AN INFORMED LCMV FILTER BASED ON MULTIPLE INSTANTANEOUS DIRECTION-OF-ARRIVAL ESTIMATES

Oliver Thiergart and Emanuël A. P. Habets

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen, Germany
{oliver.thiergart, emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

Extracting sound sources in noisy and reverberant conditions remains a challenging task that is commonly found in modern communication systems. In this work, we consider the problem of obtaining a desired spatial response for at most L simultaneously active sound sources. The proposed spatial filter is obtained by minimizing the diffuse plus self-noise power at the output of the filter subject to L linear constraints. In contrast to earlier works, the L constraints are based on instantaneous narrowband direction-of-arrival estimates. In addition, a novel estimator for the diffuse-to-noise ratio is developed that exhibits a sufficiently high temporal and spectral resolution to achieve both dereverberation and noise reduction. The presented results demonstrate that an optimal tradeoff between maximum white noise gain and maximum directivity is achieved.

***Index Terms*—** microphone array processing, optimal beamforming, diffuse-to-noise ratio estimation

1. INTRODUCTION

Extracting sound sources in noisy and reverberant conditions is commonly found in modern communication systems. In the last four decades, a large variety of spatial filtering techniques have been proposed to accomplish this task. Existing spatial filters are optimal when the assumed signal model is valid and when the information required to compute the filters is accurate. In practice, however, the signal model is often violated and estimating the required information becomes a major challenge.

Existing spatial filters can be broadly classified into linear spatial filters [1–4] and parametric spatial filters [5–8]. In general, linear spatial filters require an estimate of the propagation vector(s) or the second-order statistics (SOS) of the desired source(s) plus the SOS of the interference. Some spatial filters are designed to extract a single source signal (either reverberant or dereverberated) [9–16], while others have been designed to extract the sum of two or more reverberant source signals [17–19]. The aforementioned methods require prior knowledge of the direction of the desired source(s) or a period in which only the desired sources are active (either separately or simultaneously). A drawback of these methods is the inability to adapt sufficiently quickly to new situations (e. g., source movements, competing speakers that become active when the desired source is active). Parametric spatial filters are often based on a relatively simple signal model (i. e., the received signal in the time-frequency domain consists of a single plane wave plus diffuse sound) and are computed based on instantaneous estimates of the model parameters. The principle advantages of parametric spatial filters are a highly flexible directional response, a comparatively strong suppression of diffuse sound and interferers, and the ability to quickly adapt to new situations. However, as shown in [20], the underlying single plane wave signal model can easily be violated in practice which strongly degrades the performance of the parametric spatial filters.

In this work, we consider the problem of obtaining a desired, arbitrary spatial response for at most L sound sources being simultaneously active per time-frequency instant. We solve this problem by incorporating instantaneous parametric information about the acoustic scene into the design of a linearly constrained minimum variance (LCMV) filter resulting in an *informed LCMV filter*. The proposed informed spatial filter is obtained by minimizing the diffuse plus self-noise power subject to L linear constraints. In contrast to earlier works, the L constraints are based on instantaneous direction-of-arrival (DOA) estimates and the resulting responses to the L DOAs correspond to the specific desired directivity. In addition, a novel estimator for the diffuse-to-noise ratio (DNR) is developed which exhibits a sufficiently high temporal and spectral resolution to reduce both reverberation and noise. The presented results demonstrate that an optimal tradeoff between maximum white noise gain (WNG) and maximum directivity is achieved. In addition, the proposed spatial filter exhibits a short response time.

The remainder of the paper is organized as follows: Section 2 formulates the problem. In Sec. 3, two existing optimal linear filters and the proposed optimal filter are described. In Sec. 4, it is shown how the L DOAs and the DNR are estimated. The performance of the proposed spatial filter is evaluated in Sec. 5. Section 6 draws the conclusions.

