A Bayesian Approach to Informed Spatial Filtering with Robustness Against DOA Estimation Errors

Soumitro Chakrabarty, *Student Member, IEEE*, and Emanuël A. P. Habets, *Senior Member, IEEE*

***Abstract*—A Bayesian approach to spatial filtering is presented, which is robust to uncertain or erroneous direction-of-arrival (DOA) information. The proposed framework aims to capture multiple sound sources at each time-frequency (TF) instant with an arbitrary direction dependent gain, while attenuating the diffuse sound and noise. For robustness, the DOA corresponding to each sound source is assumed to be a discrete random variable with a prior defined on a discrete set of candidate DOAs over the whole DOA space. With this assumption, the solution is given as a weighted sum of individual spatial filters, each corresponding to a specific combination of probable DOA values, with the weighting factors given by the joint posterior probabilities of the combination of DOA values. Assuming the whole DOA space as the support for each random variable results in redundant computations and contributes to a high computational cost. To alleviate this problem, a narrowband DOA estimate-based posterior probability approximation method is proposed, which isolates regions in the DOA space with high probability of containing the actual source DOAs to compute time adaptive supports for each random variable. Through experimental analysis, we demonstrate the robustness of the proposed framework against DOA estimation errors. Experimental evaluation with simulated and measured room impulse responses, in terms of objective performance measures, demonstrates the effectiveness of the framework to perform spatial filtering in noisy and reverberant acoustic environments.**

***Index Terms*—Spatial filtering, Bayesian beamforming, EM algorithm, DOA uncertainty, Robustness.**

I. Introduction

In hands-free communication systems, extraction of desired speech signal(s) in the presence multiple active sound sources, in noisy and reverberant environments remains a challenging task. With the advent of multiple microphones on modern devices, microphone array processing techniques have become an attractive solution to this task. One of the most common microphone array signal processing techniques is known as spatial filtering [1], that linearly combines the microphone signals to extract the desired source signal while suppressing the undesired signal components. In recent decades, a large variety of spatial filtering techniques have been proposed [2]–[7].

The salience of spatial filtering techniques lies in the utilization of spatial information within an estimation framework, to extract the desired speech signal and suppress the undesired signal components. In acoustic environments, sound source location is one of the most prevalent and important information. The relative location of a sound source with respect to a microphone array is generally given in terms of the direction-of-arrival (DOA) of the plane wave originating from that location. The DOA is used for various purposes, e.g., to estimate the steering vector for the spatial filter [1], to determine the time-frequency (TF) bins where the desired speech source is active [4], computing the second-order statistics (SOS) of the desired source(s) [7], and the SOS of noise and reverberation [6], [8]. Information regarding the sound source DOAs is often unavailable, and an imprecise knowledge of the source DOAs leads to severe degradation of spatial filtering performance.

Over the years, in the context of array signal processing in general, extensive research has been done in the development of robust adaptive beamforming methods to achieve robustness against steering vector errors [9]. There are various probable reasons for the error in steering vectors, of which, error or lack of DOA information is one of the main reasons and our main focus in this work.

In the existing literature on robust adaptive beamforming methods, this problem of imprecise DOA information has been approached as a DOA mismatch problem. Earlier approaches to this problem involved diagonal loading based approaches [10], [11], however a proper choice of the diagonal loading factor poses a major challenge for these methods. Another approach to this problem is the imposition of multiple linear constraints along with minimum variance beamforming [12]–[15]. These methods mainly aim to broaden the main beam of the beampattern to account for the uncertainty in the DOA information. However, due to the addition of the extra constraints the degrees of freedom of the beamformer are reduced which limits its ability to suppress the undesired signal components. In [16], quadratic constraints were introduced over an uncertainty set with steering vectors corresponding to DOAs within a desired uncertainty DOA range. Statistical approaches have also been proposed to tackle the general problem of steering vector uncertainty [17]–[20]. A particular statistical 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) defined over a candidate set of DOAs. In addition to the above mentioned methods, some earlier approaches also focused on estimating the DOA of the desired and interfering sources, and proceed with the estimates as the true DOAs. These approaches have been termed as direction-finding based approaches [21]–[23]. These methods aim to provide a data-driven solution when the DOA information regarding the sources is unavailable rather than tackle uncertainty in an explicit manner, however, they suffer severe performance degradation when the estimates are not reliable.

Recently, in the field of multi-microphone speech enhancement, spatial filtering methods have been developed that use narrowband DOA estimates, within a parametric sound field model, to formulate the steering vector as well as to estimate required parametric information [5]–[8]. In a recent contribution [8], a spatial filter, known as *informed* spatial filter, with L linear directional constraints based on instantaneous narrowband DOA estimates, was proposed to capture at most L sound sources with a user-defined gain at each time-frequency (TF) instant. Such an application can be termed as *directional filtering*, where rather than a desired source the emphasis lies on capturing sounds from specific directions with specific gains. The employment of a user-defined directional response to determine the direction dependent gain at each TF instant enables different applications using the informed spatial filters, whereas the incorporation of almost instantaneous parametric information in the

Soumitro Chakrabarty and Emanuël Habets are with the International Audio Laboratories Erlangen (a joint institution between the University of Erlangen-Nuremberg and Fraunhofer IIS), Germany.

Corresponding address: Soumitro Chakrabarty, International Audio Laboratories Erlangen, University of Erlangen-Nuremberg, Am Wolfsmantel 33, 91058 Erlangen, Germany. Email: soumitro.chakrabarty@audiolabs-erlangen.de.

form of narrowband DOA estimates enables quick adaptability of the system to changes in the sound scene. Despite these advantages, a major challenge to ensure that the user-defined directional response is obtained is to have reliable DOA estimates at each TF instant. While several methods to obtain the narrowband DOA estimates are available [1], most are known to suffer from severe degradation in performance in noisy and reverberant environments [24]. In a recent contribution [25], the present authors showed the influence of DOA estimation errors on the obtained directional response at the output of the informed spatial filters. It was shown that in case of DOA estimation errors, the obtained directional response at the output of the spatial filter deviates from the desired one even when the sources are far apart, thereby providing an objective motivation for the need of robustness in the informed spatial filtering framework. The robust adaptive beamforming methods generally deal with uncertainty in DOA information regarding a *single* desired source of interest which makes their employment to add robustness to the informed spatial filtering framework unfeasible, without extensive modification.

As a robust alternative to informed spatial filters, in [26], the present authors, following [19], proposed a Bayesian approach to spatial filtering that mitigates the problems that occur in the presence of DOA estimation errors by considering the DOA parameter at each TF instant to be a random variable with a defined prior over a discrete set of points in the whole DOA space. The solution was given as a weighted sum of individual spatial filters pointed at a discrete set of DOAs, where the weights were given by the posterior probabilities for each candidate DOA in the discrete set. The proposed solution in [26], though robust to DOA estimation errors, suffered from the problem of model violation, [27], since the individual spatial filters only incorporated a single directional constraint, instead of L constraints as required by the considered multi-wave signal model in the informed spatial filtering framework. Experimental evaluation showed that the proposed spatial filtering approach in [26] achieved strong suppression of the undesired components, however the model violation introduced noticeable amount of distortion to the desired signal at the filter output in multi-talk scenarios.

Extending our previous method, we develop a Bayesian approach to spatial filtering that now considers the DOA corresponding to each of the L plane waves as a discrete random variable with a defined prior over a discrete set of DOAs. The solution is subsequently given as the weighted sum of individual spatial filters that incorporate L directional constraints, instead of a single directional constraint, as proposed in [26]. The weighting factors for the individual spatial filters are now given by the joint posterior pdfs of the L random variables. At first, we present the formulation of the Bayesian framework with the whole DOA space as the support of each of the L random variables. The support for each random variable is considered to be the whole DOA space since there is no prior information available regarding the source locations. In this work, the individual spatial filters are derived using the minimum variance criterion with L directional constraints resulting in a linearly constrained minimum variance (LCMV) spatial filter. Also, a method for estimation of the joint posterior probabilities based on the microphone signals, that employs similar computations as in [26], is presented.

One key limitation of the Bayesian framework described above lies in considering the whole DOA space as the support for each of the L discrete random variables. It leads to the computation of the weights of the individual spatial filters as well as the estimation of the joint posterior probabilities for all possible combinations of discrete DOA values in the whole DOA space, resulting in a high computational cost. Therefore, to reduce the number of computations, we propose to formulate a Bayesian framework where the weighted averaging is performed only over regions in DOA space with high probability of containing the actual source DOAs. To this end, we propose a narrowband DOA estimate-based method to approximate the joint posterior probabilities that simultaneously reduces the computational complexity of the overall system.

For evaluation, we first present a robustness experiment, where it is shown that the proposed method indeed introduces robustness to the informed spatial filtering framework. Following that, a comparative evaluation of the proposed method and the informed spatial filtering framework [6] is presented, using simulated as well as measured RIRs.

The rest of the paper is organized as follows: Section II introduces the signal model and formulates the problem. Section III presents the mathematical formulation of the Bayesian framework considering the multi-wave signal model, along with the formulation of the individual spatial filters. Section IV presents the computation of the posterior probabilities using the microphone signals, as well as a rough complexity analysis of the Bayesian formulation presented thus far to motivate the need for further improvement in the proposed framework. The proposed method for the approximation of the posterior probabilities based on narrowband DOA estimates is presented in Section V. The estimation of the required parameters is presented in Section VI. The experimental evaluation of the proposed framework is presented in Section VII. Section VIII concludes the paper.

II. Signal Model and Problem Formulation

Let us consider an array of M microphones located at $\mathbf{d}_{1\dots M}$. The spatial filtering framework in this paper is developed in the TF domain. The TF domain signals are obtained by transforming the time domain microphone signals via a short-time Fourier transform (STFT). Considering a multi-wave signal model, for each TF instant (n,k), where n is the time frame index and k is the frequency index, we assume that the sound field is composed of $L(n,k) \leq M$ plane waves propagating in an isotropic and spatially homogeneous diffuse sound field. Note that the number of plane waves per TF instant, $L(n,k)$, might be smaller than the total number of sources present in the sound scene. The $M \times 1$ vector of microphone signals, $\mathbf{y}(n,k) = [Y(n,k,\mathbf{d}_1)\dots Y(n,k,\mathbf{d}_M)]^T$, is given by

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

where $\mathbf{x}_l(n,k) = [X_l(n,k,\mathbf{d}_1)\dots X_l(n,k,\mathbf{d}_M)]^T$ contains the sound components due to the l-th plane wave, $\mathbf{x}_{\mathrm{d}}(n,k)$ contains the L diffuse sound components, which models the reverberation, and $\mathbf{x}_{\mathrm{n}}(n,k)$ contains the microphone self-noise which is assumed to be spatially uncorrelated and stationary.

Without loss of generality, we consider the first microphone as a reference microphone. Then, the sound pressure corresponding to the l-th plane wave, i.e., the directional sound $\mathbf{x}_l(n,k)$ is expressed as

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

where $X_l(n,k,\mathbf{d}_1)$ denotes the signal proportional to the l-th plane wave at the 1-st microphone and $\mathbf{a}(\theta_l,k)$ is the array propagation vector corresponding to the l-th plane wave. Assuming the plane waves propagate in the horizontal plane, the propagation vector is dependent on the azimuth of the DOA of the l-th plane wave, $\theta_l(n,k)$. For a linear array with M omnidirectional microphones, the m-th element of the propagation vector $\mathbf{a}(\theta_l,k)$ can be written as

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

where j denotes the imaginary unit, r_m is the distance between the first and the m-th microphone, and $\kappa_k = 2\pi f_k/c$ denotes the

wavenumber with f_k being the frequency corresponding to the k-th frequency bin and c is the speed of sound.

