POLITECNICO DI MILANO

Estimation of the Sound Field at Arbitrary Positions in Distributed Microphone Networks Based on Distributed Ray Space Transform

M. Pezzoli, F. Borra, F. Antonacci, A. Sarti, S. Tubaro

Outline

- Goals
- State of the art
- Sound Field Estimation
- Parameter Estimation
  - Source Localization
  - Radiance Pattern Estimation
  - Signal Reconstruction
- Sound Synthesis
- Simulations and Experiments Results
- Conclusions
- Future Works

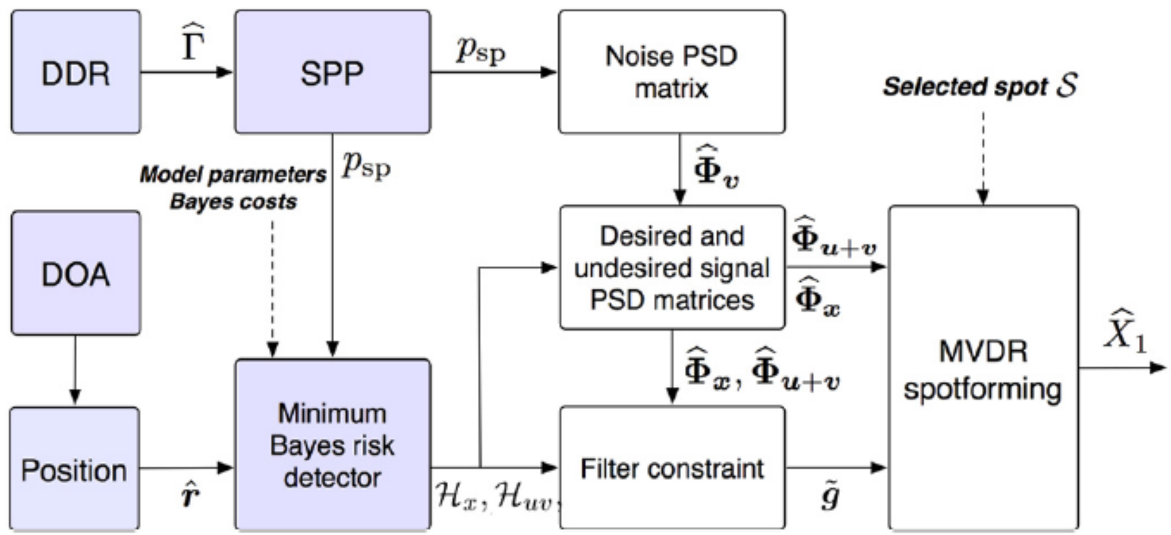
Goal of Virtual Miking

- Goal: reconstruct the sound field at an arbitrary position in the presence of multiple sources.
- Input data: signals coming from spatially-distributed compact arrays
- Assumptions:
  - Low reverberation
  - Virtual microphones has an omnidirectional sensitivity pattern
  - Coplanar arrays and sources
- Approach: sources parameterized by their signal, radiance pattern and location in space

Previous Techniques (1/2)

- Sources are extracted from a spot of interest (SOI), while reducing interferers and background noise in a fully data-dependent framework, which requires the estimation of the PSD of the sources on the scene (*)
- Sources are assumed to be omnidirectional

Spotforming method [1]