2. PROBLEM FORMULATION

Let us consider an uniform linear array (ULA) of M omnidirectional microphones located at $\mathbf{d}_{1\ldots M}$. For each time-frequency instant we assume a sound field composed of $L < M$ plane waves (directional sound) propagating in an isotropic and spatially homogenous diffuse sound field. The microphone signals $\mathbf{x}(k,n) = [X(k,n,\mathbf{d}_1)\ldots X(k,n,\mathbf{d}_M)]^{\mathrm{T}}$ at frequency index k and time index n can be written as

$$\mathbf{x}(k,n) = \sum_{l=1}^{L} \mathbf{x}_l(k,n) + \mathbf{x}_{\mathrm{d}}(k,n) + \mathbf{x}_{\mathrm{n}}(k,n), \tag{1}$$

where $\mathbf{x}_l(k,n) = [X_l(k,n,\mathbf{d}_1)\ldots X_l(k,n,\mathbf{d}_M)]^{\mathrm{T}}$ contains the microphone signals that are proportional to the sound pressure of the l-th plane wave, $\mathbf{x}_{\mathrm{d}}(k,n)$ is the measured diffuse field, and $\mathbf{x}_{\mathrm{n}}(k,n)$ is the microphone self-noise.

*A joint institution of the University Erlangen-Nuremberg and Fraunhofer IIS, Germany

Assuming the three components in (1) are mutually uncorrelated, we can express the power spectral density (PSD) matrix of the microphone signals as

$$\begin{aligned}\mathbf{\Phi}(k,n) &= \mathrm{E}\left\{\mathbf{x}(k,n)\,\mathbf{x}^{\mathrm{H}}(k,n)\right\} \\ &= \sum_{l=1}^{L}\mathbf{\Phi}_l(k,n) + \mathbf{\Phi}_{\mathrm{d}}(k,n) + \mathbf{\Phi}_{\mathrm{n}}(k,n),\end{aligned} \tag{2}$$

with

$$\mathbf{\Phi}_{\mathrm{d}}(k,n) = \phi_{\mathrm{d}}(k,n)\,\mathbf{\Gamma}_{\mathrm{d}}(k), \tag{3}$$
$$\mathbf{\Phi}_{\mathrm{n}}(k,n) = \phi_{\mathrm{n}}(k,n)\,\mathbf{I}. \tag{4}$$

Here, $\mathbf{I}$ is an identity matrix, $\phi_{\mathrm{n}}(k,n)$ is the expected power of the microphone self-noise, which is identical for all microphones, and $\phi_{\mathrm{d}}(k,n)$ is the expected power of the diffuse field, which can vary rapidly across time and frequency. The ij-th element of the coherence matrix $\mathbf{\Gamma}_{\mathrm{d}}(k)$, denoted by $\gamma_{ij}(k)$, is the diffuse field coherence between microphone i and j. For instance for a spherically isotropic diffuse field, we have $\gamma_{ij}(k) = \mathrm{sinc}(\kappa\, r_{ij})$ [21] with wavenumber κ and $r_{ij} = ||\mathbf{d}_j - \mathbf{d}_i||$.

The directional sound $\mathbf{x}_l(k,n)$ in (1) can be written as

$$\mathbf{x}_l(k,n) = \mathbf{a}(k,\varphi_l)\,X_l(k,n,\mathbf{d}_1), \tag{5}$$

where $\varphi_l(k,n)$ is the DOA of the l-th plane wave ($\varphi = 0$ denoting the array broadside) and $\mathbf{a}(k,\varphi_l) = [a_1(k,\varphi_l)\ \ldots\ a_M(k,\varphi_l)]^{\mathrm{T}}$ is the propagation vector. The i-th element of $\mathbf{a}(k,\varphi_l)$,

$$a_i(k,\varphi_l) = \exp\{\jmath\,\kappa\, r_i \sin\varphi_l(k,n)\}, \tag{6}$$

describes the phase shift of the l-th plane wave from the first to the i-th microphone. Note that $r_i = ||\mathbf{d}_i - \mathbf{d}_1||$ is equal to the distance between the first and the i-th microphone.

