DIRECTION PRESERVING WIENER MATRIX FILTERING FOR AMBISONIC INPUT-OUTPUT SYSTEMS

Adrian Herzog and Emanuël A.P. Habets

International Audio Laboratories Erlangen*, Am Wolfsmantel 33, 91058 Erlangen, Germany

ABSTRACT

We present a spatial matrix filtering framework for noise reduction in the spherical harmonics (ambisonics) domain (SHD), which outputs an SHD signal vector rather than one signal as commonly provided by beamforming approaches. We discuss two spatial matrix filtering methods: A multi-beamformer method using known propagation vectors of the desired signal components and a method preserving the directional information in an optimal way. Parametric multi-channel Wiener filter solutions for both methods are discussed and a performance evaluation is conducted. It is shown that the direction preserving method preserves the spatial distribution of the desired sounds and residual noise at the cost of less noise reduction and higher signal distortion when compared to the multi-beamformer approach. Moreover, no spatial parameters have to be estimated.

Index Terms— Spatial filtering, ambisonics, spherical harmonics, noise reduction

1. INTRODUCTION

Microphone arrays can be used to capture a desired sound source while suppressing undesired sources in the recorded sound field. A common approach is to apply a beamformer to the microphone signals and extract an estimate of the desired sound which, in this case, results in a single-channel signal. To extract several sources from the array data, one can apply several beamformers [1]. Nevertheless, the spatial information of the original sound field is lost after beamforming.

In [2] an informed beamformer is derived which yields a specific spatial response while suppressing noise using direction-of-arrival (DOA) estimates of the desired sources. One might use this technique for spatial sound reproduction by applying several such beamformers, each with the desired response of a loudspeaker [2]. However, the beamformer yields the desired spatial response only for the desired source directions. Therefore, the spatial distribution of the residual noise cannot be controlled. Moreover, instantaneous DOAs have to be estimated which is often difficult, especially for 3D scenarios. Another approach, developed in the context of binaural processing, is to calculate two beamformers for the left and right ear to directly generate a filtered binaural signal such that the binaural cues are preserved [3, 4, 5]. In a recent work, the desired signal, which is assumed to be a single directional source, is first extracted using an adaptive beamformer and then spatially reconstructed by estimating the transfer function from the source to the array [6].

The aforementioned methods can be formulated in the spherical harmonics domain (SHD) which has many benefits over other sound field representations such as scalability, i.e., the spatial resolution scales with the number of spherical harmonic coefficients used, and implementation of rotations (e.g., for head tracking) with matrix multiplications. The SHD is an efficient representation of sound fields on the surface of a sphere, which can be captured using spherical microphone arrays.

As discussed before, the spatial information of the sound field is lost after beamforming or only captured by a few parameters which might be difficult to estimate. To preserve the spatial information of the sound field we propose to use a filter matrix which extracts a signal vector in the same spatial domain as the input signal. Two approaches are discussed. The first approach uses multiple beamformers and knowledge of the number of directional sources and their propagation vectors. The second approach uses a special direction preserving form of the filter matrix, which can be related to the ambisonics directional-loudness modification discussed in [7].

In Section 2, we introduce the spherical harmonics transform and mode-strength compensation. In Section 3, we discuss the proposed spatial matrix filtering approaches. In Section 4, the corresponding solutions of the parametric multi-channel Wiener filter matrix are derived. In Section 5, the performance of these methods is evaluated.

2. SPHERICAL HARMONICS TRANSFORM

The spherical harmonics transform (SHT) is used to efficiently represent signals captured on a 2-sphere $\mathcal{S}^2$ by a countable set of SHD coefficients of order $l = 0, ..., L$ and modes $m = -l, ..., l$, where the spatial resolution scales with the maximum order L. Let a solid angle be denoted by $\Omega = (\theta, \phi)$ with elevation $\theta \in [0, \pi]$ and azimuth $\phi \in [-\pi, \pi)$. The SHT of a function $f : \mathcal{S}^2 \to \mathbb{C}$ is defined as [8, 9]:

$$f_{lm} := \int_{\mathcal{S}^2} f(\Omega) Y_{lm}^*(\Omega)\, \mathrm{d}\Omega\,, \tag{1}$$

where * denotes the complex conjugate, $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta \mathrm{d}\phi$ and $Y_{lm}(\Omega)$ is the spherical harmonic (SH) of order l and mode m. The SHs constitute an orthonormal basis for functions on $\mathcal{S}^2$. For $L \to \infty$ this basis is complete. The number of SHD coefficients up to order L is $(L+1)^2$. If we sample the sound-field with P microphones on a sphere at directions Ω_1, ..., Ω_P, one can approximate the SHT (1) by

$$f_{lm} \approx \sum_{p=1}^{P} \mathsf{q}_p f(\Omega_p) Y_{lm}^*(\Omega_p)\,, \tag{2}$$

where the *sampling weights* q_p are chosen such that

$$\sum_{p=1}^{P} \mathsf{q}_p Y_{lm}(\Omega_p) Y_{l'm'}^*(\Omega_p) = \delta_{ll'} \delta_{mm'} \tag{3}$$

is fulfilled [8, 9]. For uniform spatial sampling, one yields $\mathsf{q}_p = \frac{4\pi}{P}$ [10] if $P \geq (L+1)^2$. The SHD coefficients of a unit-amplitude

*A joint institution of the Friedrich Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

plane-wave with DOA Ω_{PW} are given by [11]

$$f_{\text{PW},lm} = b_l(\kappa r) Y_{lm}^*(\Omega_{\text{PW}}) \,, \tag{4}$$

where κ denotes the wavenumber and r the radius of the 2-sphere. The function b_l is called the mode strength of order l and depends on the spherical array type. To compensate for the array-dependency of the SHD signals one can divide the signals by the respective mode strengths [11], yielding mode-strength compensated SHD (MC-SHD) signals, which are also referred to as *ambisonic signals*. In what follows we use a vector notion for the SHs

$$\mathbf{y}(\Omega) := [Y_{00}(\Omega), Y_{1-1}(\Omega), ..., Y_{LL}(\Omega)]^T \tag{5}$$

and MC-SHD signals

$$\mathbf{x} := [X_{00}, X_{1-1}, ..., X_{LL}]^T \,, \tag{6}$$

where $(\cdot)^T$ denotes the transpose and X_{lm} the lm'th coefficient of an MC-SHD signal. Moreover, we assume that the signals captured by the P microphones are transformed to the short-time-Fourier-transform (STFT) domain, where we denote time frame and frequency bin indices by n and k, before the SHT. STFT indices are omitted for brevity when possible.

3. SPATIAL MATRIX FILTERING

To extract a desired signal S from a multi-channel signal vector $\mathbf{x}$, one usually applies a beamformer $\mathbf{w}$ to $\mathbf{x}$ resulting in a mono signal

$$Z = \mathbf{w}^H \mathbf{x} \,, \tag{7}$$

where $(\cdot)^H$ denotes the conjugate transpose, which is the estimate of S. This method has been extensively studied in literature, even in the SHD [9], and is used for e.g., denoising and dereverberation. However, as only one signal is extracted from $\mathbf{x}$, the spatial information encoded in $\mathbf{x}$ is lost. To maintain this information we propose to extract an MC-SHD signal vector $\mathbf{z}$ by applying a spatial filter matrix $\mathbf{W}$ to $\mathbf{x}$ in the MC-SHD, such that

$$\mathbf{z} = \mathbf{W}\mathbf{x} \,. \tag{8}$$

We, therefore, apply an ambisonic transformation instead of a beamformer. In the following sections, we derive explicit forms of the spatial filter matrix using two different approaches.