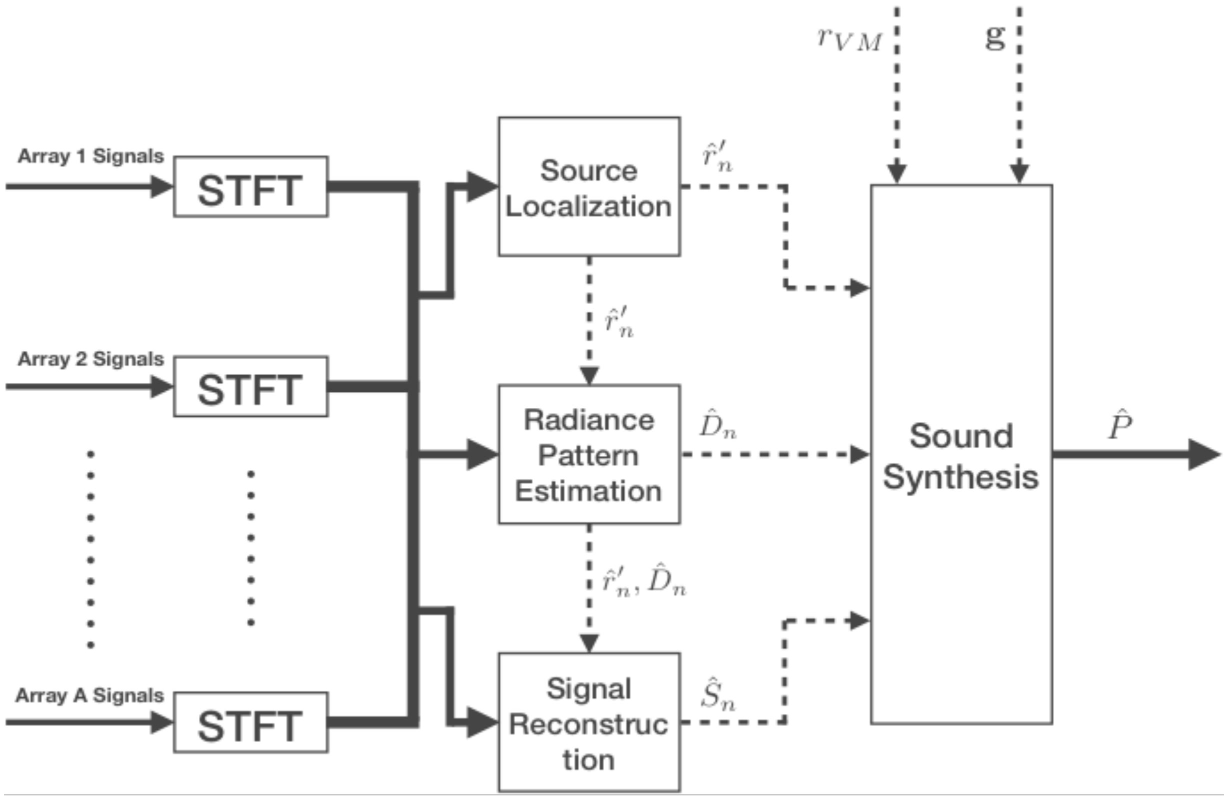
(*) M. Taseska, E.A.P. Habets, "Spotforming: Spatial Filtering With Distributed Arrays for Position-Selective Sound Acquisition," in IEEE/ACM TrASLP, Vol. 24, no. 7, pp. 1291-1304, July 2016.

Previous Techniques (2/2)

- Higher-Order Ambisonics (HOA) and Wave Field Synthesis (WFS) aim at reconstructing the sound field at a given location (HOA) or region (WFS) in space
- HOA usually has a small rendering region and WFS needs a large number of spatially distributed microphones

Our Approach to Sound Field Estimation (1/2)

We propose a **parametric sound field reconstruction approach** based on the estimation of **a limited number of acoustic source descriptors** (*position, radiation pattern and source signal*), given the signals acquired by **a few arbitrarily placed microphone arrays.**



Our Approach to Sound Field Estimation (2/2)

Signal Model

Signal acquired by the *m*-th microphone in *a*-th array

$$\boxed{P(\mathbf{r}_m^{(a)},\omega)} = \sum_{n=0}^{N-1} D_n(\theta_{m,n}^{(a)},\omega)\, g(\mathbf{r}_m^{(a)},\mathbf{r'}_n,\omega)\, S_n(\omega) + e_m^{(a)}(\omega)$$

Signal Model

Signal acquired by the *m*-th microphone in *a*-th array

$$\boxed{P(\mathbf{r}_m^{(a)}, \omega)} = \sum_{n=0}^{N-1} \boxed{D_n(\theta_{m,n}^{(a)}, \omega)}\, g(\mathbf{r}_m^{(a)}, \mathbf{r'}_n, \omega)\, S_n(\omega) + e_m^{(a)}(\omega)$$

Radiance Pattern of the *n*th source

Signal Model

Green's function

$$\frac{e^{-jk\|\mathbf{r}_m^{(a)}-\mathbf{r'}_n\|}}{4\pi\|\mathbf{r}_m^{(a)}-\mathbf{r'}_n\|}$$

Signal acquired by the *m*-th microphone in *a*-th array

$$\boxed{P(\mathbf{r}_m^{(a)},\omega)} = \sum_{n=0}^{N-1} \boxed{D_n(\theta_{m,n}^{(a)},\omega)}\,\boxed{g(\mathbf{r}_m^{(a)},\mathbf{r'}_n,\omega)}\,S_n(\omega) + e_m^{(a)}(\omega)$$