The aim of the paper is to filter the microphone signals $\mathbf{x}(k,n)$ such that directional sounds arriving from specific spatial regions are attenuated or amplified as desired, while the diffuse sound and microphone self-noise are suppressed. The desired signal can therefore be expressed as

$$Y(k,n) = \sum_{l=1}^{L} G(k,\varphi_l)\,X_l(k,n,\mathbf{d}_1), \tag{7}$$

where $G(k,\varphi)$ is a real-valued arbitrary directivity function which can be frequency dependent. Figure 1 shows the magnitude of two example directivities $G_1(k,\varphi)$ and $G_2(k,\varphi)$. When using $G_1(k,\varphi)$ (solid line), we attenuate directional sound arriving from $\varphi < 45°$ by 21 dB while directional sound from other directions is not attenuated. In principle, one can design arbitrary directivities, even functions such as $G_2(k,\varphi)$ (dashed line). Moreover, $G(k,\varphi)$ can be designed time variant, e. g., to extract moving or emerging sound sources once they have been localized.

An estimate of the signal $Y(k,n)$ is obtained by a linear combination of the microphone signals $\mathbf{x}(k,n)$, i. e.,

$$\hat{Y}(k,n) = \mathbf{w}^{\mathrm{H}}(k,n)\,\mathbf{x}(k,n), \tag{8}$$

where $\mathbf{w}(k,n)$ is a complex weight vector of length M. It follows from (5) and (7) that $\mathbf{w}(k,n)$ has to satisfy the linear constraints

$$\mathbf{w}^{\mathrm{H}}(k,n)\,\mathbf{a}(k,\varphi_l) = G(k,\varphi_l), \quad l \in \{1,2,\ldots,L\}. \tag{9}$$

Moreover, the diffuse sound power and self-noise power at the filter's output has to be minimized. The corresponding optimal weight vector $\mathbf{w}(k,n)$ is derived in the next section. In the following, the dependency of the weights $\mathbf{w}(k,n)$ on k and n is omitted for brevity.

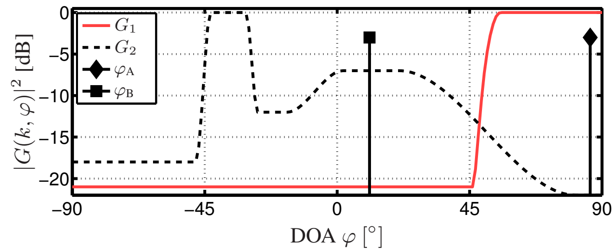


Fig. 1. Two arbitrary directivity functions & source positions

3. OPTIMAL SPATIAL FILTERS

3.1. Existing Spatial Filters

While the PSD $\phi_{\mathrm{n}}(k,n)$ can be estimated during periods of silence, $\phi_{\mathrm{d}}(k,n)$ is commonly assumed unknown and unobservable. We therefore consider two existing spatial filters that can be computed without this knowledge.

The first spatial filter is known as a delay-and-sum beamformer and minimizes the self-noise power at the filter's output (i. e., maximizes the WNG) [1]. The optimal weight vector that minimizes the mean squared error (MSE) between (7) and (8) subject to (9) is then obtained by

$$\mathbf{w}_{\mathrm{n}} = \arg\min_{\mathbf{w}}\ \underbrace{\mathbf{w}^{\mathrm{H}}\,\mathbf{\Phi}_{\mathrm{n}}(k,n)\,\mathbf{w}}_{\mathbf{w}^{\mathrm{H}}\mathbf{w}} \quad \text{s. t. (9)}. \tag{10}$$

There exists a closed-form solution to (10) [1] that allows a fast computation of $\mathbf{w}_{\mathrm{n}}$. It should be noted that this filter does not necessarily provide the largest directivity index (DI).