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

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

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

where the matrix $\mathbf{A}(n,k) = [\mathbf{a}(\theta_1,k), \mathbf{a}(\theta_2,k), \ldots, \mathbf{a}(\theta_L,k)]$ contains the propagation vector corresponding to the L plane waves. The PSD matrix of the L plane waves is given by $\mathbf{\Phi}_{\mathbf{x}}(n,k) = \mathrm{E}\{\mathbf{x}(n,k)\mathbf{x}^{\mathrm{H}}(n,k)\}$, where $\mathbf{x}(n,k) = [X_1(n,k,\mathbf{d}_1), \ldots, X_L(n,k,\mathbf{d}_1)]$ denotes the signal vector of the L plane waves as received by the reference microphone.The PSD matrices corresponding to the diffuse sound and microphone self-noise, $\mathbf{\Phi}_{\mathrm{d}}(n,k)$ and $\mathbf{\Phi}_{\mathrm{n}}(n,k)$, are defined similarly using $\mathbf{x}_{\mathrm{d}}(n,k)$ and $\mathbf{x}_{\mathrm{n}}(n,k)$, respectively. Considering the plane waves to be mutually uncorrelated, $\mathbf{\Phi}_{\mathbf{x}}(n,k)$ is a diagonal matrix with the powers of the L plane waves on its diagonal, $\mathrm{diag}\{\mathbf{\Phi}_{\mathbf{x}}(n,k)\} = [\phi_1(n,k), \ldots, \phi_L(n,k)]$. The diffuse sound PSD matrix can be written as

$$\mathbf{\Phi}_{\mathrm{d}}(n,k) = \phi_{\mathrm{d}}(n,k)\ \mathbf{\Gamma}_{\mathrm{d}}(k), \tag{6}$$

where $\phi_{\mathrm{d}}(n,k)$ denotes the power of the diffuse sound at each TF instant. Since we assume the diffuse sound field to be homogeneous, this power is assumed to be identical for all microphones. The ij-th element of the diffuse sound coherence matrix $\mathbf{\Gamma}_{\mathrm{d}}(k)$, denoted by γ_{ij}, denotes the spatial coherence between i-th and j-th microphone in a purely diffuse sound field. In this work, we assume a spherically isotropic diffuse sound field, which gives $\gamma_{ij}(k) = \mathrm{sinc}(\kappa r_{ij})$ [28], 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

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

where $G(\theta_l,k)$ is the direction dependent gain corresponding to the l-th plane wave, whose value is determined based on a directional response function $\mathrm{g}(\theta,k)$. The function can be complex-valued and frequency-dependent. The design of the desired directional response function depends on the application. Two example directional response functions are shown in Fig.1. The function represented by the dashed line is designed to attenuate plane waves from 60° by 17 dB while capturing plane waves from 110° with unit gain. Such a response function can be designed for an application where the task is to equalize the spatial loudness of the directional sound components [29]. The solid line corresponds to a directional filtering application where the task is to capture the sound source coming from the array broadside with unit gain while attenuating all other directional components by 21 dB. For further details on the directional response function, the readers are referred to [8].

III. Bayesian Spatial Filter

In this section, we introduce the Bayesian approach to spatial filtering for estimating the desired signal given by (7). In Section III-A, we present the mathematical formulation of the solution with the Bayesian approach. The proposed approach involves estimating the desired signal as a weighted sum of directional estimates of the desired signal. In Section III-B, we present the formulation of the spatial filter that provides the directional estimates.

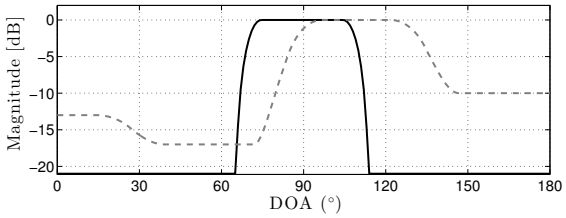


Fig. 1: Examples of directional response functions.

A. Estimate of the desired signal

An estimate of the desired signal $Z(n,k)$ can be given as a linear combination of the microphone signals $\mathbf{y}(n,k)$ at the current TF instant. This estimate, $\hat{Z}(n,k)$, can be written as

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

where $\mathbf{w}$ is a complex weight vector of length M. From the definition of the desired signal provided in (7), it can be seen that to estimate the desired signal we require spatial information in the form of the DOAs of the L plane waves. In conventional spatial filtering approaches, the source DOAs are either assumed to be known or estimated using the microphone signals. However, in noisy and reverberant conditions, it is difficult to obtain accurate DOA estimates using state-of-the-art estimators. To overcome this problem, in this work, we develop a Bayesian approach to spatial filtering that is robust against DOA estimation errors.

Let us consider the DOA of each of the L plane waves at each TF instant, $\theta_l,\ \ \forall\ \ l \in \{1,\ldots,L\}$, to be a random variable with a prior pdf $p(\theta_l)$ over the whole DOA space Θ. Please note that the TF indices on the DOAs of the L plane waves have been omitted for the sake of brevity. Following the Bayesian approach [26], the proposed spatial filter is now given as a weighted sum of individual spatial filters, with each spatial filter corresponding to a specific combination of L probable DOA values. In this work, the joint *a posteriori* pdf of the L random variables is used as the weighting factors for the individual spatial filters that provide the directional estimates of the desired signal. Then, the estimate of the desired signal can be written as

$$\hat{Z}(n,k) = \int_{\Theta}\cdots\int_{\Theta} p(\theta_1,\ldots,\theta_L|\mathbf{y}(n,k)) \\ \mathbf{w}^{\mathrm{H}}(\theta_1,\ldots,\theta_L,n,k)\mathbf{y}(n,k)\ \ \mathrm{d}\theta_1\cdots\mathrm{d}\theta_L, \tag{9}$$

where $p(\theta_1,\ldots,\theta_L|\mathbf{y}(n,k))$ is the joint posterior pdf of the L plane wave DOAs given the microphone signals $\mathbf{y}(n,k)$. It should be noted that since all the L random variables are defined over the same support Θ, we consider the joint probability for all possible combinations of the L plane wave DOAs over the whole DOA space.

For simplicity, we consider a discrete setting, where we assume that the priors $p(\theta_l)$ are defined only for a discrete set of I points $\Theta = \{\bar{\theta}_1,\ldots,\bar{\theta}_I\}$, where each element of the set is a discrete sample in the whole DOA space. With this assumption, the estimate of the desired signal becomes

$$\hat{Z}(n,k) = \underbrace{\sum_{\Theta}\cdots\sum_{\Theta}}_{L\ \text{summations}} p(\theta_1,\ldots,\theta_L|\mathbf{y}(n,k)) \\ \mathbf{w}^{\mathrm{H}}(\theta_1,\ldots,\theta_L,n,k)\mathbf{y}(n,k). \tag{10}$$

Since the computation of the posterior pdfs as well as the corresponding directional estimates is done for possible combinations of the DOAs of the L plane waves, it is worthwhile to express the estimate of the desired signal in terms of the combinations of possible DOA

values. For this, let us introduce the set of all possible combinations of DOAs, $\bar{\Theta} = \{\bar{\mathbf{\Theta}}_1, \ldots, \bar{\mathbf{\Theta}}_J\}$, where J is the total number of possible combinations of the L DOAs over the set of I possible values. Each individual element in this set is a vector of length L and can be expressed as $\bar{\mathbf{\Theta}}_j = \left[\bar{\Theta}_j(1), \ldots, \bar{\Theta}_j(L)\right]^{\mathrm{T}}$, where the individual elements of this vector correspond to one of the I possible DOA values, i.e., $\bar{\Theta}_j(l) \in \Theta, \quad \forall \quad l \in \{1, \ldots, L\}$. To clarify, θ_l denotes the random variable corresponding to the DOA of the l-th plane wave whereas $\bar{\Theta}_j(l)$ denotes a particular value taken by the random variable θ_l in the j-th combination. With this notation in place, (10) can be written as

$$\hat{Z}(n,k) = \sum_{j=1}^{J} p(\bar{\mathbf{\Theta}}_j|\mathbf{y}(n,k))\mathbf{w}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j, n, k)\mathbf{y}(n,k), \tag{11}$$

where $\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k)$ is the weight vector of the spatial filter that provides the directional estimate for the specific combination of the L DOAs, $\bar{\mathbf{\Theta}}_j$. Therefore, the complex weight vector $\mathbf{w}$ in (8) can be expressed as

$$\mathbf{w}(n,k) = \sum_{j=1}^{J} p(\bar{\mathbf{\Theta}}_j|\mathbf{y}(n,k))\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k). \tag{12}$$

With this formulation in place, we now need to compute the directional weight vectors for each combination and their corresponding joint posterior pdf.

B. Individual spatial filter weights

The weight vectors of the individual spatial filters can be computed considering some optimization criterion, such as minimum mean square error (MMSE), minimum variance distortionless response (MVDR), linearly constrained minimum variance (LCMV), etc. In this work, we use the LCMV criterion to obtain the directional weight vectors. The weights of this LCMV filter can be found by minimizing the sum of the diffuse sound power and the self-noise power at the output, i.e.,

$$\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k) = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{\Phi}_{\mathrm{u}}\mathbf{w} \tag{13}$$

subject to

$$\mathbf{a}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j(l), k)\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k) = G(\bar{\mathbf{\Theta}}_j(l), k) \ \ \forall \ \ l \in \{1, \ldots, L\}. \tag{14}$$

It should be noted that for each combination $\bar{\mathbf{\Theta}}_j$, the cost function to minimize remains the same. It is only the constraints of the directional LCMV filter that change for each combination. The solution is given by

$$\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k) = \mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)\left[\mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)\right]^{-1}\mathbf{g}. \tag{15}$$

where $\mathbf{A}(\bar{\mathbf{\Theta}}_j) = \left[\mathbf{a}(\bar{\Theta}_j(1), k), \ldots, \mathbf{a}(\bar{\Theta}_j(L), k)\right]$ contains the propagation vectors corresponding to the j-th combination of probable DOA values. The corresponding directional gains are given by $\mathbf{g} = \left[G(\bar{\Theta}_j(1), k), \ldots, G(\bar{\Theta}_j(L), k)\right]$.

Given the directional estimates corresponding to all the possible combinations for probable DOAs, it can be seen from (12) that we need to compute the corresponding *a posteriori* pdf to obtain the final estimate of the weight vector in (8).

IV. Estimation of posterior probabilities

In this section, we first present a method for estimation of the posterior probabilities that follows from the approach presented in [26]. Then, we present a rough complexity estimate of the complete proposed spatial filtering framework to motivate further improvements.

A. Microphone signal based posterior pdf estimation

Using Bayes theorem, the posterior pdf for the j-th combination can be expressed as

$$p(\bar{\mathbf{\Theta}}_j|\mathbf{y}(n,k)) = \frac{p(\bar{\mathbf{\Theta}}_j)p(\mathbf{y}(n,k)|\bar{\mathbf{\Theta}}_j)}{\sum_{j=1}^{J} p(\bar{\mathbf{\Theta}}_j)p(\mathbf{y}(n,k)|\bar{\mathbf{\Theta}}_j)}, \tag{16}$$

where $p(\mathbf{y}(n,k)|\bar{\mathbf{\Theta}}_j)$ is the likelihood of the observed data $\mathbf{y}(n,k)$ given the source DOA combination $\bar{\mathbf{\Theta}}_j$. The joint prior distribution of the source DOAs $p(\bar{\mathbf{\Theta}}_j)$ is based on the prior information of the source directions.