Radiance Pattern of the *n*th source

Signal Model

Green's function

$$\frac{e^{-jk\|\mathbf{r}_m^{(a)}-\mathbf{r}'_n\|}}{4\pi\|\mathbf{r}_m^{(a)}-\mathbf{r}'_n\|}$$

Signal acquired by the *m*-th microphone in *a*-th array

$$\boxed{P(\mathbf{r}_m^{(a)},\omega)} = \sum_{n=0}^{N-1} \boxed{D_n(\theta_{m,n}^{(a)},\omega)}\,\boxed{g(\mathbf{r}_m^{(a)},\mathbf{r}'_n,\omega)}\,\boxed{S_n(\omega)} + e_m^{(a)}(\omega)$$

Radiance Pattern of the *n*th source

Signal of *n*th source

Signal Model

Green's function

$$\frac{e^{-jk\|\mathbf{r}_m^{(a)}-\mathbf{r}'_n\|}}{4\pi\|\mathbf{r}_m^{(a)}-\mathbf{r}'_n\|}$$

Signal acquired by the *m*-th microphone in *a*-th array

Microphone self-noise

$$\boxed{P(\mathbf{r}_m^{(a)},\omega)} = \sum_{n=0}^{N-1} \boxed{D_n(\theta_{m,n}^{(a)},\omega)}\boxed{g(\mathbf{r}_m^{(a)},\mathbf{r}'_n,\omega)}\boxed{S_n(\omega)} + \boxed{e_m^{(a)}(\omega)}$$

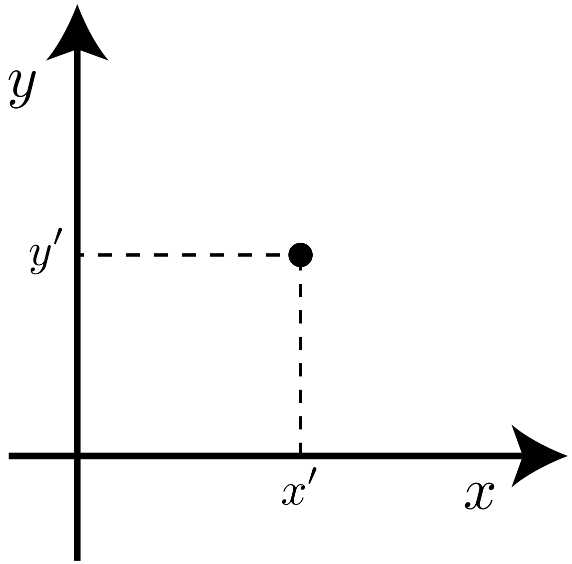
Radiance Pattern of the *n*th source

Signal of *n*th source

Source Localization (1/2)