3.1. Multi-Beamformer Approach

Let us denote the desired MC-SHD signal vector consisting of J sources $\check{S}_j$ with propagation vectors $\mathbf{d}_j$, $j = 1, ..., J$ by

$$\mathbf{s} = \sum_{j=1}^{J} \check{S}_j \mathbf{d}_j =: \mathbf{D}\check{\mathbf{s}} \,, \tag{9}$$

where $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_J]$ and $\check{\mathbf{s}} = [\check{S}_1, ..., \check{S}_J]^T$. A source extraction approach is to apply J beamformers $\mathbf{w}_j$ to the signal $\mathbf{x}$ which yield estimates Z_j for the J signals S_j:

$$\check{\mathbf{z}} := [\check{Z}_1, ..., \check{Z}_J]^T = [\mathbf{w}_1, ..., \mathbf{w}_J]^H \mathbf{x} =: \check{\mathbf{W}}\mathbf{x} \tag{10}$$

A commonly used beamformer is the linearly-constrained-minimum-variance (LCMV) beamformer [12]. Given the signal model in (9) and estimates of the separate sources $\check{\mathbf{z}}$ the estimate of $\mathbf{s}$ is

$$\mathbf{z} = \mathbf{D}\check{\mathbf{z}} = \mathbf{D}\check{\mathbf{W}}\mathbf{x} \,. \tag{11}$$

Therefore, the corresponding filter matrix is $\mathbf{W} = \mathbf{D}\check{\mathbf{W}}$. This approach is well suitable if the propagation vectors are available or can be estimated. Nevertheless, all signal components which can not be modelled with (9) such as diffuse sound will not maintain their spatial characteristics after filtering.

3.2. Direction Preserving Approach

To ensure that the spatial information of $\mathbf{x}$ is maintained, we want that applying $\mathbf{W}$ to a plane-wave from an arbitrary direction Ω the plane-wave maintains its direction. This yields the following condition in the MC-SHD:

$$\mathbf{W}\mathbf{y}^*(\Omega) = \alpha(\Omega)\mathbf{y}^*(\Omega) \quad \forall \Omega \in \mathcal{S}^2 \,, \tag{12}$$

where $\alpha(\Omega)$ is a directional gain function. For order limited SHs this condition can only be fulfilled for a finite number of directions Ω. Nevertheless, we can find the optimal solution for $\mathbf{W}$ by minimizing the following cost function:

$$\mathcal{J}_{\text{DP}}(\mathbf{W}) := \int_{\mathcal{S}^2} \|\mathbf{W}\mathbf{y}^*(\Omega) - \alpha(\Omega)\mathbf{y}^*(\Omega)\|^2 \, \mathrm{d}\Omega \,, \tag{13}$$

where $\|\cdot\|$ denotes the ℓ^2-norm. The optimal solution is

$$\mathbf{W} = \int_{\mathcal{S}^2} \alpha(\Omega)\mathbf{y}^*(\Omega)\mathbf{y}^T(\Omega) \, \mathrm{d}\Omega \,. \tag{14}$$

Comparing this solution with the solution provided in [7] we can see that this is precisely the directional-loudness ambisonic transformation matrix but for order truncated SHs. To avoid the integration, we can make use of the fact that the tensor product of SH vectors up to order L can be expressed as a weighted sum of $(2L+1)^2$ basis matrices, which are related to the Clebsch-Gordan coefficients [13]. The weights are the SHs up to order $2L$. Therefore, if the matrix consisting of SHs up to order $2L$ sampled at $Q \geq (2L+1)^2$ directions Ω_1, ... Ω_Q is of maximum rank, the matrices $\mathbf{y}^*(\Omega_q)\mathbf{y}^T(\Omega_q)$ with $q = 1, ..., Q$ provide an (over-)complete basis for $\{\mathbf{y}^*(\Omega)\mathbf{y}^T(\Omega) | \Omega \in \mathcal{S}^2\}$ and (14) can be replaced by

$$\mathbf{W} = \sum_{q=1}^{Q} \mathfrak{q}_q \alpha_q \mathbf{y}^*(\Omega_q)\mathbf{y}^T(\Omega_q) \tag{15}$$

with discrete directional gains α_q and sampling weights $\mathfrak{q}_q$ corresponding to the sampling scheme.