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

$$\begin{aligned} p(\mathbf{y}(n,k)|\bar{\mathbf{\Theta}}_j) =& \frac{1}{\pi^M|\mathbf{\Phi}_{\mathbf{y}}(\bar{\mathbf{\Theta}}_j)|} \\ & \times \exp\left(-\mathbf{y}^{\mathrm{H}}(n,k)\mathbf{\Phi}_{\mathbf{y}}^{-1}(\bar{\mathbf{\Theta}}_j)\mathbf{y}(n,k)\right). \end{aligned} \tag{17}$$

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

$$\begin{aligned} |\mathbf{\Phi}_{\mathbf{y}}(\bar{\mathbf{\Theta}}_j)| =& |\mathbf{\Phi}_{\mathrm{u}}||\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)| \\ & \times |\mathbf{\Phi}_{\mathrm{x}}^{-1}(\bar{\mathbf{\Theta}}_j) + \mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)|, \end{aligned} \tag{18}$$

where $\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)$ is the direction dependent PSD of the L plane waves for $\bar{\mathbf{\Theta}}_j$. Since we consider a multi-wave signal model, the estimate of the power of the plane waves is given by the minimum variance spatial spectral estimate [30] as

$$\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j) = \left[\mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{y}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)\right]^{-1}. \tag{19}$$

Since we consider the plane waves to be mutually uncorrelated, the matrix $\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)$ should be a diagonal matrix. However, the estimate of $\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)$ obtained using (19) is not a diagonal matrix. Therefore, the final estimate of the PSD matrix $\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)$ is obtained by setting the off-diagonal elements of the matrix computed using (19) to zero. Using the matrix inversion lemma [31], the inverse term $\mathbf{\Phi}_{\mathbf{y}}^{-1}(\bar{\mathbf{\Theta}}_j)$ is given by

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{y}}^{-1}(\bar{\mathbf{\Theta}}_j) =& \mathbf{\Phi}_{\mathrm{u}}^{-1} - \mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j) \\ & \times [\mathbf{\Phi}_{\mathrm{x}}^{-1}(\bar{\mathbf{\Theta}}_j) + \mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)] \ \ \mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{u}}^{-1} \end{aligned} \tag{20}$$

Using the estimate of the signal power $\mathbf{\Phi}_{\mathbf{x}}(\bar{\mathbf{\Theta}}_j)$ from (19), we compute the determinant $|\mathbf{\Phi}_{\mathbf{y}}(\bar{\mathbf{\Theta}}_j)|$ using (18), the inverse term $\mathbf{\Phi}_{\mathbf{y}}^{-1}(\bar{\mathbf{\Theta}}_j)$ using (20), and substitute the values into (17) to obtain the likelihood.

Thus far in our formulation, we considered the finite support for each random variable θ_l to be the discrete set of points over the *whole* DOA space Θ. Recall that initially we assumed $L \leq M$. With this assumption, we can consider the L sources to be sparse in the DOA space. Hence, computing the weighted average of the directional spatial filters over all possible DOA combinations is not efficient, and can potentially result in large number of redundant computations. Therefore, it would be more suitable to formulate a Bayesian framework where the weighted averaging is done only over isolated regions in the DOA space with high probability of containing the actual source DOAs. Note that with the formulation presented thus far, this is not possible. In the following, we provide a rough estimate of the computational complexity of the Bayesian spatial filtering framework presented so far to further motivate the need for a modified formulation.

B. Complexity analysis

The complexity analysis presented in this section is for the case of $L \leq M$. The computational complexity of the directional LCMV filters is mainly dominated by the two matrix inversions $\mathbf{\Phi}_{\mathrm{u}}^{-1}$ and $\left[\mathbf{A}^{\mathrm{H}}(\bar{\mathbf{\Theta}}_j)\mathbf{\Phi}_{\mathrm{u}}^{-1}\mathbf{A}(\bar{\mathbf{\Theta}}_j)\right]^{-1}$ [32]. However, the outer inversion is the

only one that needs to be recomputed for each DOA combination. Taking into account the computational complexity due to the matrix multiplication involved in the inversion, the computational cost for each computation of the directional weight vectors can be given as $\mathcal{O}(M^2L)$.

The complexity corresponding to the computation of the posterior probabilities is dominated by the matrix multiplications and for each computation the complexity can be roughly given as $\mathcal{O}(M^2L)$. Therefore, the total computational complexity at each TF instant can be given as $\mathcal{O}(JM^2L)$.

One of the major contributors to the high complexity is the total number of combinations J for which we need to compute the directional spatial filter weights and the corresponding posterior probabilities. The total number of combinations is given by considering each combination as an L element subset of the I element set Θ, where the elements are not ordered. The total number of such combinations can be given by

$$J = \binom{I}{L} = \frac{I!}{(I-L)!L!}. \tag{21}$$

To put the computational complexity into perspective, let us consider the simplest case of $L = 2$. To obtain an accurate estimate of the desired signal, we need $I \gg L$, therefore sampling the DOA space of a ULA, $\{0, 180\}$, with a resolution of 10 degrees, we obtain $I = 19$. Substituting these values in (21), we obtain $J = 171$, i.e., we need to go through 171 iterations of computing the direction dependent parameters at each TF instant.

The analysis presented here shows that with the present formulation of the Bayesian approach to spatial filtering, the computational complexity is too huge to make it an effecient algorithm. As stated above, the main contributor to this high computation cost is the number of combinations J. Therefore, in the next section we propose a DOA estimate-based method with lower computational complexity, to approximate the posterior probabilities, and simultaneously reduce the number of iterations J for which each spatial filter weight $\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k)$ needs to be computed. Please note that the estimation of the posterior probabilities using the microphone signals is presented for completeness. The performance of the Bayesian framework with this estimation procedure is not evaluated in this paper due to its evident limitations from the computational stand point.

V. Proposed Approximation of Posterior Probabilities

In this section, we will present the proposed narrowband DOA estimate-based method for approximation of the posterior pdfs, based on which we also aim to reduce the total number of combinations J for which the individual spatial filter weights $\mathbf{w}(\bar{\mathbf{\Theta}}_j, n, k)$, $\forall\, j \in \{1, \ldots, J\}$ need to be computed.

A. Posterior pdf approximation

Let us consider each of the L random variables θ_l to be an independent and identically distributed (i.i.d) random variable. With this assumption, the joint posterior pdf of the L random variables can be written as

$$p(\theta_1, \ldots, \theta_L|\mathbf{y}(n,k)) = \prod_{l=1}^{L} p(\theta_l|\mathbf{y}(n,k)), \tag{22}$$

It is worth noting here that following the assumption that the random variables are i.i.d, the individual posterior probabilities $p(\theta_l|\mathbf{y}(n,k))$ can be computed using the formulation presented in [26]. However, such a solution does not help in reducing the computational complexity of the overall framework. Therefore, opting for an estimate-based

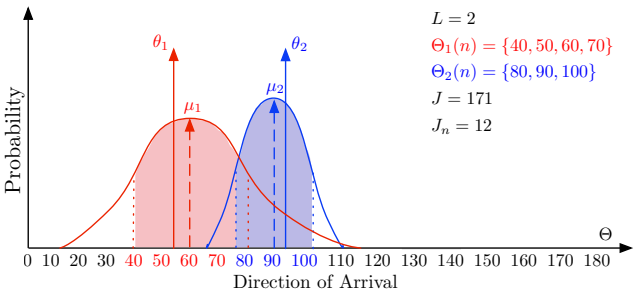


Fig. 2: Illustrative example of the Gaussian parameters based reduction of number of combinations J. θ_1 and θ_2 are the true source DOAs.

method, we propose a narrowband DOA estimate-based approximation of the posterior pdf of θ_l, given by

$$p(\theta_l|\mathbf{y}(n,k)) \approx p(\theta_l|\hat{\Theta}(n,k)), \tag{23}$$

where we now approximate the posterior pdfs based on the DOA estimates as our observations rather than the microphone signals. Since we aim to reduce the computational complexity of our spatial filtering framework, as a first step, we propose to compute the posterior pdf of each θ_l for each time frame and using the approximated values across the whole frequency range rather than computing it at each TF instant.

At each time frame, it is assumed that the fullband distribution of the narrowband DOA estimates, $\hat{\Theta}(n)$, is modeled by a Gaussian mixture (GM) with L components, given by

$$p(\hat{\Theta}(n)) = \sum_{l=1}^{L} \alpha_l \;\; \mathcal{N}(\hat{\Theta}(n); \mu_l, \sigma_l^2), \tag{24}$$

where α_l is the mixing parameter and $\mathcal{N}(\hat{\Theta}; \mu_l, \sigma_l^2)$ denotes a univariate Gaussian distribution with mean μ_l and variance σ_l^2. For our proposed approximation method, the main parameters of interest are the individual Gaussian parameters $\{\mu_l, \sigma_l^2\}\ \forall\ l \in \{1, \ldots, L\}$.

Given the L Gaussian parameters, we consider that the posterior probability distribution of θ_l is parametrized by the parameters of the l-th Gaussian component in (24). Since θ_l is a discrete random variable defined over the finite set of points $\Theta = \{\bar{\theta}_1, \ldots, \bar{\theta}_I\}$, the posterior pdf for each probable value of θ_l is then given by

$$p(\theta_l = \bar{\theta}_i|\hat{\Theta}(n)) = \frac{\mathcal{N}(\bar{\theta}_i; \mu_l, \sigma_l^2)}{\sum_{i'=1}^{I} \mathcal{N}(\bar{\theta}_{i'}; \mu_l, \sigma_l^2)}\ \forall i \in \{1, \ldots, I\}, \tag{25}$$

where

$$\mathcal{N}(\bar{\theta}_i; \mu_l, \sigma_l^2) = \frac{1}{\sigma_l\sqrt{2\pi}} \exp\left(\frac{-(\bar{\theta}_i - \mu_l)^2}{2\sigma_l^2}\right), \tag{26}$$

and the denominator term in (25) ensures the pdf sums to one. Thus, now we have posterior probabilities associated with each element in Θ for each θ_l. In the following, we describe the proposed method for reduction of J based on the posterior probabilities $p(\theta_l = \bar{\theta}_i|\hat{\Theta}(n))$.