The second spatial filter is known as the robust superdirective (SD) beamformer and minimizes the diffuse sound power at the filter's output (i. e., maximizes the DI) with a lower-bound on the WNG [22]. The lower-bound on the WNG increases the robustness to errors in the propagation vector and limits the amplification of the self-noise [22]. The optimal weight vector that minimizes the MSE between (7) and (8) subject to (9) and satisfies the lower-bound on the WNG is then obtained by

$$\mathbf{w}_{\mathrm{d}} = \arg\min_{\mathbf{w}}\ \underbrace{\mathbf{w}^{\mathrm{H}}\,\mathbf{\Phi}_{\mathrm{d}}(k,n)\,\mathbf{w}}_{\mathbf{w}^{\mathrm{H}}\,\mathbf{\Gamma}_{\mathrm{d}}(k,n)\,\mathbf{w}} \quad \text{s. t. (9)} \tag{11}$$

and subject to a quadratic constraint $\mathbf{w}^{\mathrm{H}}\,\mathbf{w} < \beta$. The parameter β^{-1} defines the minimum WNG and determines the achievable DI of the filter. In practice, it is often difficult to find an optimal trade-off between a sufficient WNG in low signal-to-noise ratio (SNR) situations, and a sufficiently high DI in high SNR situations. Moreover, solving (11) leads to a non-convex optimization problem due to the quadratic constraint, which is time-consuming to solve. This is especially problematic in our application, since the complex weight vector needs to be recomputed for each k and n due to the time-varying constraints (9).

3.2. Proposed Spatial Filters

The proposed spatial filter combines the benefits of the spatial filters in the previous subsection, i. e., providing a high DI in situations with high DNR, and a high WNG otherwise. The spatial filter is only linearly constrained, which allows a fast computation of the weights.

The optimal weights $\mathbf{w}(k,n)$ to solve our problem in (8) are found by minimizing the sum of the self-noise power and diffuse sound power at the filter's output, i. e.,

$$\mathbf{w}_{\text{nd}} = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\text{H}} \ [\boldsymbol{\Phi}_{\text{d}}(k,n) + \boldsymbol{\Phi}_{\text{n}}(k,n)] \ \mathbf{w} \quad \text{s. t. (9)}. \tag{12}$$

Using (3) and (4), the optimization problem can be expressed as

$$\mathbf{w}_{\text{nd}} = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\text{H}} \ \underbrace{[\Psi(k,n)\, \boldsymbol{\Gamma}_{\text{d}}(k) + \mathbf{I}]}_{=\mathbf{J}(k,n)} \mathbf{w} \quad \text{s. t. (9)}, \tag{13}$$

where

$$\Psi(k,n) = \frac{\phi_{\text{d}}(k,n)}{\phi_{\text{n}}(k,n)} \tag{14}$$

is the time-varying input DNR at the array microphones. The solution to (13) given the constraints (9) is [23]

$$\mathbf{w}_{\text{nd}} = \mathbf{J}^{-1}\mathbf{A}\left[\mathbf{A}^{\text{H}}\mathbf{J}^{-1}\mathbf{A}\right]^{-1}\mathbf{g}, \tag{15}$$

where $\mathbf{A}(k,n) = [\mathbf{a}(k,\varphi_1) \ \ldots \ \mathbf{a}(k,\varphi_L)]$ contains the propagation vectors for the L plane waves. The corresponding gains are given by $\mathbf{g}(k,n) = [G(k,\varphi_1) \ \ldots \ G(k,\varphi_L)]^{\text{T}}$. The estimation of $\Psi(k,n)$ is discussed in the next section.

4. PARAMETER ESTIMATION

Several parameters need to be estimated for the proposed approach in Sec. 3.2. The DOAs $\varphi_l(k,n)$ of the L plane waves can be obtained with well-known narrowband DOA estimators such as ESPRIT [24] or root MUSIC [25]. In the following, we discuss the estimation of the input DNR $\Psi(k,n)$.

To estimate $\Psi(k,n)$, we propose to use an additional spatial filter which cancels the L plane waves such that only diffuse sound is captured. The weights of this spatial filter are found by maximizing the WNG of the array, i. e.,

$$\mathbf{w}_{\Psi} = \arg\min_{\mathbf{w}} \ \mathbf{w}^{\text{H}}\mathbf{w} \tag{16}$$

subject to

$$\mathbf{w}^{\text{H}}\,\mathbf{a}(k,\varphi_l) = 0, \quad l \in \{1,2,\ldots,L\}, \tag{17}$$

$$\mathbf{w}^{\text{H}}\,\mathbf{a}(k,\varphi_0) = 1. \tag{18}$$

Constraint (18) ensures non-zero weights $\mathbf{w}_{\Psi}$. The propagation vector $\mathbf{a}(k,\varphi_0)$ corresponds to a specific direction $\varphi_0(k,n)$ being different from the DOAs $\varphi_l(k,n)$ of the L plane waves. In the following, we choose for $\varphi_0(k,n)$ the direction which has the largest distance to all $\varphi_l(k,n)$, i. e.,