Review of the **Projective Ray Space representation** of a **point source**

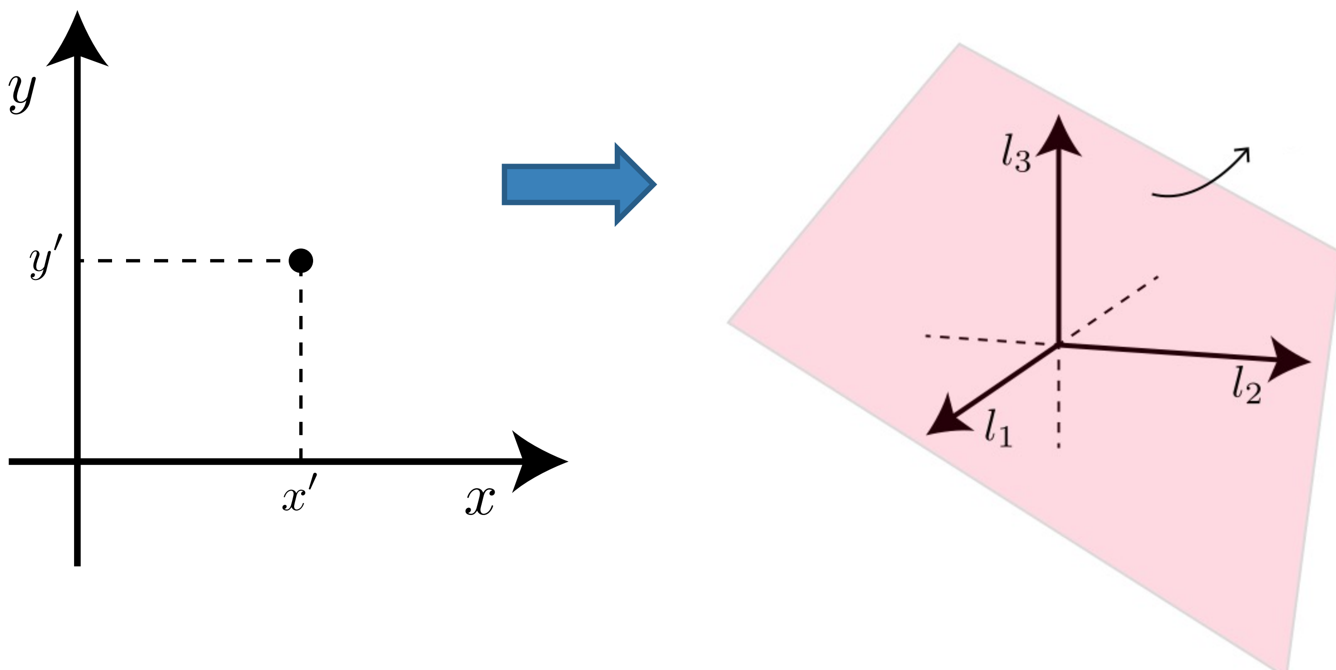
Geometric Space



Source Localization (1/2)

Review of the **Projective Ray Space representation** of a **point source**

Geometric Space **Projective Ray Space**

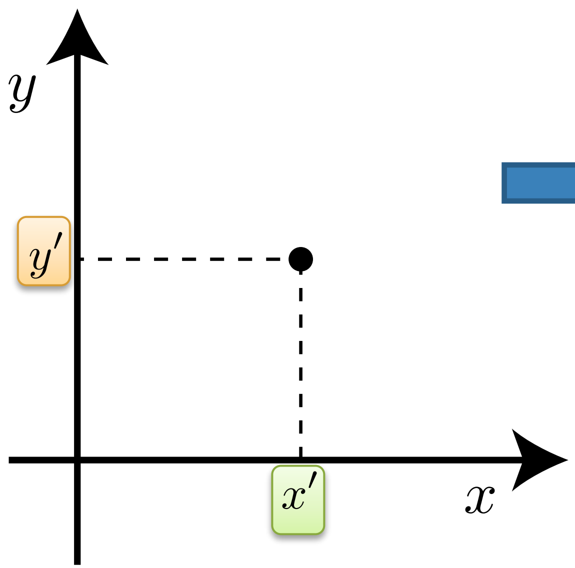


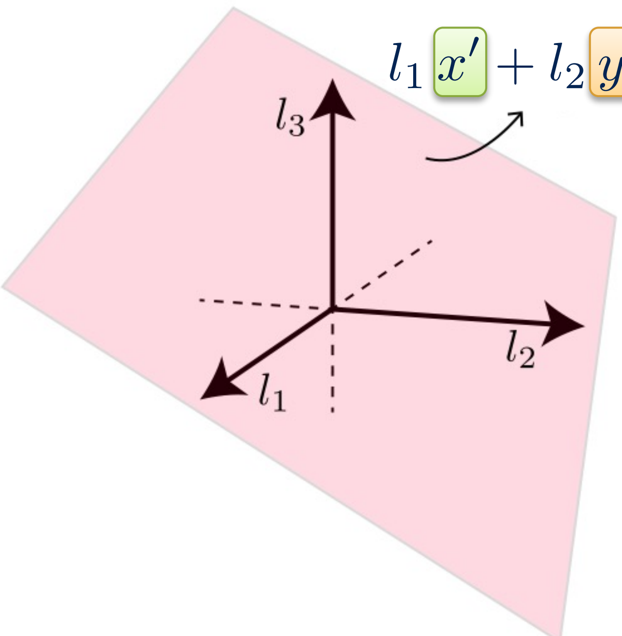
Source Localization (1/2)

