INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

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

Soumitro Chakrabarty, Oliver Thiergart, Emanuel. A. P. Habets

International Audio Laboratories Erlangen

E-Mail: soumitro.chakrabarty@audiolabs-erlangen.de

1. Introduction

Motivation

- Spatial filters require information regarding the direction-of-arrival (DOA) of the sound source(s).
- In [1], a spatial filter was proposed based on instantaneous DOA estimates to capture L sound sources with arbitrary spatial response.
- Difficult to get reliable DOA estimates in noisy and reverberant conditions.

Proposed Method

- We reformulate the approach in [1] and propose a Bayesian approach, to account for uncertainities in the estimated DOAs.
- The resulting Bayesian filter is a weighted sum of spatial filters pointed at a discrete set of DOAs.
- A probabilistic approach to diffuse sound power estimation is presented that gives a sufficiently accurate estimate to achieve joint dereverberation and noise reduction.

2. Problem Formulation

- **Aim**: Capture the directional sounds from a specific spatial region with a specific gain while attenuating the diffuse sound and microphone self-noise.

- Consider a uniform linear array (ULA) of M microphones located at $\mathbf{d}_{1\ldots M}$. The microphone signals in STFT domain

time index frequency index

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

direct sound diffuse sound microphone noise

- Assuming signal components are mutually uncorrelated

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

- Desired signal

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

Definitions

- Microphone signal of l-th plane wave
  $\mathbf{x}_l(n,k) = [X_l(n,k,\mathbf{d}_1)\ldots X_l(n,k,\mathbf{d}_M)]^T$
- Sound pressure of l-th plane wave
  $\mathbf{x}_l(n,k) = \mathbf{a}(\theta_l,k) X_l(n,k,\mathbf{d}_1)$
- m-th element of steering vector
  $a_m(\theta_l,k) = \exp\{-j\kappa r_m \cos\theta_l(n,k)\}$
- Power spectral density (PSD) matrix
  $\mathbf{\Phi}_{\mathbf{y}}(n,k) = \mathrm{E}\{\mathbf{y}(n,k)\mathbf{y}^{\mathrm{H}}(n,k)\}$
- Undesired PSD matrix
  $\mathbf{\Phi}_{\mathrm{d}}(n,k) = \phi_{\mathrm{d}}(n,k)\,\mathbf{\Gamma}_{\mathrm{d}}(k)$,
  $\mathbf{\Phi}_{\mathrm{n}}(n,k) = \phi_{\mathrm{n}}(n,k)\,\mathbf{I}$
- $G(\theta_l,k)$ is a real-valued arbitrary directivity function which can be designed based on the target application.

3. Proposed Method

Bayesian spatial filter

- DOA is modelled as a discrete random variable with a prior over the candidate set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_I\}$, with $I \gg L$.
- Approximation of the desired signal