4. PARAMETRIC MULTICHANNEL WIENER FILTER MATRIX

Suppose the signal $\mathbf{x}$ consists of a desired signal $\mathbf{s}$ and an undesired/noise signal $\mathbf{v}$, i.e.,

$$\mathbf{x} = \mathbf{s} + \mathbf{v} \,, \tag{16}$$

which are assumed to be uncorrelated. We search for an estimate $\mathbf{z} = \mathbf{W}\mathbf{x}$ of $\mathbf{s}$. The parametric multichannel Wiener filter (PMWF) [14, 15] solution for the filter matrix $\mathbf{W}$ is derived by minimizing the following cost function:

$$\mathcal{J}_{\text{PMWF}}(\mathbf{W}) = E\{\|\mathbf{W}\mathbf{s} - \mathbf{s}\|^2\} + \mu E\{\|\mathbf{W}\mathbf{v}\|^2\} \,, \tag{17}$$

where $E\{\cdot\}$ denotes the expectation operator. The first term of $\mathcal{J}_{\text{PMWF}}$ is a measure for the desired signal distortion, the second term a measure for noise reduction and the parameter μ adjusts the trade-off between these two measures.

4.1. Multi-Beamformer Solution

Inserting (9) and $\mathbf{W} = \mathbf{D}\check{\mathbf{W}}$ in the cost function (17) we get:

$$\mathcal{J}_{\text{PMWF}}(\mathbf{W}) = E\left\{(\check{\mathbf{W}}\mathbf{D}\check{\mathbf{s}} - \check{\mathbf{s}})^H \mathbf{D}^H \mathbf{D}(\check{\mathbf{W}}\mathbf{D}\check{\mathbf{s}} - \check{\mathbf{s}})\right\} + \mu E\left\{(\check{\mathbf{W}}\mathbf{v})^H \mathbf{D}^H \mathbf{D}(\check{\mathbf{W}}\mathbf{v})\right\} . \quad (18)$$

Minimizing w.r.t. $\check{\mathbf{W}}^H$ and assuming that $\mathbf{D}^H\mathbf{D}$ is invertible yields the following optimal solution:

$$\check{\mathbf{W}} = \mathbf{\Phi}_{\check{\mathbf{s}}}\mathbf{D}^H \left(\mathbf{D}\mathbf{\Phi}_{\check{\mathbf{s}}}\mathbf{D}^H + \mu\mathbf{\Phi}_{\mathbf{v}}\right)^{-1} , \quad (19)$$

where $\mathbf{\Phi}_{(\cdot)} := E\{(\cdot)(\cdot)^H\}$ denotes the power spectral density (PSD) matrix of the respective signal vector. The same solution for $\mu = 1$ has been derived in [1] and it is shown that (19) can be decomposed into J LCMV beamformers $\mathbf{w}_{\text{LCMV},j}$ and a multi-source (parametric) Wiener postfilter matrix $\mathbf{W}_{\text{WPF}}$. A lower bound of the Wiener postfilters can be implemented by further decomposing the non-diagonal $\mathbf{W}_{\text{WPF}}$ (see [1] for more details). The filter matrix is derived using $\mathbf{W} = \mathbf{D}\check{\mathbf{W}}$.

4.2. Direction Preserving Solution

For the derivation of the direction preserving PMWF matrix it is useful to define the following quantities:

$$\begin{aligned}\Phi_{(\tilde{\cdot}),q'q} &:= \mathbf{y}^T(\Omega_{q'})\mathbf{\Phi}_{(\cdot)}\mathbf{y}^*(\Omega_q)\\ \tilde{\Delta}_{qq'} &:= \mathfrak{q}_q^*\mathfrak{q}_{q'}\mathbf{y}^T(\Omega_q)\mathbf{y}^*(\Omega_{q'}) .\end{aligned} \quad (20)$$