Review of the **Projective Ray Space representation** of a **point source**

Geometric Space

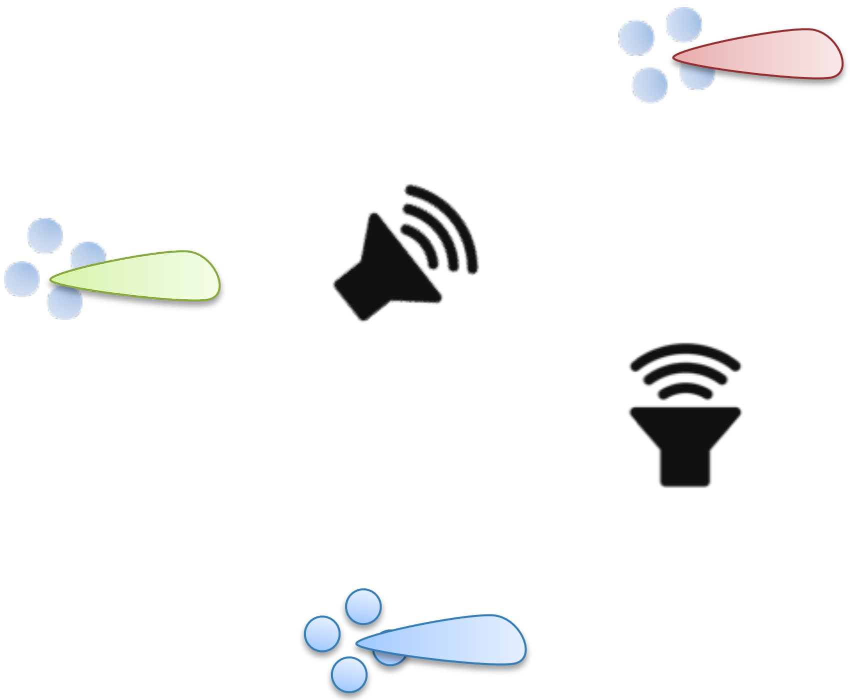
Projective Ray Space

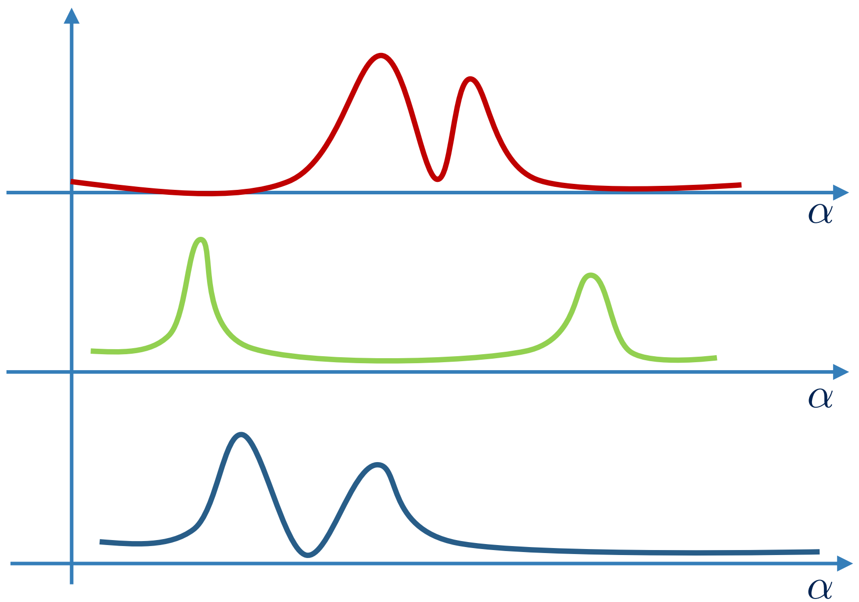




$$l_1 \boxed{x'} + l_2 \boxed{y'} + l_3 = 0$$

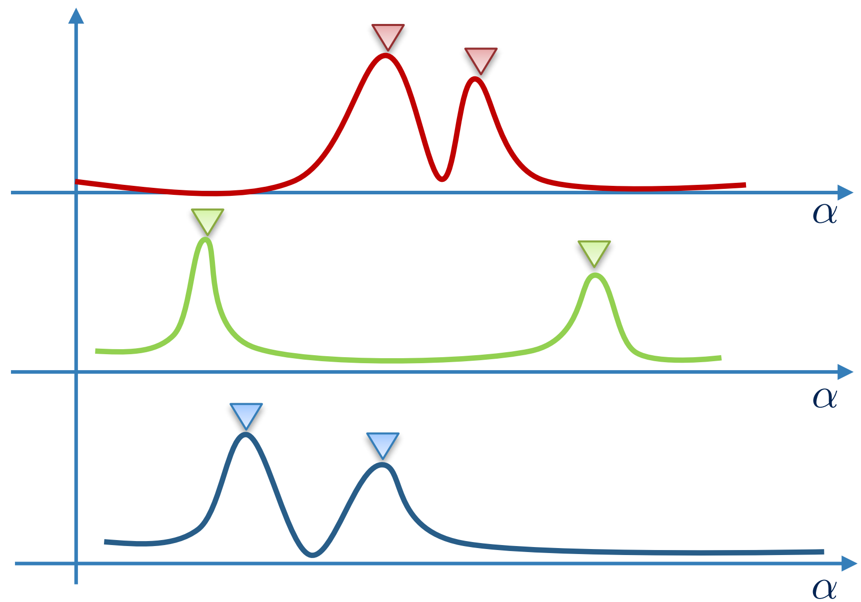
Source Localization (2/2)