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

posterior probabilities

desired signal estimate from direction $\theta_i \in \Theta$

- Conditional estimate of the desired signal

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

- Individual spatial filter weights

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

- Posterior probabilities

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

Likelihood

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

Signal power estimate

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

Determinant

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

Inverse

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

Diffuse sound power estimation

- Auxiliary spatial filter

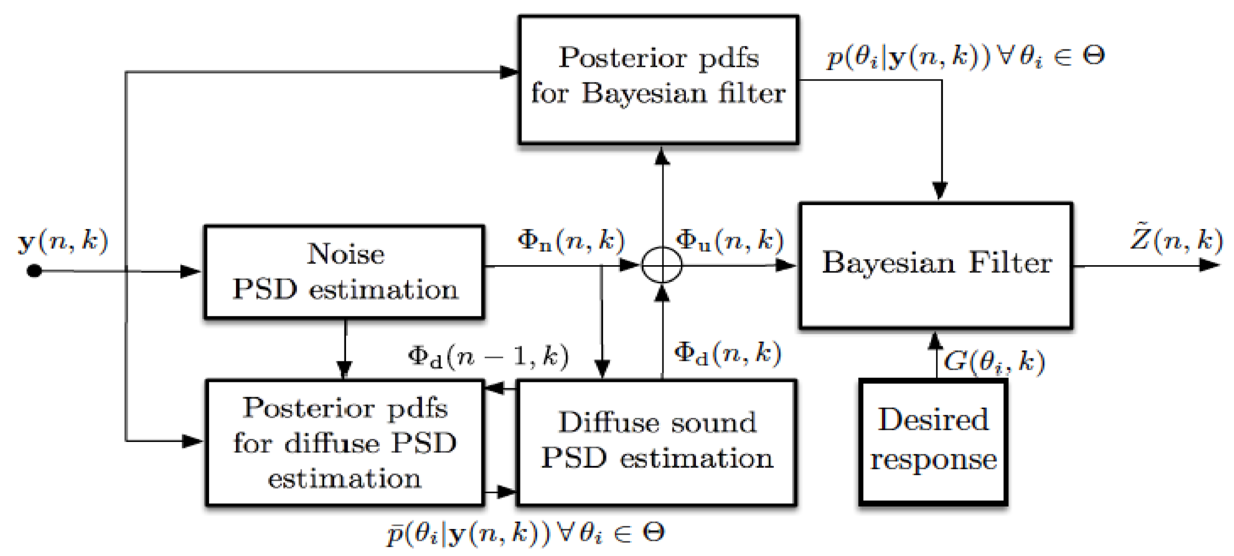
$$\mathbf{w}_{\mathrm{d}} = \arg\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{w}$$
s.t.
$$\mathbf{w}^{\mathrm{H}}\mathbf{a}(\theta^l_{\max}) = 0, \quad l \in \{1,\ldots L\},$$
$$\mathbf{w}^{\mathrm{H}}\mathbf{a}(\theta_{\min}) = 1,$$

$\theta^l_{\max}$ are the DOAs of the L plane waves, obtained by selecting L local maximas in $\bar{p}(\theta_i|\mathbf{y}(n,k))$

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

- Diffuse sound power

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



4. Results

Simulation Setup

- Room: 7.0m X 5.4m X 2.5m.
- ULA with M = 5.
- L=2, I = 37.
- Fs = 16 kHz.
- 512-point STFT, 50% overlap.
- $\theta_A = 135^\circ$ $\theta_B = 45^\circ$
- Scenario I:
  RT_{60} = 260 ms, SSNR = 27dB
- Scenario II:
  RT_{60} = 450 ms, SSNR = 17dB

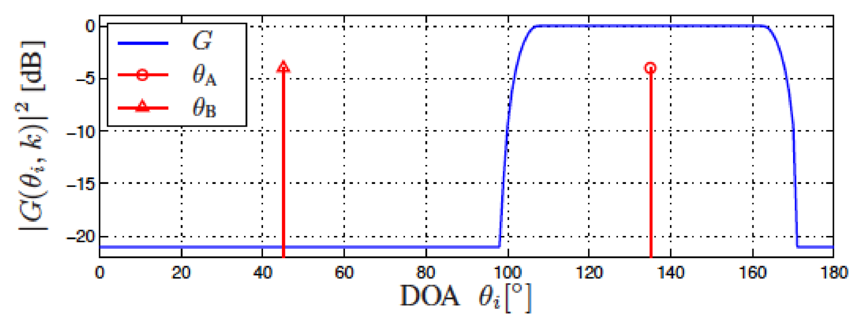
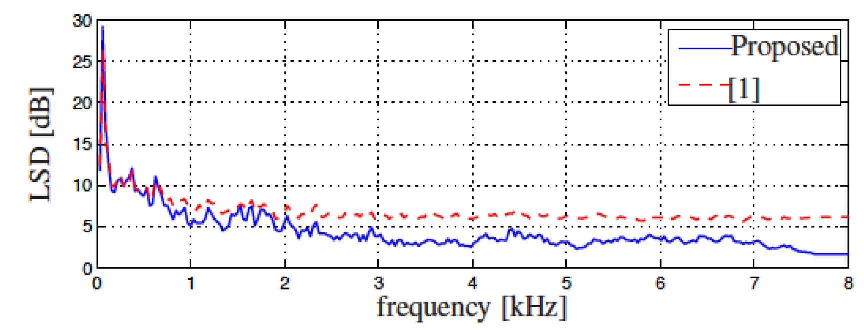


Figure:Directivity function and source positions

Diffuse power estimation performance

- Evaluated only for Scenario II.
- Figure shows estimation error in terms of log-spectral distance (LSD) across frequencies averaged over time.



Bayesian spatial filter performance

Scenario I

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

Scenario II

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

5. Conclusion

- Proposed Bayesian spatial filter provides an overall better performance, especially for the more challenging Secnario II.
- Probabilistic approach to diffuse power estimation has lower eestimation error than the method proposed in [1].
- Proposed filter introduces small amount of distortion to the desired signal.

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