A BAYESIAN APPROACH TO SPATIAL FILTERING AND DIFFUSE POWER ESTIMATION FOR JOINT DEREVERBERATION AND NOISE REDUCTION

Soumitro Chakrabarty, Oliver Thiergart, and Emanuël A. P. Habets

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

ABSTRACT

A spatial filter, with L linear constraints that are based on instantaneous narrowband direction-of-arrival (DOA) estimates, was recently proposed to obtain a desired spatial response for at most L sound sources. In noisy and reverberant environments, it becomes difficult to get reliable instantaneous DOA estimates and hence obtain the desired spatial response. In this work, we develop a Bayesian approach to spatial filtering that is more robust to DOA estimation errors. The resulting filter is a weighted sum of spatial filters pointed at a discrete set of DOAs, with the relative contribution of each filter determined by the posterior distribution of the discrete DOAs given the microphone signals. In addition, the proposed spatial filter is able to reduce both reverberation and noise. In this work, the required diffuse sound power is estimated using the posterior distribution of the discrete set of DOAs. Simulation results demonstrate the ability of the proposed filter to achieve strong suppression of the undesired signal components with small amount of signal distortion, in noisy and reverberant conditions.

***Index Terms*—** microphone array processing, bayesian beamforming, dereverberation

1. INTRODUCTION

In modern communication systems, signal extraction in noisy and reverberant environments plays an important role. Various spatial filtering techniques have been proposed in the past decades to accomplish this task. Existing spatial filters can be broadly classified into classical spatial filters [1–4] and parametric spatial filters [5–7]. Both classes of spatial filters often require information regarding direction-of-arrival (DOA) of the sound source(s). Classical spatial filters generally require this information for estimating the propagation vectors (cf. [1]) or use the information to determine the time-frequency bins where the desired source(s) are active (cf. [4]). In parametric spatial filters, instantaneous information regarding the source DOAs is used to compute a spatial filter that provides an arbitrary desired spatial response [6–8].

In most practical cases, the DOA is often unknown and needs to be estimated. DOA estimation methods such as MUSIC [9] and ESPRIT [10] are often used for this task. These methods yield good estimates when the acoustic condition is favorable, however their performance degrades severely in noisy and reverberant conditions [11]. The aforementioned spatial filters are known to be sensitive to such estimation errors, which generally leads to signal distortions and degradation of the filter performance.

Many adaptive beamforming algorithms have been developed for increasing robustness to look direction uncertainty. Some popular approaches are diagonal loading based approach [12] and constrained minimum variance beamforming [13–16]. In these methods, robustness is achieved at the cost of reduction in noise and interference suppression. Statistical approaches have also been proposed to tackle this problem [17–20]. One particular approach of interest is Bayesian beamforming [19, 20], which focuses on DOA uncertainty and models the DOA as a discrete random variable with a prior probability density function (pdf) over a candidate set of DOAs.

Recently in [6], a spatial filter, with L linear constraints which are based on instantaneous DOA estimates, was proposed to capture at most L sound sources with a desired, arbitrary spatial response at each time-frequency instant. However, in noisy and reverberant environments, it becomes difficult to get reliable instantaneous DOA estimates. In this paper, we reformulate the approach presented in [6] and develop a Bayesian approach based on [19, 20], that provides robustness against DOA estimation errors. In contrast to [19, 20], we aim to capture sound sources with an arbitrary spatial response rather than extract a single source. The resulting Bayesian filter is a weighted sum of spatial filters pointed at a discrete set of DOAs, where the relative contribution of each spatial filter is determined according to the posterior pdf of the discrete DOAs given the microphone signals. The individual spatial filters are designed to suppress the diffuse sound and microphone self-noise while capturing the direct sound(s) with a desired spatial response, as presented in [6]. Furthermore, a probabilistic approach to diffuse sound power estimation is presented that improves upon the estimator presented in [6] and gives a sufficiently accurate estimate to achieve joint dereverberation and noise reduction. The presented simulation results demonstrate the improvement we achieve over the filter presented in [6] and also shows the ability of the proposed filter to achieve strong suppression of the undesired signal components with small amount of signal distortion, in noisy and reverberant conditions.