We insert the optimum direction preserving form of $\mathbf{W}$ (15) in the cost function (17) and minimize it w.r.t. α_q^* for $q = 1, ..., Q$. This yields the following set of equations:

$$\sum_{q'=1}^{Q} \tilde{\Delta}_{qq'}(\Phi_{\tilde{\mathbf{s}},q'q} + \mu\Phi_{\tilde{\mathbf{v}},q'q})\alpha_{q'} = \sum_{q'=1}^{Q} \tilde{\Delta}_{qq'}\Phi_{\tilde{\mathbf{s}},q'q} . \quad (21)$$

Solving this set of equations for the directional gains α_q requires a matrix inversion per time and frequency bin (recall that time-frame and frequency bin indices are omitted for brevity). To reduce the computational complexity, we can make use of the fact that $\mathbf{y}^T(\Omega)\mathbf{y}^*(\Omega')$ has a maximum at $\Omega = \Omega'$ and decays similarly to a sinc-function for increasing angular distance [8]. Therefore, we propose to neglect the off-diagonal contributions of $\tilde{\Delta}_{qq'}$ in (21). Using this approximation, one can derive

$$\alpha_q = \frac{\Phi_{\tilde{\mathbf{s}},qq}}{\Phi_{\tilde{\mathbf{s}},qq} + \mu\Phi_{\tilde{\mathbf{v}},qq}} , \quad (22)$$

which has the form of a single-channel parametric Wiener filter. A lower bound can easily be implemented by lower-bounding the α_q. Finally, the filter matrix $\mathbf{W}$ is derived by inserting (22) into (15).

4.3. Desired Signal PSD Estimation

Given an estimate of the noise PSD matrix $\mathbf{\Phi}_{\mathbf{v}}$, the signal PSDs $\Phi_{\check{\mathbf{s}},jj}$ for the multi-beamformer approach and $\Phi_{\tilde{\mathbf{s}},qq}$ for the direction preserving approach can be recursively estimated using the decision-directed method [16]. This yields the following update equations:

$$\begin{aligned}\Phi_{\check{\mathbf{s}},jj}(n,k) &= \beta|\check{Z}_j(n-1,k)|^2\\ &\quad + (1-\beta)\max\left(|Z_{\text{LCMV},j}(n,k)|^2 - \Psi_{\mathbf{v},jj}(n,k), 0\right) ,\\ \Phi_{\tilde{\mathbf{s}},qq}(n,k) &= \beta|\tilde{Z}_q(n-1,k)|^2\\ &\quad + (1-\beta)\max\left(|\tilde{X}_q(n,k)|^2 - \Phi_{\tilde{\mathbf{v}},qq}(n,k), 0\right)\end{aligned} \quad (23)$$

for $j = 1, ..., J$ and $q = 1, ..., Q$, where we have defined $Z_{\text{LCMV},j} = \mathbf{w}_{\text{LCMV},j}^H\mathbf{x}$, $\mathbf{\Psi}_{\mathbf{v}} = (\mathbf{D}^H\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{D})^{-1}$, $\tilde{X}_q = \mathbf{y}^T(\Omega_q)\mathbf{x}$ and $\tilde{Z}_q = \alpha_q\tilde{X}_q$. The parameter $\beta \in [0, 1]$ is a recursive smoothing parameter.

5. EVALUATION

To compare the noise reduction performance of the proposed matrix spatial filters, we apply both approaches to synthesized sound fields with two directional sources and diffuse noise.

5.1. Setup

Two different scenarios (Sc1 and Sc2) were examined:
Sc1: Two english speech files, 1 male and 1 female, of 20 seconds length, were placed at $\Omega_{s_1} = [75°, 70°]$ and $\Omega_{s_2} = [115°, -60°]$.
Sc2: 100 sound fields with closely spaced directional sources were generated using the following procedure: One male and 1 female speech file are randomly selected from a set of 10 male and 10 female english speech files of 5 seconds length. The directions Ω_{s_1} and Ω_{s_2} are randomly generated such that $\angle(\Omega_{s_1}, \Omega_{s_2}) \leq 20°$.