B. Reduction of number of combinations

As mentioned in Section IV-B, the number of combinations J is dependent on the finite support of each random variable θ_l, which thus far was assumed to be the whole DOA space Θ. The main aim is to reduce this support of each θ_l from the set of points over the whole DOA space Θ to a time-dependent subset $\Theta_l(n) \subseteq \Theta$. Given the posterior pdf of θ_l at each $\bar{\theta}_i$, we select the points in the set Θ whose pdf is above a predefined threshold. This can be

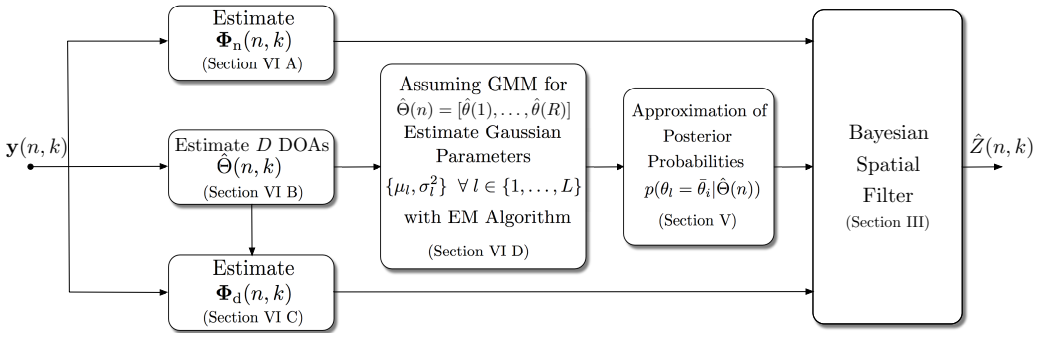


Fig. 3: Block diagram of the proposed framework.

mathematically given by

$$\Theta_l(n) = \{\bar{\theta}_i : p(\theta_l = \bar{\theta}_i|\hat{\Theta}(n)) \geq \delta\}, \tag{27}$$

where δ is the predefined threshold that is considered to be same for all L. For closely placed sources, there exists the possibility of overlap between the computed subsets $\Theta_l(n),\ \forall\, l \in \{1, \ldots, L\}$. To ensure the reduction in the number of redundant computations, we impose the restriction that the L subsets need to be disjoint, i.e.,

$$\Theta_1(n) \cap \Theta_2(n) \cap \cdots \Theta_L(n) = \emptyset. \tag{28}$$

To ensure this, if there are overlapping elements at time frame n, an overlapping element $\bar{\theta}_i$ is retained in the set $\Theta_l(n)$ for which

$$l(n) = \arg\max_{l}\ p(\theta_l = \bar{\theta}_i|\hat{\Theta}(n)). \tag{29}$$

With this time-dependent truncation of the support for each θ_l, the set of possible combinations is redefined as a time-dependent set $\bar{\Theta}(n) = \{\bar{\Theta}_1(n), \ldots, \bar{\Theta}_{J_n}(n)\}$, where the number of combinations, denoted by J_n, is now given by

$$J_n = \prod_{l=1}^{L} |\Theta_l(n)|, \tag{30}$$

where $|\cdot|$ denotes the cardinality operator, and the expression follows from the disjoint set restriction. It should be noted that once the adaptive subsets are computed, the posterior probabilities of the elements in each subset should be renormalised to ensure that the individual pdfs sum to one.

An illustrative example is presented in Fig. 2, to explain the proposed method via visualization. Sampling the DOA space of a ULA, Θ, with a resolution of 10 degrees, we get $I = 19$ discrete candidate DOAs, given on the X-axis. For the n-th time frame, we consider $L = 2$ impinging plane waves with DOAs θ_1 and θ_2. The means of the 2 Gaussian distributions representing the fullband distribution of the narrowband DOA estimates $\hat{\Theta}(n)$ are given by μ_1 and μ_2. Using (27), only the elements on the X-axis of the plot that lie within the shaded regions of the individual distributions are assigned to either of the time-dependent subsets $\Theta_1(n)$ or $\Theta_2(n)$, represented in red and blue, respectively. In the figure it can be seen that the element "80" lies within the shaded regions of both the Gaussians, however since the posterior probability of this element lying in the second Gaussian is higher, it is retained in Θ_2 and eliminated from Θ_1, thus giving the 2 disjoint subsets as shown in the figure. Using our proposed method, it can be seen that the total number of combinations reduces from $J = 171$ to $J_n = 12$.

To further clarify the redefinition of the set of combinations, similar to the original definition in Section III-A, we express the j-th element of $\bar{\Theta}(n)$ as a vector of length L, $\bar{\mathbf{\Theta}}_j(n) = \left[\bar{\mathbf{\Theta}}_j(1,n), \ldots, \bar{\mathbf{\Theta}}_j(L,n)\right]^{\mathrm{T}}$. Please recall that $\bar{\Theta}(n)$ denotes the set of all possible combinations of DOAs at time frame n. Note that now the l-th element of this vector $\bar{\mathbf{\Theta}}_j(l,n)$ corresponds to one of the possible DOA values contained in the redefined support of the l-th plane wave, i.e. $\bar{\mathbf{\Theta}}_j(l,n) \in \Theta_l(n)$, rather than the complete set of possible DOA values Θ. With the redefined notations in place, the weight vector computation given in (12) can be reformulated as

$$\mathbf{w}(n,k) = \sum_{j=1}^{J_n} p(\bar{\mathbf{\Theta}}_j(n)|\hat{\Theta}(n))\mathbf{w}(\bar{\mathbf{\Theta}}_j(n), n, k), \tag{31}$$

where the number of elements within the summation are now time-dependent. With this modified formulation of the framework finally in place, the only remaining computations are the estimation of the required parameters in the framework, which is presented in the following.

VI. Parameter Estimation

In the presented framework, several distinct parameters need to be computed, namely, noise PSD matrix $\mathbf{\Phi}_{\mathrm{n}}(n,k)$, diffuse sound PSD matrix $\mathbf{\Phi}_{\mathrm{d}}(n,k)$, narrowband estimates of the DOAs, and the L Gaussian mixture parameters $\{\mu_l, \sigma_l^2\}\ \forall\ l \in \{1, \ldots, L\}$. These parameters, except the Gaussian mixture parameters, are computed for each TF instant. The Gaussian mixture parameters are estimated for each time frame. In this work, we assume that L is known and fixed for all TF bins. In this section, we explain the computation of each of the rest of the parameters.

A. Estimation of noise PSD $\mathbf{\Phi}_n(n,k)$

In this work, we consider the noise statistics to be stationary. Considering this, the noise PSD matrix $\mathbf{\Phi}_{\mathrm{n}}(n,k)$ is estimated from the time frames where the speech sources are silent and no diffuse sound component is present. There also exist several other methods for noise PSD estimation in literature that can be employed within the proposed framework. For further details regarding such methods, the reader is referred to [33]–[35] and the references therein.

B. DOA estimation

For the proposed estimate-based posterior probability approximation method, we need to estimate the narrowband DOAs, $\hat{\Theta}(n,k)$ at each TF bin.

The narrowband DOA estimates can be obtained with subspace based narrowband DOA estimators such as ESPRIT [36] or MUSIC [37]. In this work, we choose to use ESPRIT due to its computational efficiency. ESPRIT requires the microphone array to posses displacement invariance i.e., there should be matched pairs of microphones with identical displacement vectors. This condition is satisfied by a ULA, which is used for the experimental evaluations presented

in Section VII. The DOA estimator requires the microphone signal PSD $\mathbf{\Phi}_{\mathbf{y}}(n,k)$, defined in (4), as input. In this work, $\mathbf{\Phi}_{\mathbf{y}}(n,k)$ is estimated by approximating the expectation operation in (4) by recursive temporal averaging,

$$\hat{\mathbf{\Phi}}_{\mathbf{y}}(n,k) = \alpha_t \mathbf{y}(n,k)\mathbf{y}^{\mathrm{H}}(n,k) + (1-\alpha_t)\hat{\mathbf{\Phi}}_{\mathbf{y}}(n-1,k), \quad (32)$$

where α_t is the temporal smoothing factor. With the narrowband DOA estimates at each TF instant, for each time frame we obtain $R = KD$ DOA estimates, where K is the total number of frequency bins, and D denotes the number of DOA estimates obtained per TF bin. This forms the data set to which the L Gaussians are fitted, as explained in Section VI-D.

C. Estimation of diffuse sound PSD $\mathbf{\Phi}_d(n,k)$

For the computation of the diffuse sound power ϕ_d, which is used to compute the diffuse sound PSD matrix $\mathbf{\Phi}_{\mathrm{d}}(n,k)$ with (6), in this work, we directly employ the method proposed in [6], where an auxiliary spatial filter is employed to estimate the diffuse sound power. The auxiliary filter aims to maximize the white noise gain (WNG) of the array while canceling out the L plane waves corresponding to the direct sound sources by pointing a null towards the L estimated source DOAs at each TF instant. For further details, we refer the reader to [6].

D. Estimation of Gaussian mixture parameters

In the following, we assume the fullband distribution of the DOA estimates to be a mixture of L univariate Gaussians. As explained in Section V, for the DOA estimate-based approximation of the posterior probabilities, we need to estimate the parameters of the L Gaussians. In this work, we use the standard maximum-likelihood based expectation maximization (EM) algorithm to estimate the Gaussian parameters. The estimation of Gaussian parameters is done at each time frame n, however for clarity, in the following we omit the time index.

Since the number of Gaussians, L, is known, we initialize the means of the Gaussians with the K-means algorithm [38]. Alternatively, an adaptive K-means algorithm could be used if the number of Gaussians is unknown. After the initialization, we propagate through the E- and M-step of the EM algorithm, which are repeated until convergence.

Given the set of R narrowband DOA estimates at each time frame $\hat{\Theta} = [\hat{\theta}(1),\ldots,\hat{\theta}(R)]$, the Gaussian parameters $\mathcal{G} = \{\alpha_l, \mu_l, \sigma_l^2\}\ \forall\ l \in \{1,\ldots,L\}$ can be found by maximizing the likelihood function,

$$p(\hat{\Theta}|\mathcal{G}) = \prod_{r=1}^{R}\sum_{l=1}^{L} \alpha_l\ \mathcal{N}(\hat{\theta}(r);\mu_l,\sigma_l^2), \quad (33)$$

which, in EM algorithm, is done by alternating between the E- and M-step until convergence . In the E-step of the algorithm, using the current parameter estimates, the membership probabilities of each data point $\hat{\theta}(n,r)$ is computed using

$$\Lambda_l^r = \frac{\alpha_l\ \mathcal{N}(\hat{\theta}(r);\mu_l,\sigma_l^2)}{\sum_{l'=1}^{L}\alpha_{l'}\ \mathcal{N}(\hat{\theta}(r);\mu_{l'},\sigma_{l'}^2)}. \quad (34)$$

In the M-step, the mixture parameters are updated as

$$\mu_l \leftarrow \frac{\sum_{r=1}^{R}\Lambda_l^r\hat{\theta}(r)}{\sum_{r=1}^{R}\Lambda_l^r}, \quad (35)$$

$$\sigma_l^2 \leftarrow \frac{\sum_{r=1}^{R}\Lambda_l^r(\hat{\theta}(r)-\mu_l)(\hat{\theta}(r)-\mu_l)^{\mathrm{T}}}{\sum_{r=1}^{R}\Lambda_l^r}, \quad (36)$$

$$\alpha_l \leftarrow \frac{1}{R}\sum_{r=1}^{R}\Lambda_l^r. \quad (37)$$

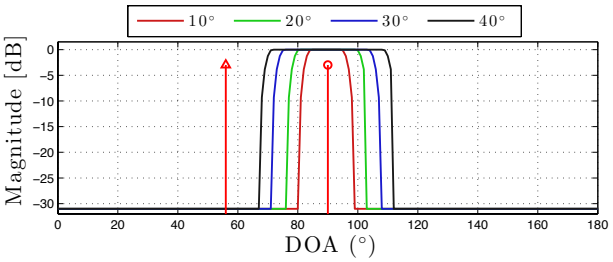


Fig. 4: Directional response function $\mathrm{g}(\theta,k)$, with varying widths, considered in the experimental analysis presented in Section VII-A. The markers represent the direction of Source A (circle) and Source B (triangle). The different window widths are plotted with different colors.