2. PROBLEM FORMULATION

Let us consider a uniform linear array (ULA) of M microphones located at $\mathbf{d}_{1\ldots M}$. For each time-frequency instant we assume that the sound field is composed of at most $L < M$ plane waves propagating in an isotropic and spatially homogeneous diffuse sound field. The vector of received signals, $\mathbf{y}(n,k) = [Y(n,k,\mathbf{d}_1)\ldots Y(n,k,\mathbf{d}_M)]^T$, at time frame n and frequency bin k is given by

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

* A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS).

where $\mathbf{x}_l(n,k) = [X_l(n,k,\mathbf{d}_1) \dots X_l(n,k,\mathbf{d}_M)]^T$ contains the microphone signals corresponding to the l-th plane wave, $\mathbf{x}_d(n,k)$ denotes the diffuse sound field, which models the reverberation, and $\mathbf{x}_n(n,k)$ is the spatially uncorrelated and stationary microphone self-noise. The sound pressure corresponding to the l-th plane wave, i.e., the directional sound $\mathbf{x}_l(n,k)$ is given by

$$\mathbf{x}_l(n,k) = \mathbf{a}(\theta_l,k) X_l(n,k,\mathbf{d}_1), \tag{2}$$

where $\theta_l(n,k)$ is the DOA of the l-th plane wave ($\theta = 90$ denotes the array broadside). For a ULA with omnidirectional microphones, the m-th element of the steering vector $\mathbf{a}(\theta_l,k)$ is given by

$$a_m(\theta_l,k) = \exp\{-j\kappa r_m \cos\theta_l(n,k)\}, \tag{3}$$

where r_m is the distance between the first and the m-th microphone, and κ denotes the wavenumber.

Assuming the three components in (1) to be mutually uncorrelated, the power spectral density (PSD) matrix of the microphone signals can be expressed as

$$\boldsymbol{\Phi}_{\mathbf{y}}(n,k) = \mathrm{E}\{\mathbf{y}(n,k)\mathbf{y}^{\mathrm{H}}(n,k)\} \tag{4}$$

$$= \boldsymbol{\Phi}_{\mathbf{x}}(n,k) + \underbrace{\boldsymbol{\Phi}_{\mathrm{d}}(n,k) + \boldsymbol{\Phi}_{\mathrm{n}}(n,k)}_{\boldsymbol{\Phi}_{\mathrm{u}}(n,k)}, \tag{5}$$

with

$$\boldsymbol{\Phi}_{\mathbf{x}}(n,k) = \sum_{l=1}^{L} \phi_l(n,k)\mathbf{a}^{\mathrm{H}}(\theta_l,k)\mathbf{a}(\theta_l,k), \tag{6}$$

$$\boldsymbol{\Phi}_{\mathrm{d}}(n,k) = \phi_{\mathrm{d}}(n,k)\,\boldsymbol{\Gamma}_{\mathrm{d}}(k), \tag{7}$$

$$\boldsymbol{\Phi}_{\mathrm{n}}(n,k) = \phi_{\mathrm{n}}(n,k)\,\mathbf{I}, \tag{8}$$

where $\mathbf{I}$ is an identity matrix, $\phi_{\mathrm{n}}(n,k)$ denotes the expected self-noise power, which is identical for all microphones, $\phi_{\mathrm{d}}(n,k)$ is the expected power of the diffuse sound field and $\phi_l(n,k)$ denotes the expected power of the l-th plane wave as received by the first microphone. The diffuse sound PSD $\boldsymbol{\Phi}_{\mathrm{d}}(n,k)$ and the microphone self-noise PSD $\boldsymbol{\Phi}_{\mathrm{n}}(n,k)$ are summed to get $\boldsymbol{\Phi}_{\mathrm{u}}(n,k)$, which denotes the PSD matrix of the undesired signals. The ij-th element of the coherence matrix $\boldsymbol{\Gamma}_{\mathrm{d}}(k)$, denoted by $\gamma_{ij}(k)$, is the diffuse field coherence between microphone i and j. In this work, we assume a spherically isotropic diffuse sound field, which gives $\gamma_{ij}(k) = \mathrm{sinc}(\kappa r_{ij})$ [21] with wavenumber κ and $r_{ij} = \|\mathbf{d}_i - \mathbf{d}_j\|_2$.