All speech files had a sampling frequency of $f_s = 16$ kHz and were transformed in the STFT domain using a square-root-Hann window with 512 samples (32 ms) length, 50% overlap (16 ms) and a discrete Fourier transform size of 1024. Let us denote these STFT signals with $S_1(n, k)$ and $S_2(n, k)$. Assuming no reverberation, the MC-SHD plane-wave signals were generated by multiplying the single-channel STFT signals with SH vectors at source directions Ω_{s_1} and Ω_{s_2}, yielding

$$\mathbf{s}(n, k) = S_1(n, k)\mathbf{y}^*(\Omega_{s_1}) + S_2(n, k)\mathbf{y}^*(\Omega_{s_2}) . \quad (24)$$

The diffuse noise signal vector $\mathbf{v}$ was calculated by almost uniformly sampling the sphere at $Q_v = 50$ positions and forming a random superposition of the corresponding SH vectors at each n and k:

$$\mathbf{v}(n, k) = \sqrt{\frac{\sigma_v^2}{Q_v}} \sum_{q_v=1}^{Q_v} R_{q_v}(n, k)\mathbf{y}^*(\Omega_{q_v}) , \quad (25)$$

where the R_{q_v} are complex white Gaussian noise processes of unit variance in the STFT domain and σ_v^2 is the noise variance. Almost uniform sampling schemes were derived using a repelling-charged-particles-on-sphere model [17]. Assuming perfect uniform sampling, the noise PSD matrix is given by $\mathbf{\Phi}_{\mathbf{v}} = (\sigma_v^2/4\pi)\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The noise variance σ_v^2 is calculated from a chosen input signal-to-noise ratio iSNR and the speech variance σ_s^2, which is derived from the speech signals $S_1(n, k)$ and $S_2(n, k)$ as follows:

$$\sigma_s^2 = \underset{n\in\mathcal{N}}{\text{mean}}\, E(n) \;\text{ with }\; E(n) = \underset{k}{\text{mean}}\left(\sum_{j=1}^{2}|S_j(n, k)|^2\right) \quad (26)$$

and

$$\mathcal{N} := \left\{n \in \{1, ..., N\} \mid E(n) > 0.01 \cdot \max_n(E(n))\right\} . \quad (27)$$

For the direction preserving PMWF, the sphere was sampled almost uniformly at $Q = (2L + 1)^2$ directions. For the multi-beamforming approach, we incorporated knowledge of Ω_{s_1} and Ω_{s_2} in the algorithm. Therefore, for our non-reverberant scenario, the propagation vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ are given by $\mathbf{d}_j = \sqrt{4\pi}\mathbf{y}^*(\Omega_{s_j})$ for $j = 1, 2$. We chose a maximum SHD order of $L = 3$, iSNR $= 3$ dB, a lower bound for the PMWFs of 0.1, $\mu = 1$ and $\beta = 0.8$ for the decision-directed PSD estimation.

5.2. Performance Measures

Let us first denote the residual noise by $\mathbf{v}_{\text{res}} := \mathbf{W}\mathbf{v}$ and the residual desired sound by $\mathbf{s}_{\text{res}} := \mathbf{W}\mathbf{s}$. We then define the *mean noise reduction* (NR) as follows:

$$\text{NR} := 10 \log_{10} \left(\frac{\underset{n,k}{\text{mean}} \|\mathbf{v}(n,k)\|^2}{\underset{n,k}{\text{mean}} \|\mathbf{v}_{\text{res}}(n,k)\|^2} \right) . \tag{28}$$

Next, we define the *mean desired signal distortion* (SD):

$$\text{SD} := 10 \log_{10} \left(\frac{\underset{n,k}{\text{mean}} \|\mathbf{s}_{\text{res}}(n,k) - \mathbf{s}(n,k)\|^2}{\underset{n,k}{\text{mean}} \|\mathbf{s}(n,k)\|^2} \right) . \tag{29}$$