As mentioned earlier in Section V, in our framework, the individual Gaussian parameters $\{\mu_l, \sigma_l^2\}$ are used to approximate the posterior probabilities. The computational blocks associated with the proposed framework are shown in Fig. 3.

VII. Experiments and Performance Evaluation

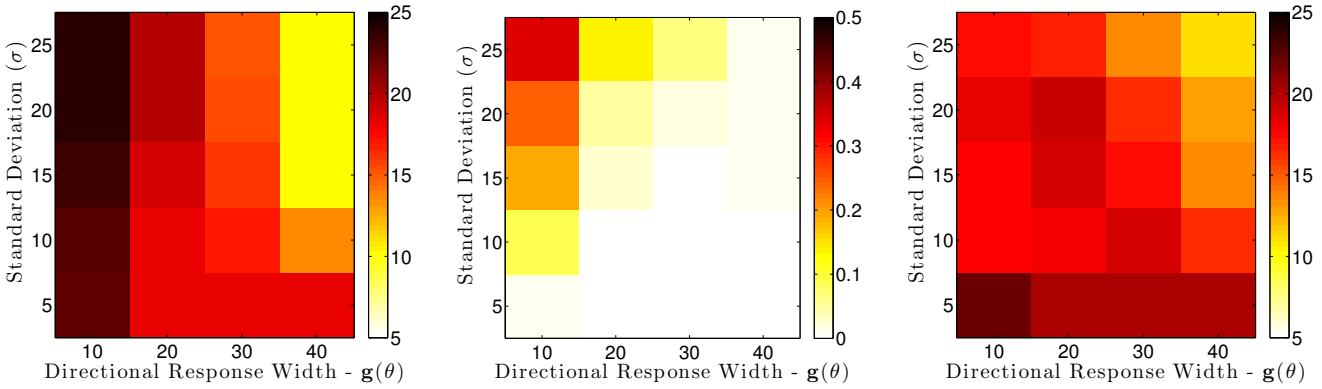
The performance of the proposed method was evaluated using both simulated and measured data. For all the experimental evaluations presented in this section, we consider the application of directional filtering where we define a desired directional response function that aims to capture direct sound source(s) from a specific region without distortion, while attenuating the direct sound sources from all other angular regions by a specific level. In the following, sound source(s) lying in the directional region with unit gain are referred as the desired source(s) whereas any sound source(s) outside this region are termed as interfering sources(s). The number of plane waves for all the experiments is the same, i.e., $L = 2$. First, an experimental analysis to demonstrate the robustness of the proposed framework in comparison to the informed LCMV filter [6] is presented in Section VII-A. The performance of the proposed framework is evaluated under different acoustic conditions using simulated room impulse responses (RIRs) in Section VII-B and using measured RIRs in Section VII-C.

A. Robustness Experiment

We first present an experimental analysis where the DOA estimation errors are introduced in a controlled manner.

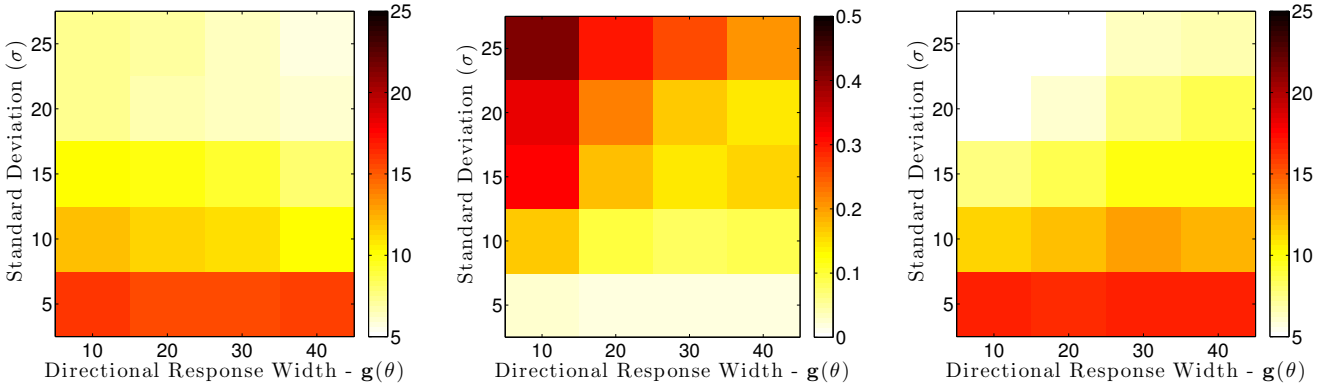
1) Experimental setup: For this experiment, we consider a ULA with $M = 4$ elements with an inter-element spacing of 3 cm. Two speech sources, Source A and B, were placed in front of the microphone array. The desired source, Source A, was positioned at the array broadside, $\theta_{\mathrm{A}} = 90^{\circ}$, with the interfering speaker, Source B positioned at $\theta_{\mathrm{B}} = 56^{\circ}$. The input signal of 6 s duration consists of two simultaneously active speakers, with a sampling rate of $F_s = 16$ kHz. A 512 point short-time Fourier transform (STFT) with 50% overlap is used to transform the signals into the TF domain. The acoustic environment was considered to be anechoic, with spatially white noise added to the speech signals resulting in an input segmental signal to noise ratio (segSNR) of 24 dB.

In this experiment, we model the DOAs at each TF instant, by a zero-mean Gaussian process with a standard deviation σ_{DOA}, centered around the known DOAs of Source A and B. Then, the narrowband



(a) segIR (b) segSDI (c) segSIR

Fig. 5: Objective measures for the proposed framework, for the experiment described in Section VII-A.



(a) segIR (b) segSDI (c) segSIR

Fig. 6: Objective measures for Informed LCMV filter [6], for the experiment described in Section VII-A.

DOA estimates can be expressed as

$$\hat{\theta}_{\mathrm{A}}(n,k) = \theta_{\mathrm{A}} + \Delta\theta_{\mathrm{A}}(n,k),$$
$$\hat{\theta}_{\mathrm{B}}(n,k) = \theta_{\mathrm{B}} + \Delta\theta_{\mathrm{B}}(n,k), \tag{38}$$

where $\Delta\theta_{\mathrm{A}}(n,k)$ and $\Delta\theta_{\mathrm{B}}(n,k)$ are the absolute DOA errors corresponding to Source A and B, respectively, with $\sigma^2_{\mathrm{DOA}} = \mathrm{E}\{\Delta\theta^2\}$ as the error variance. For the proposed method, the DOA space was sampled with a resolution of 10°, resulting in $I = 19$ discrete points. The performance of the proposed framework was compared to the informed LCMV filter for different standard deviation values of the introduced error and varying widths of the directional response window. The choice of the directional response function(s) for this experiment is shown in Fig. 4, where it can be seen that the aim is to acquire Source A with unit gain while suppressing Source B by 31 dB.

For both the proposed framework as well as the informed LCMV (iLCMV) filter, the DOA estimation errors directly affect the ability of the spatial filters to suppress the interfering speaker to the desired level as well as its ability to acquire the desired speaker without distortion. Therefore, to evaluate the robustness of the spatial filters, we use the following three objective performance measures: i) segmental interference reduction (segIR), which signifies the amount of interference suppression achieved by the spatial filter, ii) segmental signal distortion index (segSDI), which indicates the distortion of the desired signal, and iii) segmental signal-to-interference ratio (segSIR) at the output of the spatial filter.

2) Results: The results for this experiment are presented as 2D plots in Fig. 5 and 6, for the proposed method and the iLCMV filter, respectively, with the directional response width on the x-axis and the different considered values for standard deviation of the DOA error on the y-axis. The presented results were averaged over 5 experiments for each standard deviation value of the DOA estimation error. We analyze the results from the perspective of the DOA estimation errors. It can be seen that for $\sigma = 5^\circ$, the interference reduction (IR) performance, Fig. 5(a) and 6(a), of both spatial filters is slightly affected by the width of the spatial window, with the best performance for a narrow spatial window width. However, as the standard deviation of the error is increased, the degradation in IR performance is more noticeable for the iLCMV filter compared to the proposed method. Such a trend is expected for the iLCMV filter as it directly utilizes the DOA estimates at each TF bin to formulate the directional constraints of the filter which leads to degradation in performance as the estimation error increases. On the other hand, the proposed framework involves a soft decision approach based on the approximated posterior probabilities which adds robustness against such errors.

In terms of the segSIR performance, Fig. 5(c) and 6(c), both filters achieve the best performance when the error is small and the width of the window is narrow. However, an interesting thing to note is that the worst performance of the iLCMV filter is obtained for high estimation error and narrow window width whereas for the proposed method the worst case is high estimation error with a broad window width. To understand this, it is necessary to look at the plot for the segSDI measure, Fig. 5(b) and 6(b). For both filters, the maximum amount of distortion is introduced to the desired signal when the estimation error is high and the window width is narrow, which is the worst

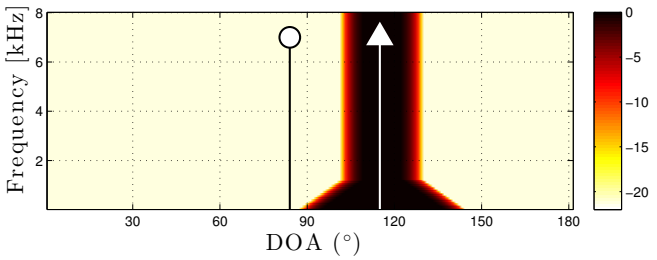


Fig. 7: Directional response function $\mathrm{g}(\theta, k)$ considered for the simulation experiment. The markers represent the direction of Source A (triangle) and Source B (circle).

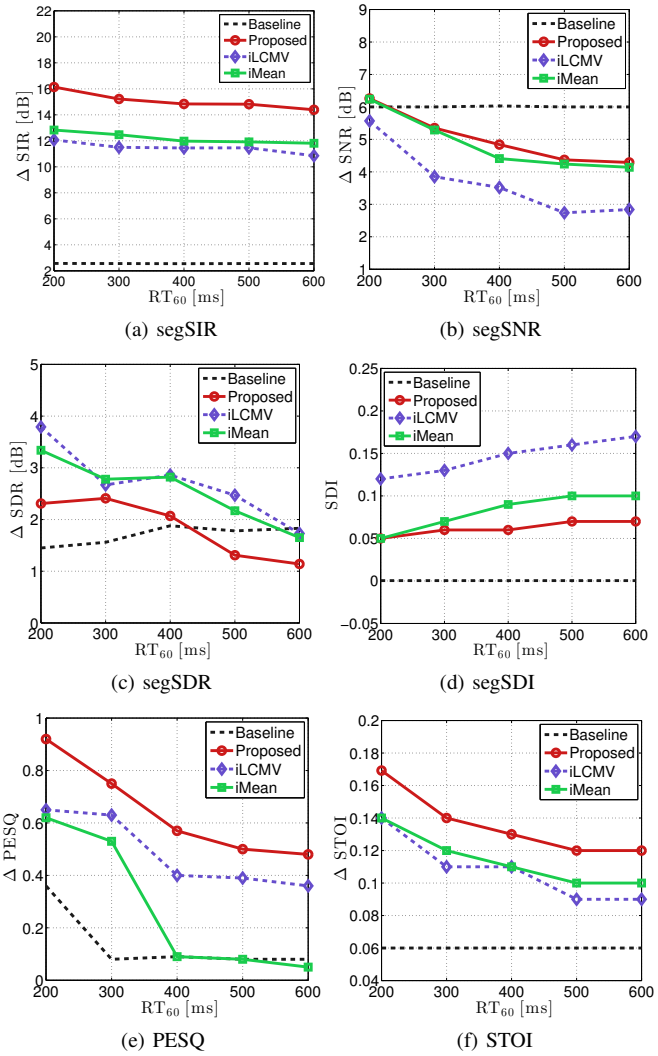
performance of the filters in terms of segSDI. Now, looking at the segIR plot, it can be seen that for the iLCMV filter the performance does not vary considerably over different window widths, for a high estimation error, whereas for the proposed method this difference is significant. Looking at the performance of the filters in terms of segIR and segSDI together, the trend noticed in the segSIR plot becomes evident.