The aim of this work is to capture the directional sounds from a specific spatial region with a specific gain while attenuating the diffuse sound and microphone self-noise. The desired signal can be expressed as

$$Z(n,k) = \sum_{l=1}^{L} G(\theta_l,k) X_l(n,k,\mathbf{d}_1), \tag{9}$$

where $G(\theta_l,k)$ is a real-valued arbitrary directivity function which can be designed based on the target application. The proposed approach for estimation of the desired signal is presented in the next section. A block scheme of the complete system is depicted in Fig.1.

3. BAYESIAN SPATIAL FILTER

Given the TF dependent DOAs $\theta_1,\dots,\theta_L$, the desired signal $Z(n,k)$ can be estimated using a linearly constrained minimum variance (LCMV) filter with L constraints which are based on instantaneous narrowband DOA estimates. As the DOAs are TF dependent, a narrowband DOA estimator was employed in [6]. Any DOA estimation errors can result in an error in the estimation of $Z(n,k)$. To increase the robustness in noisy and reverberant environments in which the estimation of the DOAs is more challenging, we propose a Bayesian approach for spatial filtering.

In the following, the DOA is modeled as a discrete random variable with a prior pdf $p(\theta)$ over a candidate set $\Theta = \{\theta_1,\theta_2,\dots,\theta_I\}$, where $I \gg L$. An approximation of the desired signal in (9) is given as a weighted sum of spatial filters pointed at a discrete set of DOAs, which are combined according to the value of the posterior pdf for each look direction [19], i.e.,

$$\tilde{Z}(n,k) = \sum_{i=1}^{I} p(\theta_i|\mathbf{y}(n,k))\widehat{Z}(\theta_i,n,k), \tag{10}$$

where $p(\theta_i|\mathbf{y}(n,k))$ is the *a posteriori* pdf of the direction θ_i given the microphone signals, and $\widehat{Z}(\theta_i,n,k)$ is an estimate of the directional sound(s), with a specific gain, from the direction $\theta_i \in \Theta$. Since the spatial filtering is performed for each DOA in the discrete set Θ, the directional information corresponding to the L plane waves are not incorporated into the design of the filters. Instead, a single directional constraint, based on the value of the directivity function G corresponding to each element in the set Θ, is incorporated for the individual spatial filters. Therefore, the formulation in (10) only provides an approximation of the desired signal, given by (9). This formulation can potentially suffer from performance degradation, especially in terms of speech distortion, for applications that require high suppression of directional interferences.

The computation of the parameters in (10) is presented in the subsequent sections. In the following, the dependency of the steering vector $\mathbf{a}(\theta,n,k)$ and the PSD matrices $\boldsymbol{\Phi}(n,k)$ on n and k is omitted, wherever possible, for brevity.

3.1. Spatial filter

The conditional estimate of the desired signal $\widehat{Z}(\theta_i,n,k)$ can be given by a weighted sum of the microphone signals $\mathbf{y}(n,k)$, i.e.,

$$\widehat{Z}(\theta_i,n,k) = \mathrm{E}\{Z(n,k)|\theta_i\} = \mathbf{w}^{\mathrm{H}}(\theta_i,n,k)\mathbf{y}(n,k), \tag{11}$$

where $\mathbf{w}(\theta_i,n,k)$ is a complex weight vector of length M for a spatial filter pointed at the direction θ_i. The weights $\mathbf{w}(\theta_i,n,k)$ to compute $\widehat{Z}(\theta_i,n,k)$ can be found by minimizing the sum of the diffuse sound power and the self-noise power at the filter's output [6], i.e.,