Finally, we are interested in the *spatial distribution of the residual noise power*:

$$P_{\mathbf{v}_{\text{res}}}(\Omega) := \underset{n,k}{\text{mean}} \left(\frac{|\mathbf{y}^T(\Omega)\mathbf{v}_{\text{res}}(n,k)|^2}{\mathbf{y}^T(\Omega)\mathbf{y}^*(\Omega)} \right) . \tag{30}$$

To analyse the *similarity of the spatial distributions* of two MC-SHD signal vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, we use the following similarity measure:

$$\sigma(\mathbf{x}_1, \mathbf{x}_2) := \frac{2}{\pi} \sin^{-1} \left(\frac{\sum_i P_{\mathbf{x_1}}(\Omega_i) P_{\mathbf{x_2}}(\Omega_i)}{\sqrt{\sum_i P_{\mathbf{x_1}}(\Omega_i)^2} \sqrt{\sum_i P_{\mathbf{x_2}}(\Omega_i)^2}} \right) , \tag{31}$$

where $P_{\mathbf{x}_{1/2}}(\Omega)$ is defined analogous to (30) and the sum goes over a densely sampled finite subset of $\mathcal{S}^2$. The similarity measure takes values $\sigma(\mathbf{x}_1, \mathbf{x}_2) \in [0, 1]$ and can be identified as the angular similarity [18] of the two spatial distributions.

5.3. Results

In Table 1 the NR, SD and similarity results for Sc1 are summarized. The multi-beamforming method yields higher noise reduction which is expected as the denoised signal $\mathbf{z} = \mathbf{D}\check{\mathbf{z}}$ is a superposition of two directional signals. Therefore, apart from the directions Ω_{s_1} and Ω_{s_2}, the noise is highly suppressed. Nevertheless, the residual noise power at those directions is actually slightly higher compared to the direction preserving approach (see Fig. 1). The multi-beamforming method yields less desired signal distortion, which is probably due to the fact that, with this approach, noise is already reduced with distortion-less LCMV beamformers before the Wiener postfilters are applied.

In Fig. 1 the spatial distribution of the residual noise power for Sc1 is shown. The direction preserving approach almost completely preserves the isotropic character of the diffuse noise, while for the multi-beamforming approach the residual noise is concentrated at source directions Ω_{s_1} and Ω_{s_2}.

In Fig. 2 the noise reduction is shown as a function of the angular distance of the two sources (Sc2). With decreasing angular distance, the multi-beamforming approach yields less noise reduction, while for the direction preserving approach the noise reduction does not significantly depend on the angular distance.

Note that for the multi-beamforming approach the source propagation vectors $\mathbf{d}_1, \mathbf{d}_2$ have to be known while for the direction preserving method this information is not used and no spatial properties of the desired signal $\mathbf{s}$ are assumed.

Method	NR [dB]	SD [dB]	$\sigma(\mathbf{v}_{\text{res}}, \mathbf{v})$	$\sigma(\mathbf{s}_{\text{res}}, \mathbf{s})$
Multi-Beam.	17.96	−21.24	0.30	1.00
Dir. Pres.	11.66	−17.29	0.93	0.99

Table 1. Mean noise reduction and signal distortion results (Sc1)