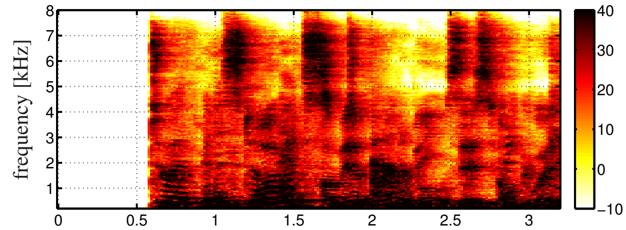
$$\varphi_0(k,n) = \arg\max_{\varphi}\big(\min_{l}|\varphi - \varphi_l(k,n)|\big), \tag{19}$$

where $\varphi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Given the weights $\mathbf{w}_{\Psi}$, the output power of the additional spatial filter is given by

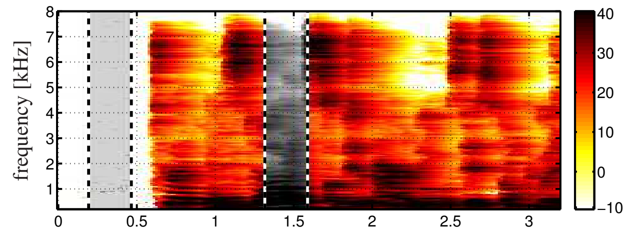
$$\mathbf{w}_{\Psi}^{\text{H}}\,\boldsymbol{\Phi}(k,n)\,\mathbf{w}_{\Psi} = \phi_{\text{d}}(k,n)\,\mathbf{w}_{\Psi}^{\text{H}}\,\boldsymbol{\Gamma}_{\text{d}}(k)\,\mathbf{w}_{\Psi} + \phi_{\text{n}}(k,n)\,\mathbf{w}_{\Psi}^{\text{H}}\,\mathbf{w}_{\Psi}. \tag{20}$$

The input DNR can now be computed with (14) and (20), i. e.,

$$\Psi(k,n) = \frac{\mathbf{w}_{\Psi}^{\text{H}}\,\boldsymbol{\Phi}(k,n)\,\mathbf{w}_{\Psi} - \phi_{\text{n}}(k,n)\,\mathbf{w}_{\Psi}^{\text{H}}\,\mathbf{w}_{\Psi}}{\phi_{\text{n}}(k,n)\,\mathbf{w}_{\Psi}^{\text{H}}\,\boldsymbol{\Gamma}_{\text{d}}(k)\,\mathbf{w}_{\Psi}}. \tag{21}$$



(a) True $\Psi(k,n)$ [dB]



(b) Estimated $\Psi(k,n)$ [dB]

Fig. 2. True and estimated DNR $\Psi(k,n)$. The two marked areas indicate respectively a silent and active part of the signal.

The required expected power of the microphone self-noise $\phi_{\text{n}}(k,n)$ can for example be estimated during silence assuming that the power is constant over time. Note that the proposed DNR estimator does not necessarily provide the lowest estimation variance in practice due to the chosen optimization criteria (16), but provides unbiased results.