$$\mathbf{w}(\theta_i,n,k) = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{u}}\mathbf{w} \tag{12}$$

subject to

$$\mathbf{w}^{H}(\theta_i,n,k)\mathbf{a}(\theta_i) = G(\theta_i,k). \tag{13}$$

The solution is given by

$$\mathbf{w}(\theta_i,n,k) = \frac{\boldsymbol{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(\theta_i)}{\mathbf{a}^{\mathrm{H}}(\theta_i)\boldsymbol{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(\theta_i)} G(\theta_i,k). \tag{14}$$

Substituting (14) in (11), we obtain the estimated desired signal for each look direction $\theta_i \in \Theta$.

3.2. Computation of the posterior probability density functions

Using Bayes theorem, the posterior pdf for each θ_i is given by

$$p(\theta_i|\mathbf{y}(n,k)) = \frac{p(\theta_i)p(\mathbf{y}(n,k)|\theta_i)}{\sum_{i=1}^{I} p(\theta_i)p(\mathbf{y}(n,k)|\theta_i)}, \tag{15}$$

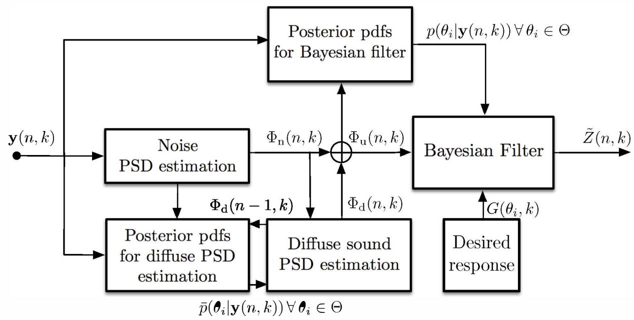

Fig. 1. The Bayesian spatial filter processing blocks.

where $p(\mathbf{y}(n,k)|\theta_i)$ is the likelihood of the observed data samples given the look direction is θ_i. The prior probability $p(\theta_i)$ is based on the available prior information regarding the source direction. In this work, we consider it to be uniform over the DOA candidate set Θ, since we assume no prior knowledge regarding the DOAs.

Assuming the microphone signals are generated from a complex Gaussian random process, the likelihood is given by

$$p(\mathbf{y}(n,k)|\theta_i) = \frac{1}{\pi^M |\mathbf{\Phi}_{\mathbf{y}}(\theta_i)|} \times \exp\left(-\mathbf{y}(n,k)^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{y}}^{-1}(\theta_i)\mathbf{y}(n,k)\right). \tag{16}$$

The determinant $|\mathbf{\Phi}_{\mathbf{y}}(\theta_i)|$ is given by

$$|\mathbf{\Phi}_{\mathbf{y}}(\theta_i)| = |\mathbf{\Phi}_{\mathrm{u}}|(1+\sigma_{\mathbf{x}}^2(\theta_i)\mathbf{a}^{\mathrm{H}}(\theta_i)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(\theta_i)), \tag{17}$$

where $\sigma_{\mathbf{x}}^2(\theta_i)$ is the power of the signal from θ_i. In this work, we estimate $\sigma_{\mathbf{x}}^2(\theta_i)$ using the minimum variance spatial spectral estimate at θ_i [22], given by

$$\sigma_{\mathbf{x}}^2(\theta_i) = \frac{1}{\mathbf{a}(\theta_i)^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{a}(\theta_i)}. \tag{18}$$

Using the matrix inversion lemma [23], we can write the inverse term $\mathbf{\Phi}_{\mathbf{y}}^{-1}(\theta_i)$ as

$$\mathbf{\Phi}_{\mathbf{y}}^{-1}(\theta_i) = \mathbf{\Phi}_{\mathrm{u}}^{-1} - \frac{\sigma_{\mathbf{x}}^2(\theta_i)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(\theta_i)\mathbf{a}^{\mathrm{H}}(\theta_i)\mathbf{\Phi}_{\mathrm{u}}^{-1}}{(1+\sigma_{\mathbf{x}}^2(\theta_i)\mathbf{a}^{\mathrm{H}}(\theta_i)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{a}(\theta_i))}. \tag{19}$$