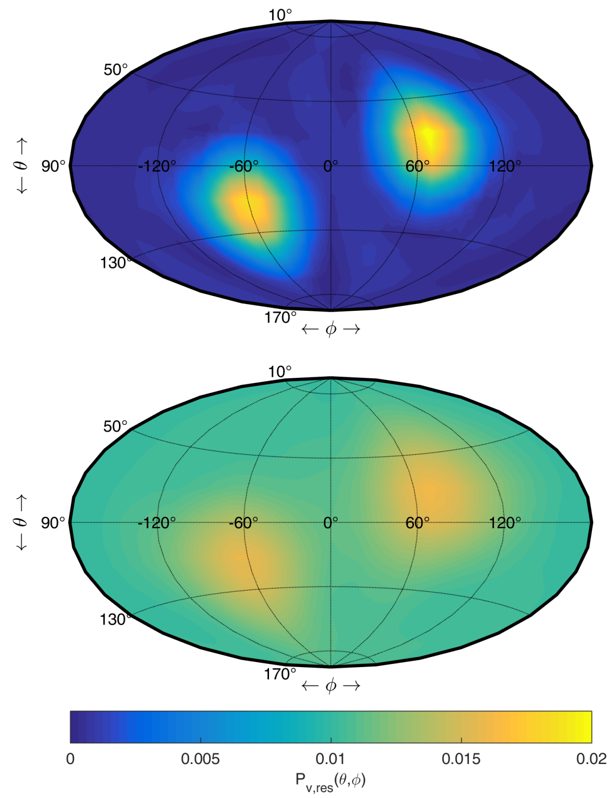

Fig. 1. Residual spatial noise power for multi-beamforming (top) and direction-preserving (bottom) matrix PMWF (Sc1)

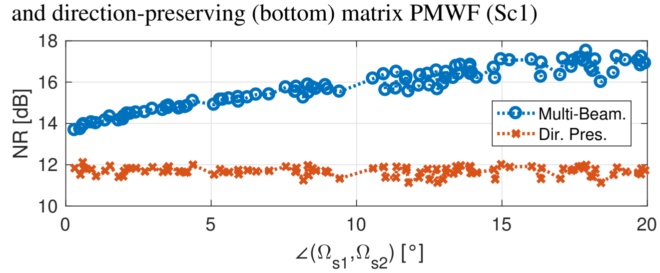

Fig. 2. Noise reduction over angular distance (Sc2)

6. CONCLUSION

To preserve the spatial information of a signal vector in the MC-SHD after spatial filtering, we proposed to apply a filter matrix instead of a beamformer to the signal vector which outputs an MC-SHD signal. We discussed two noise reduction methods using such a filter matrix. The first method uses the multi-source multi-channel Wiener filter method [1] and the propagation matrix of the desired sources. The second method was derived by demanding that the filter-matrix should preserve the directional information of any sound field in an optimal way. In the evaluation, we have seen that the first method yields higher noise reduction and less speech distortion, if the propagation vectors are known, but, in contrast to the direction preserving method, the spatial distribution of the noise is not preserved.

7. REFERENCES

[1] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multispeaker LCMV beamformer and postfilter for source separation and noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 940–951, May 2017.

[2] O. Thiergart, M. Taseska, and E.A.P. Habets, "An informed parametric spatial filter based on instantaneous direction-of-arrival estimates," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 12, pp. 2182–2196, Dec. 2014.

[3] N. R. Shabtai and B. Rafaely, "Spherical array beamforming for binaural sound reproduction," in *Proc. 27th Conv. of Electrical and Electronics Engineers in Israel (IEEEI)*, Eilat, Israel, 2012.

[4] N. R. Shabtai and B. Rafaely, "Binaural sound reproduction beamforming using spherical microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[5] N. R. Shabtai and B. Rafaely, "Generalized spherical array beamforming for binaural speech reproduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 238–247, Jan. 2014.

[6] C. Borrelli, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "A denoising methodology for higher order ambisonics recordings," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Tokyo, Japan, Sept. 2018.

[7] M. Kronlachner and F. Zotter, "Spatial transformations for the enhancement of ambisonic recordings," in *Proc. Intl. Conf. on Spatial Audio*, Jan. 2014.

[8] B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8, Springer, 2015.

[9] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, vol. 9, Springer, 2017.

[10] J. Meyer and G. W. Elko, "Spherical microphone arrays for 3d sound recordings," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds., pp. 67–89. Kluwer Academic Publishers, Norwell, MA, USA, 2004.

[11] B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.

[12] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[13] D. M. Brink and G. R. Satchler, *Angular Momentum*, Oxford University Press, second edition, 1968.

[14] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7–8, pp. 636–656, Aug. 2007.

[15] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 260–276, 2010.

[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[17] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, Apr. 1999.

[18] T. Apaydin and H. Ferhatosmanoglu, "Access structures for angular similarity queries," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 1512–1525, Nov. 2006.