The experiment presented in this section showed that the proposed method indeed introduces robustness against DOA estimation errors to the informed spatial filtering framework. In the following, we will demonstrate it applicability in adverse acoustic scenarios.

B. Experiment using simulated RIRs

In this section, we evaluate the performance of the proposed method as well as the iLCMV filter for different simulated room conditions. As a baseline, we evaluate the performance of a simple delay-and-sum beamformer (DSB) with perfect knowledge of the DOA of the desired source, which from hereon is referred to as the baseline. In our proposed method, we assume the fullband distribution of the narrowband DOA estimates to be modeled by L Gaussian mixtures, whose parameters are estimated using the standard EM algorithm. Therefore, in addition to the three mentioned methods, we also evaluated the performance of a variant of the iLCMV filter where the directional constraints are computed using the estimated means of the Gaussians. We would refer to this filter as iMean.

1) Experimental Setup: For the simulations, we consider a shoe-box room of dimensions $6 \times 5 \times 2$ m^3, where the reverberation time was varied from 0.2 to 0.6 s in steps of 0.1 s. We consider a ULA with $M = 4$ elements with an inter-element spacing of 3 cm. Two speech sources, Source A and B, were placed in front of the microphone array at a distance of 1.2 m. The desired source, Source A, was positioned at $\theta_{\mathrm{A}} = 115°$, with the interfering speaker, Source B was positioned at $\theta_{\mathrm{B}} = 84°$. The aim was to acquire Source A with unit gain while suppressing the interfering Speaker B by 21 dB, as shown in Fig. 7. The input signal consists of 1 s of silence followed by two speech signals active simultaneously for a 5 s duration, with a sampling rate of $F_s = 16$ kHz. A 512 point short-time Fourier transform (STFT) with 50% overlap is used to transform the signals into the TF domain. White noise was added to the speech signals resulting in an input segSNR of 10 dB. The microphone signal PSD matrix, $\mathbf{\Phi}_{\mathbf{y}}(n, k)$, was computed using the recursive temporal averaging filter given in (32), with a time constant of 50 ms. The noise PSD matrix was estimated from the silence portion. The diffuse sound PSD matrix was computed as described in Section VI-C. At each TF bin, $D = 2$ narrowband DOA estimates were obtained using ESPRIT. For the proposed method, the DOA space was sampled with a resolution of 10°, resulting in $I = 19$ discrete points. For computing the time-dependent supports for each



(a) segSIR (b) segSNR (c) segSDR (d) segSDI (e) PESQ (f) STOI

Fig. 8: Performance of the spatial filters for different reverberation times (using simulated RIRs). The plots show improvement with respect to a reference microphone signal, except for segSDI.

DOA random variable, the threshold was empirically set to $\delta = 0.15$. For our case, the EM algorithm was found to converge within 10 iterations, preceded by 5 iterations of the k-means initialization method. For all the experiments, these number of iterations were kept fixed.

We evaluated the performance of the spatial filters in terms of the following objective measures: segSIR improvement (ΔSIR), segSNR improvement (ΔSNR), segmental signal-to-diffuse ratio improvement (ΔSDR), segSDI [39, Eq. 4.44], improvement in PESQ score (ΔPESQ) [40] and improvement in short-time objective intelligibility (ΔSTOI) [41]. The performance measures are computed for non-overlapping segments of length $T = 30$ ms. For a segment i, the input SNR, SDR and SIR at a reference microphone are computed as

$$\mathrm{iSNR}(i) = 10\log_{10}{}^{\langle |x_A(t)|^2 \rangle}/_{\langle |x_{\mathrm{n}}(t)|^2 \rangle}, \quad (39)$$

$$\mathrm{iSDR}(i) = 10\log_{10}{}^{\langle |x_A(t)|^2 \rangle}/_{\langle |x_{\mathrm{d}}(t)|^2 \rangle}, \quad (40)$$

$$\mathrm{iSIR}(i) = 10\log_{10}{}^{\langle |x_A(t)|^2 \rangle}/_{\langle |x_B(t)|^2 \rangle}, \quad t \in ((i-1)T, iT] \quad (41)$$

where $\langle \cdot \rangle$ denotes average over t. The signals $x_A(t)$ and $x_B(t)$ are the received time-domain signals corresponding to the speakers located at θ_A and θ_B, respectively. The terms $x_{\mathrm{d}}(t)$ and $x_{\mathrm{n}}(t)$ denote the

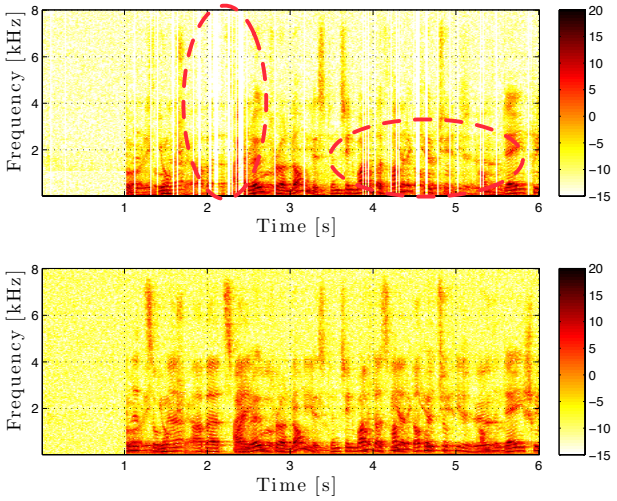


Fig. 9: Output of the iMean spatial filter (upper figure) and the corresponding input signal at a reference microphone (lower figure) for RT_{60} = 400 ms, from the experiment with simulated RIRs presented in Section VII-B.

diffuse sound and the microphone self-noise component, respectively. The output values oSNR, oSDR and oSIR are computed similarly, by using the filtered versions of the signals. The corresponding improvements, ΔSNR, ΔSDR, and ΔSIR, are computed by taking the difference between the corresponding output and the input measures.

2) Results: The results for this experiment are presented in Fig. 8. It can be seen that for all the filters, except the baseline, there is deterioration in performance as the reverberation time increases. The baseline method provides almost the same performance for all reverberation times. As shown by Fig. 8(a), the proposed method provides more suppression of the interfering speaker compared to other methods while introducing less distortion to the desired signal, except for the baseline method which has perfect knowledge of the desired source DOA (Fig. 8(d)). However, due to the weighted spatial averaging incorporated in the proposed method to account for the unreliability of the DOA estimates in adverse acoustic conditions, it is unable to suppress the diffuse noise as well as the other informed spatial filters (Fig. 8(c)). In terms of noise suppression (Fig. 8(b)), the baseline method performs the best, as expected, followed by the proposed and the iMean filter, which have a similar performance. In terms of STOI improvement, the proposed filter performs better than the other methods, with the iMean and the iLCMV filters having a similar performance.

In Fig. 8(d), it can be seen that the iMean filter introduces less distortion to the desired signal compared to the iLCMV filter. However, in terms of PESQ (Fig. 8(e)), its performance severely deteriorates beyond 300 ms reverberation time. The reason for this deterioration can be seen in Fig. 9, which presents the output of the iMean filter and the corresponding input signal for RT_{60} = 400 ms, from the above mentioned experiment. Since the iMean filter employs the estimated means of the Gaussians as the DOAs of the plane waves, if the mean of the Gaussian corresponding to the desired source is estimated to be outside the designed spatial window, both sources get suppressed by 21 dB for that whole time frame which leads to the clipping artifacts visible in the output of the iMean filter in Fig. 9. Since it is more difficult to have reliable DOA estimates at higher reverberation times, these clipping artifacts severely deteriorate the performance of the iMean filter in terms of PESQ. This shows that

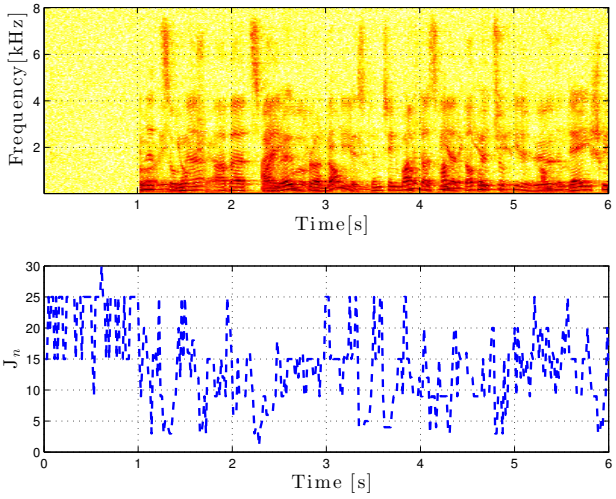


Fig. 10: Example of variation of J_n over time for RT_{60} of 300 ms.

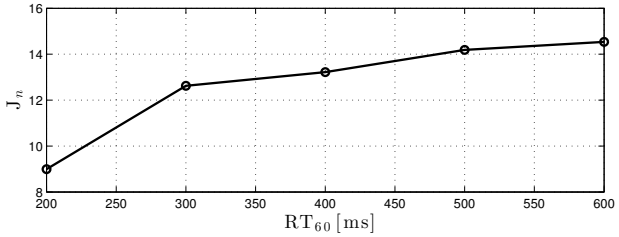
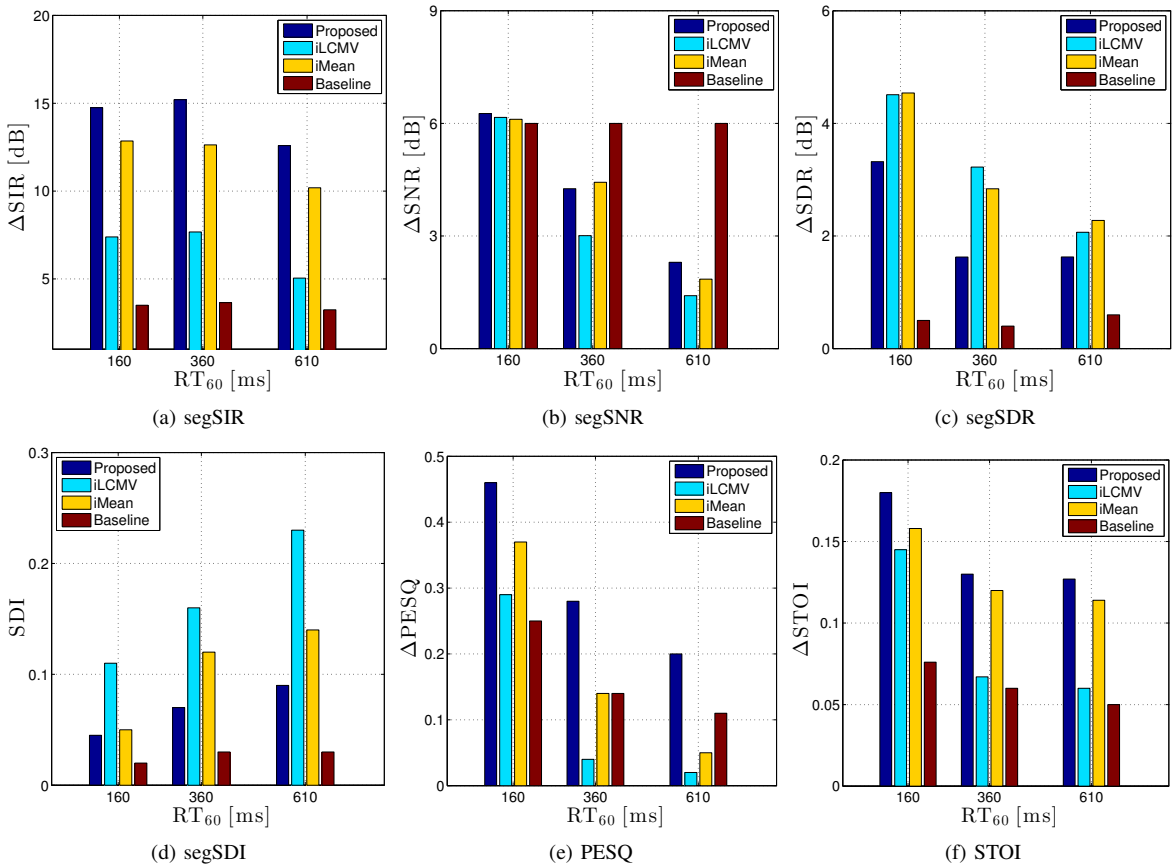