Using the estimate of the signal power $\sigma_{\mathbf{x}}^2(\theta_i)$ from (18), we compute the determinant $|\mathbf{\Phi}_{\mathbf{y}}(\theta_i)|$ using (17), the inverse term $\mathbf{\Phi}_{\mathbf{y}}^{-1}(\theta_i)$ using (19), and substitute the values into (16) to obtain the likelihood.

For computing the weights for the individual spatial filters, presented in Section 3.1, and the posterior probabilities for each θ_i, we need to estimate the PSD matrix for the undesired signals $\mathbf{\Phi}_{\mathrm{u}}(n,k)$. Given the formulation in (7) and (8), we need to estimate the diffuse sound power $\phi_{\mathrm{d}}(n,k)$ and the microphone self-noise power $\phi_{\mathrm{n}}(n,k)$. The estimation of the diffuse sound power is discussed in the next section.

4. DIFFUSE SOUND POWER ESTIMATION

To estimate the diffuse sound power $\phi_{\mathrm{d}}(n,k)$, we use an approach similar to the diffuse-to-noise ratio (DNR) estimation method presented in [6], where an auxiliary spatial filter was used which cancels the L plane waves such that only diffuse sound is captured. The direction of the L plane waves was found using the subspace method ESPRIT [10]. Based on these estimates, the look direction of the filter was computed as the direction that has the largest distance to all the estimated DOAs. In this work, we develop a probabilistic approach to find these directions and estimate the diffuse sound power, as explained in the following.

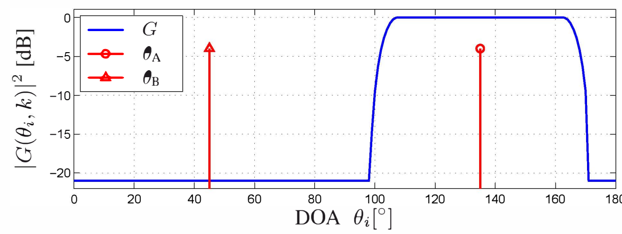

Fig. 2. Directivity function $G(\theta_i,k)$ and the source positions.

The weights of this spatial filter are found by maximizing the white noise gain (WNG) of the array, i.e.,

$$\mathbf{w}_{\mathrm{d}} = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{w} \tag{20}$$

subject to

$$\mathbf{w}^{\mathrm{H}}\mathbf{a}(\theta_{\mathrm{max}}^l) = 0, \qquad l \in \{1,\ldots L\}, \tag{21}$$

$$\mathbf{w}^{\mathrm{H}}\mathbf{a}(\theta_{\mathrm{min}}) = 1, \tag{22}$$

where the directions θ_{max}^l are the DOAs of the L plane waves, and θ_{min} denotes the direction from which we want to capture the diffuse sound. To obtain these directions, we compute an approximation of the posterior pdfs $p(\theta_i|\mathbf{y}(n,k))$, explained in Section 3.2, by replacing $\mathbf{\Phi}_{\mathrm{u}}$ with $\mathbf{\Phi}_{\mathrm{d}}(n-1)+\mathbf{\Phi}_{\mathrm{n}}$, where $\mathbf{\Phi}_{\mathrm{d}}(n-1) = \phi_{\mathrm{d}}(n-1,k)\mathbf{\Gamma}_{\mathrm{d}}(k)$. The computed approximation is denoted by $\bar{p}(\theta_i|\mathbf{y}(n,k))$.

With this approximation, θ_{min} is obtained as the direction with the lowest probability of being one of the L plane wave DOAs, i.e.,

$$\theta_{\mathrm{min}} = \arg\min_{\theta_i \in \Theta} \bar{p}(\theta_i|\mathbf{y}(n,k)). \tag{23}$$

The DOAs of the L plane waves θ_{max}^l are obtained by selecting the directions $\theta_i \in \Theta$ corresponding to the L local maximas in $\bar{p}(\theta_i|\mathbf{y}(n,k))$.