5. EXPERIMENTAL RESULTS

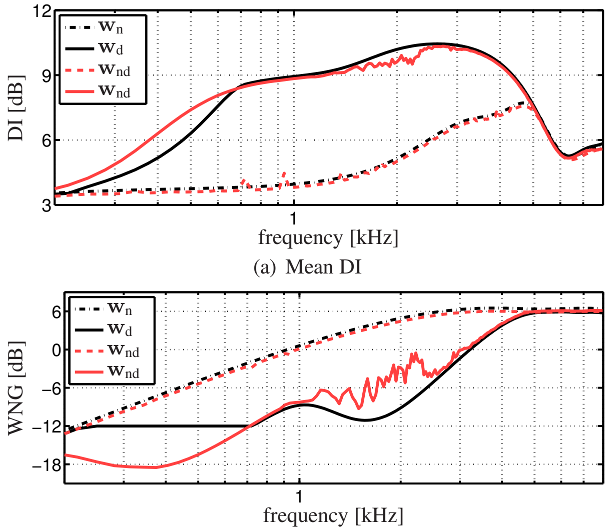
Let us assume $L=2$ plane waves in the model in (1) and an ULA with $M=4$ microphones with an inter-microphone spacing of 3 cm. A reverberant shoebox room ($7.0 \times 5.4 \times 2.4\,\text{m}^3$, $\text{RT}_{60} \approx 380\,\text{ms}$) was simulated using the source-image method [26, 27] with two speech sources at $\varphi_{\text{A}} = 86°$ and $\varphi_{\text{B}} = 11°$, respectively (distance $1.75\,\text{m}$, cf. Fig. 1). The signals consisted of $0.6\,\text{s}$ silence followed by double talk. White Gaussian noise was added to the microphone signals resulting in a segmental signal-to-noise ratio (SSNR) of $26\,\text{dB}$. The sound was sampled at $16\,\text{kHz}$ and transformed into the time-frequency domain using a 512-point STFT with 50% overlap.

We consider the directivity function $G_1(\varphi)$ in Fig. 1, i. e., we aim at extracting source A without distortions while attenuating the power of source B by $21\,\text{dB}$. We compare the two spatial filters in Sec. 3.1 and the proposed spatial filter in Sec. 3.2. For the robust SD beamformer (11), we set the minimum WNG to $-12\,\text{dB}$. For the proposed spatial filter (13), we estimate the DNR $\Psi(k,n)$ as explained in Sec. 4. The self-noise power $\phi_{\text{n}}(k,n)$ is computed from the silent signal part at the beginning. The expectation in (2) is approximated by a recursive temporal average over $\tau = 50\,\text{ms}$.

5.1. Time-Invariant Directional Constraints

For this simulation, we assume prior knowledge about the two source positions φ_{A} and φ_{B}. In all processing steps we used $\varphi_1(k,n) = \varphi_{\text{A}}$ and $\varphi_2(k,n) = \varphi_{\text{B}}$. Therefore, the directional constraints in (9) and (17) do not vary over time.

Figure 2 shows the true and estimated DNR $\Psi(k,n)$ as a function of time and frequency. We obtain a relatively high DNR during



(a) Mean DI

(b) Mean WNG

Fig. 3. DI and WNG of the spatial filters in Sec. 3. For $\mathbf{w}_\mathrm{d}$, the minimum WNG was set to $-12\,\mathrm{dB}$ to make the spatial filter robust against the microphone self-noise.

speech activity due to the reverberant environment. The estimated DNR in Fig. 2(b) possesses a limited temporal resolution due to the incorporated temporal averaging process. Nevertheless, the $\Psi(k,n)$ estimates are sufficiently accurate as shown by the following results.

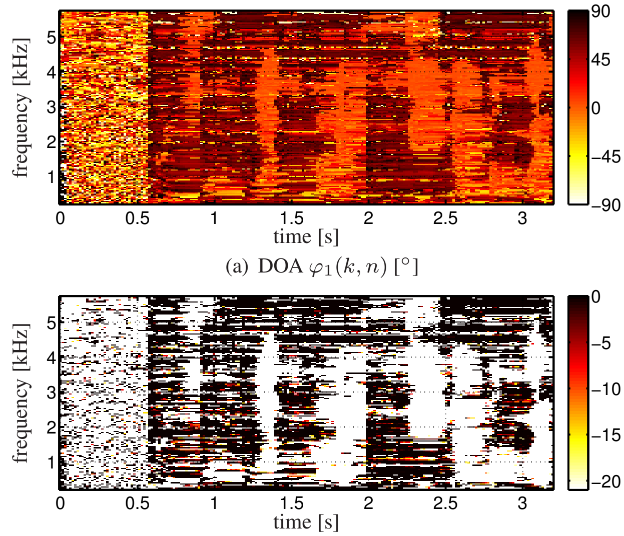
Figure 3(a) depicts the mean DI for $\mathbf{w}_\mathrm{n}$ and $\mathbf{w}_\mathrm{d}$ (which are both signal-independent), and for the proposed spatial filter $\mathbf{w}_\mathrm{nd}$ (which is signal-dependent). For the proposed spatial filter, we show the DI for a silent part of the signal and during speech activity [both signal parts marked in Fig. 2(b)]. During silence, the proposed spatial filter (dashed line $\mathbf{w}_\mathrm{nd}$) provides the same low DI as $\mathbf{w}_\mathrm{n}$. During speech activity (solid line $\mathbf{w}_\mathrm{nd}$), the obtained DI is as high as for the robust SD beamformer ($\mathbf{w}_\mathrm{d}$). Figure 3(b) shows the corresponding WNGs. During silence, the proposed spatial filter (dashed line $\mathbf{w}_\mathrm{nd}$) achieves a high WNG, while during signal activity, the WNG is relatively low.