J_n
RT_{60} [ms]

Fig. 11: Average J_n for different reverberation times. The averaging was done over all time frames with speech present.

though the estimated means of the Gaussians can possibly provide better DOA estimates, in case of an error, the deterioration in terms of the perceptual quality of the signal is even more severe compared to utilizing the obtained narrowband DOA estimates directly in the spatial filter.

In general, the results in Fig. 8 show that incorporating the narrowband DOA estimates directly into a spatial filter does not provide a good overall performance, especially in highly noisy and reverberant conditions, as shown by the results for the iLCMV filter. By using the estimated means of the Gaussians as the DOA estimates, slight improvement in performance can be achieved, however as we saw, significant deviation of the estimated means from the true source DOAs leads to severe degradation in terms of estimated perceptual quality. The proposed method, by accounting for the uncertainty in the DOA estimates, provides the best overall performance, with the only limitation being its ability to reduce the diffuse sound component.

3) Number of combinations J_n: We also analyzed the reduction of the total number of combinations, which is performed along with the approximation of the posterior probabilities in the proposed method (see Section V-B). In Fig. 10, an example of the variation of the total number of combinations J_n over time, for RT_{60} = 300 ms, is presented. It can be seen that when there is silence, the value of J_n is high since the DOA estimates do not correspond to a specific direction in this case. We aim to address this problem of extra computations during silence in the future. Once the speakers become active, based on the accuracy of the DOA estimates, there is variation in the value of J_n over time.



(a) segSIR (b) segSNR (c) segSDR

(d) segSDI (e) PESQ (f) STOI

Fig. 12: Objective measures with measured RIRs for the experiment described in Section VII-C for the source-array distance of 1 m. The plots show improvement with respect to a reference microphone signal, except for segSDI.

Baseline	iLCMV	iMean	Proposed
1	2.2	22.3	32.1

TABLE I: Average run-times of the different methods normalized by the run-time of the baseline method.

In Fig. 11, the average value of J_n for different reverberation times is presented. As the reverberation time increases, the narrowband DOA estimates becomes less accurate which leads to a larger standard deviation in their fullband distribution. Due to this, the average number of combinations for which the proposed spatial filter needs to be computed increases as the reverberation time increases, thereby signifying that the complexity of the proposed method is dependent on the reliability of the DOA estimates, which also is the main mechanism for incorporating robustness against DOA estimation errors.

4) Average run-time of compared methods: In this section, the average run-times of the different methods are presented. The numbers in Table I represent the run-time of the respective methods as a factor of the computation time for the baseline DSB. All the methods were implemented using MATLAB®. The run-time of the methods was computed by averaging over all the experiments with different reverberation times.

From the numbers it is clear that though the proposed method has a superior performance, its run-time is significantly higher for our implementation. In future work, we would like to explore optimized implementations as well as methods to reduce the computational cost of the proposed method.

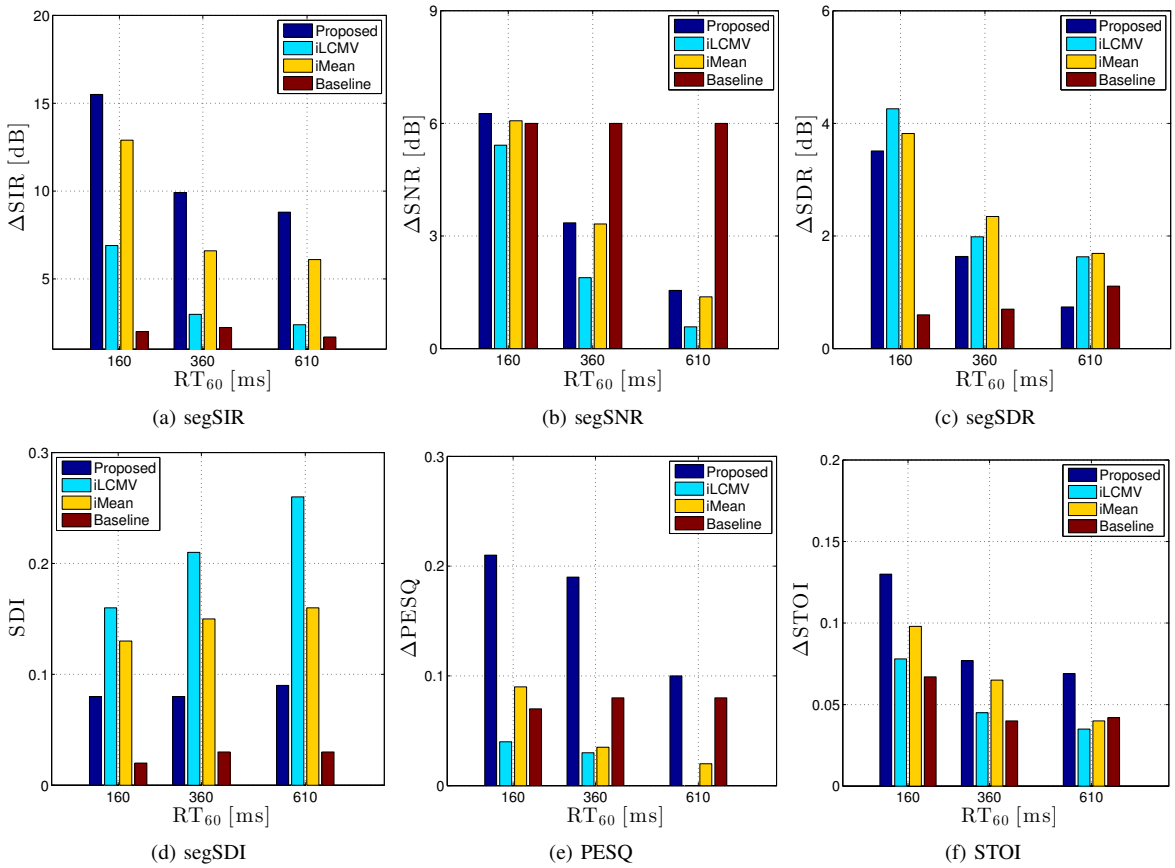
C. Experiment with measured RIRs

To verify the efficiency of the proposed method in practical acoustic scenarios, a comparative analysis of all the spatial filters was performed with measured RIRs.

1) Experimental setup: For our experiments with measured RIRs, we used the Multichannel Impulse Response Database from Bar-Ilan university [42]. The database consists of RIRs measured at Bar-Ilan university's acoustics lab, of size $6 \times 5 \times 2$ m^3, for three different acoustic scenarios with reverberation times of RT_{60} = 160, 360, and 610 ms. The recordings were done for several source positions placed on a spatial grid of semi-circular shape covering the whole angular range for a linear array, i.e., $[0^\circ, 180^\circ]$, in steps of 15° at distances of 1 m and 2 m from the center of the microphone array.

The recordings were done with a linear microphone array with three different microphone spacings. For our experiment, we chose the [3, 3, 3, 8, 3, 3, 3] cm setup [42], which consists of eight microphones where the distance between the 2 center microphones is 8 cm with the distance between the other microphones being 3 cm. We selected a subset of four microphones out of the total eight microphones used in the original setup, to have a ULA with $M = 4$ elements with an inter-microphone distance of 3 cm, which corresponds to the setup used in Section VII-B.

For this experiment, when the sources were placed on the grid 1 m away from the microphone array, the true source DOAs were, $\theta_{\mathrm{A}} = 109^\circ$ and $\theta_{\mathrm{B}} = 79^\circ$, and for 2 m distance, $\theta_{\mathrm{A}} = 107^\circ$ and



(a) segSIR (b) segSNR (c) segSDR

(d) segSDI (e) PESQ (f) STOI

Fig. 13: Objective measures with measured RIRs for the experiment described in Section VII-C for the source-array distance of 2 m. The plots show improvement with respect to a reference microphone signal, except for segSDI.

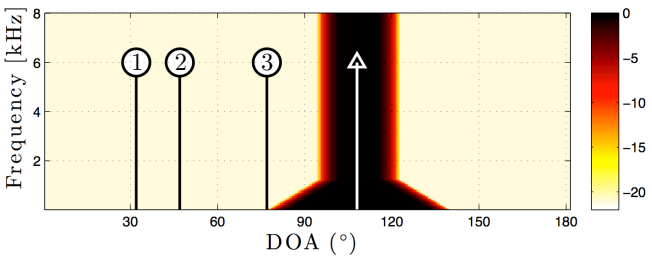


Fig. 14: Directional response function $g(\theta, k)$ considered for the experiment with moving interfering speaker. The markers represent the direction of Source A (triangle) and Source B (circle).

$\theta_B = 77°$. The directional response function was considered to be the one given in Fig. 7, with the directional window shifted according to the true DOA of Source A. The input signal consisted of 1 s silence, followed by 5 s of double talk. White noise was added to the input signal for an input segSNR of 10 dB.

The remaining experimental parameters are the same as in the simulation experiments, described in Section VII-B1. The performance of the filters was evaluated for the three different reverberation times.

2) Results: The results for this experiment are presented in Figs. 12 and 13, for source-array distances of 1 m and 2 m, respectively.

When the sources are relatively close to the array, i.e., 1 m, it can be seen that the degradation in performance of the informed filters is not prominent in terms of interference suppression (Fig. 12(a)). Due to the highly noisy environment, the iLCMV filter already suffers from incorporating unreliable DOA estimates in the filtering framework, whereas the performance of the iMean filter is close to the proposed approach. However, in terms of signal distortion (Fig. 12(d)), as the reverberation time increases, the difference in performance of the proposed approach and the iMean filter becomes significant. A similar trend can also be noticed in the performance of the spatial filters in terms PESQ score improvement (Fig. 12(e)). In terms of STOI improvement (Fig. 12(f)), the performance of the proposed approach is slightly better than the other methods.