Given the filter weights $\mathbf{w}_{\mathrm{d}}$ for the auxiliary spatial filter, the output power of the filter is given by

$$\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}}\mathbf{w}_{\mathrm{d}} = \phi_{\mathrm{d}}(n,k)\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Gamma}_{\mathrm{d}}(k)\mathbf{w}_{\mathrm{d}} + \phi_{\mathrm{n}}(n,k)\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{w}_{\mathrm{d}}. \tag{24}$$

Therefore, the diffuse sound power is given by

$$\phi_{\mathrm{d}}(n,k) = \frac{\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}}\mathbf{w}_{\mathrm{d}} - \phi_{\mathrm{n}}(n,k)\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{w}_{\mathrm{d}}}{\mathbf{w}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Gamma}_{\mathrm{d}}(k)\mathbf{w}_{\mathrm{d}}}. \tag{25}$$

5. PERFORMANCE EVALUATION

In this section, we first compare the performance of the proposed approach for diffuse sound power estimation (Section 4) to the approach presented in [6]. Then, we compare the performance of the proposed Bayesian spatial filter (Section 3) with the informed LCMV filter [6], and a variant of the informed LCMV filter where the DOAs of the L plane waves are estimated by selecting the directions $\theta_i \in \Theta$ corresponding to the L local maximas in the posterior pdfs $p(\theta_i|\mathbf{y}(n,k))$. The variant of the informed LCMV filter is presented to demonstrate the advantage of incorporating the posterior pdfs over the whole set of DOAs Θ in our proposed approach. The performance of these spatial filters is evaluated for two different acoustic conditions, Scenario I and II.

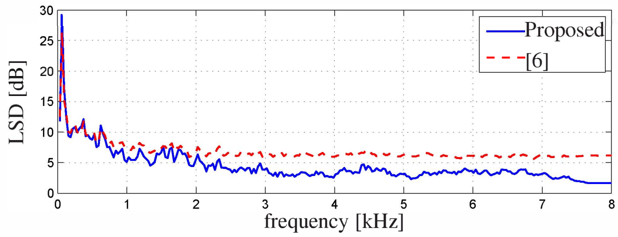

Fig. 3. Diffuse sound power estimation error w.r.t the true $\phi_{\mathrm{d}}(n,k)$, in terms of LSD, for the proposed method and the method in [6].

5.1. Setup and parameters

We assume $L = 2$ plane waves in the signal model in (1) and a ULA with $M = 5$ omnidirectional microphones where the inter-microphone distance is 3.5 cm. We simulate a reverberant shoebox room ($7.0 \times 5.4 \times 2.5$ m^3), with RT$_{60} \approx$ 260ms for Scenario I and RT$_{60} \approx$ 450ms for Scenario II, using the image-source method [24,25]. White Gaussian noise was added to the microphone signals resulting in a segmental signal-to-noise ratio (SSNR) of 27 dB and 17 dB for Scenario I and II, respectively. For both scenarios, two speech sources are placed at a distance of 1.8 m at angles $\theta_{\mathrm{A}} = 135°$ and $\theta_{\mathrm{B}} = 45°$, respectively (cf. Fig. 2). The microphone signals consist of 1 s silence followed by 2 s single talk (Source A), 2 s double talk (Source A and B), and 2 s single talk (Source B). The sampling rate was $\mathrm{F}_s = 16$ kHz and we use a 512-point short-time Fourier transform (STFT) with 50% overlap to transform the signal into the time-frequency domain. For all the spatial filters, the diffuse sound power $\phi_{\mathrm{d}}(n,k)$ is estimated using the proposed method in Section 4. The microphone self-noise power $\phi_{\mathrm{n}}(n,k)$ is computed from the silent section at the beginning. The expectation in (4) is approximated by a recursive temporal average over $\tau = 50$ms. For the proposed Bayesian filter, we consider $I = 37$ discrete DOAs in the set Θ, i.e., we consider a resolution of 5 degrees over the whole range $\theta \in [0°, 180°]$. We consider the directivity function $G(\theta_i, k)$ given in Fig. 2, i.e., we aim to capture source A with no distortions while attenuating the power of source B by 21 dB.