In general, Fig. 3 shows that the proposed spatial filter combines the advantages of both existing spatial filters: during silent parts, a maximum WNG is provided leading to a minimal self-noise amplification, i. e., high robustness. During signal activity and high reverberation, where the self-noise is usually masked, a high DI is provided (at cost of a low WNG) leading to an optimal reduction of the diffuse sound. In this case, even rather small WNGs are tolerable. Note that for higher frequencies ($f > 5\,\mathrm{kHz}$), all spatial filters perform nearly identically since the coherence matrix $\mathbf{\Gamma}_\mathrm{d}(k)$ in (11) and (13) is approximately equal to an identity matrix.

5.2. Instantaneous Directional Constraints

For this simulation, we assume that no *a priori* information on φ_A and φ_B is available. The DOAs $\varphi_1(k,n)$ and $\varphi_2(k,n)$ are estimated with ESPRIT. Thus, the constraints (9) vary across time. Only for the robust SD beamformer ($\mathbf{w}_\mathrm{d}$) we use a single and time-invariant constraint (9) corresponding to a fixed look direction of $\varphi_\mathrm{A} = 86°$. This beamformer serves as a reference.

Figure 4 shows the estimated DOA $\varphi_1(k,n)$ and resulting gain $|G(k,\varphi_1)|^2$. The arriving plane wave is not attenuated if the DOA is inside the spatial window in Fig. 1 (solid line). Otherwise, the power of the wave is attenuated by 21 dB.



(a) DOA $\varphi_1(k,n)$ [°]

(b) $|G(k,\varphi_1)|^2$ [dB]

Fig. 4. Estimated DOA $\varphi_1(k,n)$ and resulting gains $G(k,\varphi_1)$

Table 1 summarizes the overall performance of the spatial filters in terms of signal-to-interference ratio (SIR), signal-to-reverberation ratio (SRR), and SSNR at the filter's output. In terms of SIR and SRR (source separation, dereverberation), the proposed approach ($\mathbf{w}_\mathrm{nd}$) and the robust SD beamformer ($\mathbf{w}_\mathrm{d}$) provide the highest performance. However, the SSNR of the proposed $\mathbf{w}_\mathrm{nd}$ is 6 dB higher than the SSNR of $\mathbf{w}_\mathrm{d}$, which represented a clearly audible benefit. The best performance in terms of SSNR is obtained using $\mathbf{w}_\mathrm{n}$. In terms of PESQ, $\mathbf{w}_\mathrm{nd}$ and $\mathbf{w}_\mathrm{d}$ outperform $\mathbf{w}_\mathrm{n}$. Using instantaneous directional constraints (as in this section) instead of time-invariant constrains (as in Sec. 5.1, values in brackets) mainly reduced the achievable SIR, but provides a fast adaption in case of varying source positions. Note that the computation time of all required complex weights per time frame was larger than 80 s for $\mathbf{w}_\mathrm{d}$ (CVX toolbox [28, 29]) and smaller than 0.08 s for the proposed approach (MATLAB R2012b, MacBook Pro 2008).