When the sources move further away from the microphone array, i.e., 2 m, the degradation in performance of all the informed spatial filters, in terms of interference suppression (Fig. 13(a)), becomes significant as the reverberation time increases. In terms of signal distortion (Fig. 13(d)) as well as PESQ score improvement (Fig. 13(e)), the difference in performance of the proposed approach compared to the other informed spatial filters, especially the iMean filter, is more significant for all the different acoustic conditions.

In terms of noise suppression (Fig. 12(b) and Fig.13(b)), for both source-array distances, the performance of the informed spatial filters, especially the proposed method and iMean filter, are similar and marginally better than iLCMV filter, with deterioration in performance observed as the reverberation time increases. Also, as observed in the simulation experiments, the proposed method suppresses less diffuse sound compared to the other informed spatial filters (Fig. 12(c) and Fig. 13(c)).

Overall, the results in Figs. 12 and 13 show that despite the limitation in terms of diffuse noise suppression, the proposed method

TABLE II: Performance of the spatial filters for moving undesired sound source [improvement compared to an unprocessed microphone signal]. Best values are underlined.

	segSIR	segSNR	segSDR	segSDI	PESQ	STOI
Baseline	1.7	<u>6.1</u>	1.1	<u>0.01</u>	0.13	0.04
iLCMV	5.2	3.7	3.1	0.23	0.08	0.05
iMean	9.1	3.7	<u>3.9</u>	0.15	0.17	0.07
Proposed	<u>11.4</u>	3.6	2.6	0.10	<u>0.24</u>	<u>0.09</u>

TABLE III: Performance of the spatial filters for two simultaneously active sound sources [improvement compared to an unprocessed microphone signal]. Best values are underlined.

	segSIR	segSNR	segSDR	segSDI	PESQ	STOI
Baseline	2.3	<u>6.1</u>	1.9	<u>0.01</u>	0.05	0.04
iLCMV	9.0	2.3	4.8	0.30	0.07	0.01
iMean	13.0	3.1	<u>6.3</u>	0.12	0.03	0.04
Proposed	<u>13.2</u>	4.4	5.6	0.12	<u>0.11</u>	<u>0.07</u>

provides a better overall performance than the other methods. The robustness against DOA estimation errors of the proposed method is mainly evident from the low amount of signal distortion introduced as well as the superior improvement of perceptual quality shown by the proposed method for higher reverberation times even when the sources are further away from the microphone.

3) Moving interfering speaker: We also evaluated the performance of all the spatial filters for a scenario with a moving undesired source.

For this, we used the measured RIRs from the previous experiment. We chose the acoustic scenario with RT_{60} = 360 ms, with the sources placed 2 m away from the microphone. The desired source was positioned in the same direction as before, i.e. $\theta_{\mathrm{A}} = 107^{\circ}$, however the interfering speaker is now subsequently active at the positions 1 to 3, with $\theta_{\mathrm{B1}} = 31^{\circ}$, $\theta_{\mathrm{B2}} = 47^{\circ}$ and $\theta_{\mathrm{B3}} = 77^{\circ}$, as shown in Fig. 14. The input signal for this experiment consisted of 1 s of silence, followed by both the desired and the interfering speakers being active for 12 s. The interfering speaker was active at each position for 4 s. The rest of the experimental parameters are the same as in the simulation experiments.

The results for this experiment are presented in Table II. From the results, it can be seen that even with a moving undesired source, the proposed spatial filter introduces very low signal distortion while providing a better interference suppression compared to the other methods. The improvement in terms of the estimated perceptual quality (PESQ) as well as predicted intelligibility (STOI) is also superior to the other methods. Overall, the presented results demonstrate that the proposed method is able to deal with a dynamic acoustic scenario, while incorporating robustness against DOA estimation errors. Audio examples for this experiment is available online at [43].

4) Multiple interfering speakers: Finally, we evaluate the performance of all the filters for a scenario where two interfering speakers are simultaneously active from positions 1 and 2 in Fig.14, and the desired speaker is positioned in the same direction as before. The input signal consisted of 1 s of silence followed by 7 s of simultaneous activity of all three speakers. The experimental parameters are the same as for the previous experiment. We assume $L = 2$ for this experiment to evaluate the robustness of the system to errors in our knowledge of L.

The results for this experiment are given in Table III. With two simultaneously active interfering speakers, it can be seen that the performance of all the filters are slightly reduced compared to the scenario with static single interfering source, presented in SectionVII-C2. This is mainly due to the assumption of $L = 2$, which results in a model violation when three speakers are simultaneously active. However, it can be seen that the proposed method still provides the best performance, especially in terms of estimated perceptual quality (PESQ) and predicted intelligibility (STOI), thereby showing that the proposed method is able to provide robustness to model mismatch. Audio examples for this experiment are also available online at [43].

VIII. Conclusions

A Bayesian approach to informed spatial filtering that provides robustness against DOA estimation errors was proposed. The proposed method can be viewed as a efficient way of incorporating estimated narrowband DOA information in a spatial filtering framework. In the informed spatial filters, the narrowband DOA estimates were directly employed to compute the directional constraints of the spatial filter which made it susceptible to degradation in performance due to estimation errors, whereas the Bayesian framework presented in the first part of this paper needed to compute required parameters for all possible combinations of discrete DOA values in the whole DOA space due to lack of information regarding the source DOAs which lead to a large number of redundant computations. In the proposed approximation method, the DOA estimates were employed to approximate the posterior probabilities which were subsequently used to isolate regions in the DOA space with high probability of containing the actual source DOAs, thereby incorporating the source DOA information to reduce the computational cost while still achieving robustness against DOA estimation errors. interfering speaker Experimental evaluation with both simulated and measured RIRs demonstrated the robustness of the proposed method to DOA estimation errors. It showed that the proposed method is able to suppress the sound sources(s) outisde the directional region of interest while introducing low levels of distortion to the sound source within the specified region of interest, at the cost of a slight decrease in diffuse sound reduction.

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] J.Li and P. Stoica, *Robust Adaptive Beamforming*, John Wiley & Sons, 2005.

[10] Y.I Abramovich, "Controlled method for adaptive optimization of filters usisng the criteriion of maximum snr," *Radio Eng. Electron. Phys.*, vol. 26, pp. 87–95, Mar. 1981.
[11] B.D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul 1988.
[12] 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.
[13] S. Applebaum and D. Chapman, "Adaptive arrays with main beam constraints," *IEEE Trans. Antennas Propag.*, vol. 24, no. 5, pp. 650–662, Sep 1976.
[14] B.D. Van Veen, "Minimum variance beamforming with soft response constraints," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 1964–1972, Sep 1991.
[15] 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.
[16] Chun-Yang Chen and P.P. Vaidyanathan, "Quadratically constrained beamforming robust against direction-of-arrival mismatch," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4139–4150, Aug 2007.
[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] J. Yang and A. L. Swindlehurst, "The effects of array calibration errors on df-based signal copy performance," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2724–2732, Nov 1995.
[22] B. Friedlander and B. Porat, "Performance analysis of a null-steering algorithm based on direction-of-arrival estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 37, no. 4, pp. 461–466, April 1989.
[23] M. H. Er and B. C. Ng, "A new approach to robust beamforming in the presence of steering vector errors," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1826–1829, Jul 1994.
[24] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.
[25] S. Chakrabarty, O.Thiergart, and E. A. P. Habets, "A method to analyze the spatial response of informed spatial filters," in *12.ITG Fachtagung Sprachkommunikation*, 2016.
[26] S. Chakrabarty, O. Thiergart, and E.A.P. Habets, "A Bayesian approach to spatial filtering and diffuse power estimation for joint dereverberation and noise reduction," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 753–757.
[27] O. Thiergart and E. Habets, "Sound field model violations in parametric spatial sound processing," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2012.
[28] 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.
[29] S. Braun, O. Thiergart, and E. A. P. Habets, "Automatic spatial gain control for an informed spatial filter," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 830–834.
[30] H. L. van Trees, *Detection, Estimation, and Modulation Theory*, vol. IV, Optimum Array Processing, Wiley, New York, USA, Apr. 2002.
[31] A. J. Laub, *Matrix Analysis for Scientists and Engineers*, p. 103, Society for Industrial and Applied Mathematics (SIAM), 2005.
[32] S. Markovich-Golan, S. Gannot, and I. Cohen, "Low-complexity addition or removal of sensors/constraints in LCMV beamformers," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1205–1214, March 2012.
[33] T. Gerkmann and R. C. Hendriks, "Noise power estimation based on the probability of speech presence," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2011.
[34] M. Souden, J. Chen, J. Benesty, and S. Affes, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, pp. 2159 – 2169, 2011.
[35] E. A. P. Habets, "A distortionless subband beamformer for noise reduction in reverberant environments," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, August 2010.
[36] 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.
[37] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley and Sons, second edition, 2001.
[39] J. Benesty, J. Chen, and E. A. P. Habets, "Speech enhancement in the stft domain," in *Springer Briefs in Electrical and Computer Engineering*, 2012.
[40] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2001, vol. 2, pp. 749–752.
[41] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 2010, pp. 4214–4217.
[42] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Sept 2014, pp. 313–317.
[43] S. Chakrabarty and E.A.P. Habets, "Audio examples for a bayesian approach to informed spatial filtering with robustness against doa estimation errors," [Online] Available: https://www.audiolabs-erlangen.de/resources/2016-TASLP-BayesianSF.

Soumitro Chakrabarty (S'10) received his BEng. degree from Manipal University, India in 2010 and his MSc. degree from Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland in 2013. He is currently pursuing a PhD. at International Audio Laboratories Erlangen, Germany on the topic of robust estimation of spatial information for microphone array processing.

His current research interests include spatial filtering, sound source localization, multi-microphone source extraction and machine learning for array processing.

Emanuël A.P. Habets (S'02-M'07-SM'11) is an Associate Professor at the International Audio Laboratories Erlangen (a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS), and Head of the Spatial Audio Research Group at Fraunhofer IIS, Germany. He received the B.Sc. degree in electrical engineering from the Hogeschool Limburg, The Netherlands, in 1999, and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, The Netherlands, in 2002 and 2007, respectively.

From 2007 until 2009, he was a Postdoctoral Fellow at the Technion - Israel Institute of Technology and at the Bar-Ilan University, Israel. From 2009 until 2010, he was a Research Fellow in the Communication and Signal Processing Group at Imperial College London, U.K.

His research activities center around audio and acoustic signal processing, and include spatial audio signal processing, spatial sound recording and reproduction, speech enhancement (dereverberation, noise reduction, echo reduction), and sound localization and tracking.

Dr. Habets was a member of the organization committee of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC) in Eindhoven, The Netherlands, a general co-chair of the 2013 International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York, and general co-chair of the 2014 International Conference on Spatial Audio (ICSA) in Erlangen, Germany. He was a member of the IEEE Signal Processing Society Standing Committee on Industry Digital Signal Processing Technology (2013-2015), a Guest Editor for the IEEE Journal of Selected Topics in Signal Processing and the EURASIP Journal on Advances in Signal Processing, and an Associate Editor of the IEEE Signal Processing Letters (2013-2017). He is the recipient, with S. Gannot and I. Cohen, of the 2014 IEEE Signal Processing Letters Best Paper Award. Currently, he is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, vice-chair of the EURASIP Special Area Team on Acoustic, Sound and Music Signal Processing, and Editor in Chief of the EURASIP Journal on Audio, Speech, and Music Processing.