5.2. Diffuse sound power estimation performance

The proposed approach for diffuse sound power estimation is compared with the approach presented in [6] for the more reveberant environment in Scenario II. Fig. 3 shows the estimation error with respect to the true $\phi_{\mathrm{d}}(n,k)$ for both the methods, in terms of the log spectral distance (LSD) across frequencies [26]. Since both the filters are based on canceling the L plane waves and capturing only diffuse sound, they have higher estimation errors at lower frequencies. The proposed probabilistic approach obtains an overall lower estimation error. Due to the incorporated temporal averaging process, the proposed method has a limited temporal resolution. Nevertheless, the estimated $\phi_{\mathrm{d}}(n,k)$ is sufficiently accurate as shown by the following results.

5.3. Spatial filter performance

Tables 1 and 2 summarize the performance of all the spatial filters for Scenario I and II, respectively, in terms of signal-to-interference ratio (SIR), signal-to-reverberation ratio (SRR), SSNR, PESQ and the mean log spectral distortion (mLSD) at the filter's output. The values are computed over the more difficult double talk part. The mLSD is a measure for characterizing the distortion of the desired signal by the filter and is computed as shown in [7]. For a detailed evaluation, we also present the performance of the informed LCMV filters for the single wave assumption, i.e., $L = 1$. In the following, and in Tables 1 and 2, the proposed filter is denoted by $\mathbf{w}_{\text{Bayesian}}$. The informed LCMV filter proposed in [6] is denoted by $\mathbf{w}_{\text{iLCMV1}}$ and $\mathbf{w}_{\text{iLCMV2}}$ for $L = 1$ and $L = 2$, respectively. Similarly, the variant of the informed LCMV filter where the DOA of the plane waves are estimated based on local maximas in $p(\theta_i|\mathbf{y}(n,k))$ are denoted by $\mathbf{w}_{\text{MP1}}$ and $\mathbf{w}_{\text{MP2}}$.

	SIR [dB]	SRR [dB]	SSNR [dB]	PESQ	mLSD
Unprocessed	0	-6.0	27	1.88	–
$\mathbf{w}_{\text{iLCMV1}}$	**22**	$\mathbf{2.0}$	26	2.60	2.55
$\mathbf{w}_{\text{iLCMV2}}$	18	0.0	24	2.53	$\mathbf{1.22}$
$\mathbf{w}_{\text{MP1}}$	21	1.0	26	2.42	3.30
$\mathbf{w}_{\text{MP2}}$	12	-1.7	22	2.40	2.10
$\mathbf{w}_{\text{Bayesian}}$	**22**	1.5	**27**	$\mathbf{2.75}$	2.10

Table 1. Performance of all spatial filters for Scenario I.

	SIR [dB]	SRR [dB]	SSNR [dB]	PESQ	mLSD
Unprocessed	0	-9.0	17	1.78	–
$\mathbf{w}_{\text{iLCMV1}}$	18	$\mathbf{-1.6}$	16	2.30	2.83
$\mathbf{w}_{\text{iLCMV2}}$	16	-3.4	14	2.25	$\mathbf{1.56}$
$\mathbf{w}_{\text{MP1}}$	18	-2.3	16	2.20	3.50
$\mathbf{w}_{\text{MP2}}$	10	-4.3	13	2.20	2.36
$\mathbf{w}_{\text{Bayesian}}$	**20**	-1.9	**19**	$\mathbf{2.44}$	2.16

Table 2. Performance of all spatial filters for Scenario II.