6. CONCLUSIONS

An informed linearly constrained minimum variance filter was proposed that provides a desired spatial response for L sources being simultaneously active for each time and frequency in a noisy and reverberant environment. The filter exploits instantaneous information on the direction-of-arrival of L plane waves and on the diffuse-to-noise ratio (DNR) at the filter input. The DNR information allows us to design a filter that maximizes the white noise gain when the DNR is low, and the directivity index when the DNR is high. Simulations results demonstrate the practical applicability of the proposed filter and DNR estimator.

	SIR [dB]		SRR [dB]		SSNR [dB]		PESQ	
$*$	11	(11)	−7	(−7)	26	(26)	1.5	(1.5)
$\mathbf{w}_\mathrm{n}$	21	(32)	−2	(−3)	**33**	**(31)**	2.0	(1.7)
$\mathbf{w}_\mathrm{d}$	**26**	**(35)**	0	(−1)	22	(24)	**2.1**	**(2.0)**
$\mathbf{w}_\mathrm{nd}$	25	**(35)**	**1**	(−1)	28	(26)	**2.1**	**(2.0)**

Table 1. Performance of all spatial filters [$*$ unprocessed]. Values in brackets refer to Sec. 5.1, otherwise Sec. 5.2. The signals were A-weighted before computing the SIR, SRR, and SSNR.

7. REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. Ray Liu, Eds. Wiley, 2008, ch. 9.

[3] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2008, ch. 47.

[4] J. Benesty, J. Chen, and E. A. P. Habets, *Speech Enhancement in the STFT Domain*, ser. SpringerBriefs in Electrical and Computer Engineering. Springer-Verlag, 2011.

[5] I. Tashev, M. Seltzer, and A. Acero, "Microphone array for headset with spatial noise suppressor," in *Proc. Ninth International Workshop on Acoustic, Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, 2005.

[6] M. Kallinger, G. Del Galdo, F. Kuech, D. Mahne, and R. Schultz-Amling, "Spatial filtering using directional audio coding parameters," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2009, pp. 217–220.

[7] M. Kallinger, G. Del Galdo, F. Kuech, and O. Thiergart, "Dereverberation in the spatial audio coding domain," in *Audio Engineering Society Convention 130*, London UK, May 2011.

[8] G. Del Galdo, O. Thiergart, T. Weller, and E. A. P. Habets, "Generating virtual microphone signals using geometrical information gathered by distributed arrays," in *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, United Kingdom, May 2011.

[9] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.

[10] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[11] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[12] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing: Applications to real-world problems*, ser. Signals and Communication Technology, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2003, ch. 6, pp. 155–194.

[13] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1420–1434, Sep. 2009.

[14] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 1, pp. 206–219, Jan. 2011.

[15] E. A. P. Habets and J. Benesty, "Joint dereverberation and noise reduction using a two-stage beamforming approach," in *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, 2011, pp. 191–195.

[16] M. Taseska and E. A. P. Habets, "MMSE-based blind source extraction in diffuse noise fields using a complex coherence-based a priori SAP estimator," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Sep. 2012.

[17] S. Affes, S. Gazor, and Y. Grenier, "An algorithm for multi-source beamforming and multitarget tracking," *Signal Processing, IEEE Transactions on*, vol. 44, no. 6, pp. 1512–1522, June 1996.

[18] G. Reuven, S. Gannot, and I. Cohen, "Dual source transfer-function generalized sidelobe canceller," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 4, pp. 711–727, May 2008.

[19] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[20] O. Thiergart and E. A. P. Habets, "Sound field model violations in parametric spatial sound processing," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Sep. 2012.

[21] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson Jr., "Measurement of correlation coefficients in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, 1955.

[22] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.

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

[24] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, July 1989.

[25] B. Rao and K. Hari, "Performance analysis of root-music*," in *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*, vol. 2, 1988, pp. 578–582.

[26] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[27] E. A. P. Habets. (2008, May) Room impulse response (RIR) generator. [Online]. Available: http://home.tiscali.nl/ehabets/rirgenerator.html

[28] I. CVX Research, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, September 2012.

[29] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110.