS filters was obtained by transforming the desired frequency characteristics to the time domain and subsequent pruning of the obtained impulse responses.

Initial SFR filters are obtained by the use of a non-linear numerical procedure for matrix pseudo-inversion at different frequencies [5], providing a set of loudspeaker filters with discontinuities in the frequency responses. A smoothing filter is applied to mitigate the discontinuities and to effectively shorten impulse responses. As with WFS, the filters were modeled by a combination of a delay and FIR filter. The FIR filter was obtained similarly to WFS filters—by the use of an inverse discrete Fourier transform and subsequent impulse response pruning.

The obtained WFS and SFR discrete-time filters were compared through simulations. The comparison included an extended-area sound field reproduction of a point source emitting a sinusoid and a train of low-frequency pulses, and time- and frequency-domain analysis of the impulse responses at different control points in the listening area.

The simulations have shown that SFR outperforms WFS in terms of sound field reproduction accuracy for both sinusoidal sounds and trains of pulses. Furthermore, the SFR system reproduction causes lower coloration, group-delay errors, and effectively increases the aliasing frequency for a given setup compared to WFS.

6. REFERENCES

[1] AJ Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, no. 12, pp. 977–995, 1988.

[2] 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.

[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] M. Kolundžija, C. Faller, and M. Vetterli, "Sound Field Reconstruction: An Improved Approach for Wave Field Synthesis ," in *Preprint 126th Conv. Aud. Eng. Soc.*, May. 2009.

[6] E. Corteel, "Equalization in an extended area using multichannel inversion and wave field synthesis," *Journal of the Audio Engineering Society*, vol. 54, no. 12, pp. 1140–1161, 2006.

[7] 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.

[8] 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.

[9] E.G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic Press, 1999.

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

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

[12] 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.

[13] T.I. Laakso, V. Valimaki, M. Karjalainen, and U.K. Laine, "Splitting the unit delay—tools for fractional delay filter design," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, 1996.

[14] S. Flanagan, B.C.J. Moore, and M.A. Stone, "Discrimination of group delay in clicklike signals presented via headphones and loudspeakers," *Journal of the Audio Engineering Society*, vol. 53, no. 7-8, pp. 593–611, 2005.

[15] J. Blauert and P. Laws, "Group delay distortions in electroacoustical systems," *The Journal of the Acoustical Society of America*, vol. 63, pp. 1478, 1978.