For Scenario I, $\mathbf{w}_{\text{Bayesian}}$ and $\mathbf{w}_{\text{iLCMV1}}$ provide the best performance in terms of SIR (directional interference suppression), SRR (dereverberation) and SSNR (noise reduction). However, for the more adverse acoustic condition in Scenario II, $\mathbf{w}_{\text{Bayesian}}$ outperforms all other filters in terms of SIR and SSNR, while achieving an SRR comparable to $\mathbf{w}_{\text{iLCMV1}}$. In terms of mLSD (signal distortion), for both scenarios, the proposed filter $\mathbf{w}_{\text{Bayesian}}$ performs worse compared only to $\mathbf{w}_{\text{iLCMV2}}$. In terms of PESQ, all the filters improved the signal compared to the unprocessed signal, with the most improvement achieved by $\mathbf{w}_{\text{Bayesian}}$ for both scenarios. The performance of the local maxima based informed LCMV filters $\mathbf{w}_{\text{MP1}}$ and $\mathbf{w}_{\text{MP2}}$ is inferior compared to other filters.

In general, the results in Tables 1 and 2 show that, for both scenarios, $\mathbf{w}_{\text{Bayesian}}$ achieves a stronger/equally strong suppression of the undesired signal components compared to the other filters, while introducing a higher signal distortion compared only to $\mathbf{w}_{\text{iLCMV2}}$. Therefore, the proposed Bayesian spatial filter yields the best overall performance, especially for the highly reverberant and noisy acoustic condition in Scenario II.

6. CONCLUSIONS

A Bayesian approach to spatial filtering for capturing directional sound sources with a desired, arbitrary spatial response at each time-frequency instant was presented. Furthermore, a probabilistic approach to diffuse sound power estimation was also presented, that for the considered scenario provided a slightly more accurate estimate compared to a recently proposed non-probabilistic approach. Simulation results for two different acoustic conditions demonstrated the proposed filter's robustness to DOA estimation errors and its ability to achieve strong suppression of undesired signal components, while introducing small amount of distortion to the desired signal.

7. REFERENCES

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

[2] S. Gannot and I. Cohen, "Adaptive Beamforming and Postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., chapter 48. Springer-Verlag, 2007.

[3] 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.

[4] S. Araki, H. Sawada, and S. Makino, "Blind speech separation in a meeting situation with maximum SNR beamformers," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2007.

[5] 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)*, 2009.

[6] O. Thiergart and E.AP. Habets, "An informed lcmv filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 659–663.

[7] O. Thiergart, M. Taseska, and E.AP. Habets, "An informed mmse filter based on multiple instantaneous direction-of-arrival estimates," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sept 2013, pp. 1–5.

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

[9] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[10] R. Roy and T. Kailath, "ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[11] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.

[12] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[13] K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," *IEEE Trans. Antennas Propag.*, vol. 24, no. 5, pp. 662–669, Sep 1976.

[14] S. Applebaum and D. Chapman, "Adaptive arrays with main beam constraints," *IEEE Trans. Antennas Propag.*, vol. 24, no. 5, pp. 650–662, Sep 1976.

[15] B.D. Van Veen, "Minimum variance beamforming with soft response constraints," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 1964–1972, Sep 1991.

[16] 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.

[17] O. Besson, AA Monakov, and C. Chalus, "Signal waveform estimation in the presence of uncertainties about the steering vector," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2432–2440, Sept 2004.

[18] S. Malik, J. Benesty, and J. Chen, "A bayesian framework for blind adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2370–2384, May 2014.

[19] K.L. Bell, Y. Ephraim, and H.L. Van Trees, "A bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb 2000.

[20] C.J. Lam and AC. Singer, "Bayesian beamforming for doa uncertainty: Theory and implementation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4435–4445, Nov 2006.

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

[22] S. Kay, *Modern Spectral Estimation*, Prentice Hall, first edition, 1988.

[23] A. J. Laub, *Matrix Analysis for Scientists and Engineers*, p. 103, Society for Industrial and Applied Mathematics (SIAM), 2005.

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

[25] E. A. P. Habets, "Room Impulse Response (RIR) generator," May 2008.

[26] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.