Source Localization (2/2)

Peak the maxima



Source Localization (2/2)

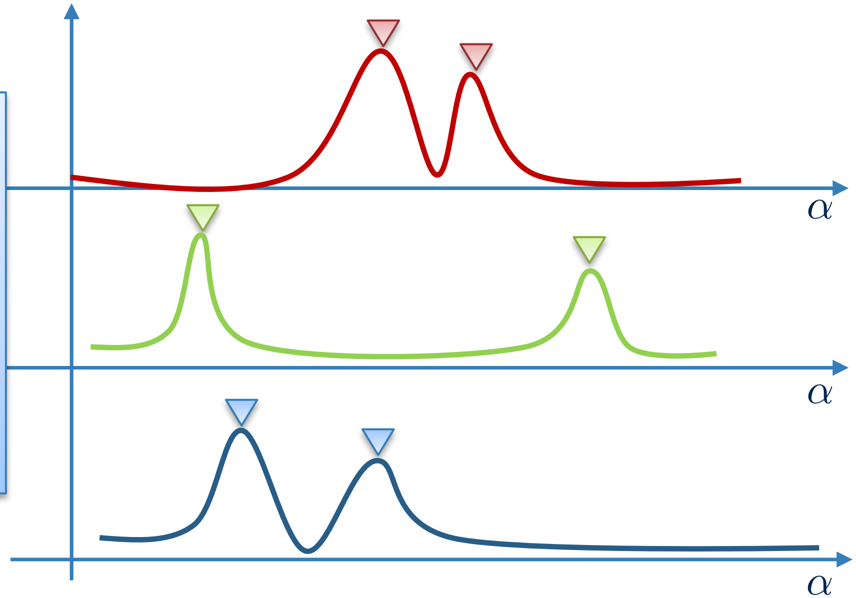
Peak the **maxima**

Map to the **Projective Ray Space**

$$l_{1,n}^{(a)} = \gamma \sin(\hat{\alpha}_n^{(a)});$$

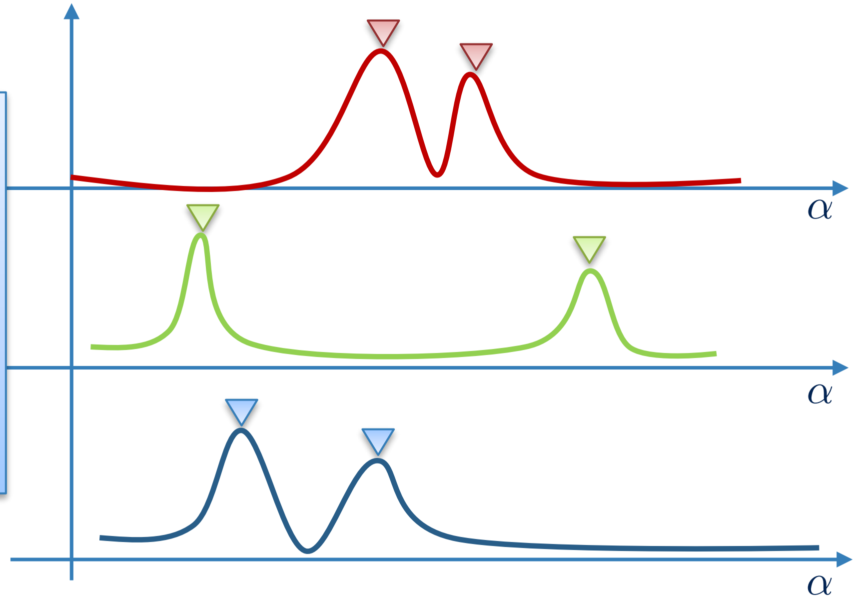
$$l_{2,n}^{(a)} = -\gamma \cos(\hat{\alpha}_n^{(a)});$$

$$l_{3,n}^{(a)} = \gamma[y^a \cos(\hat{\alpha}_n^{(a)}) - x^a \sin(\hat{\alpha}_n^{(a)})]$$



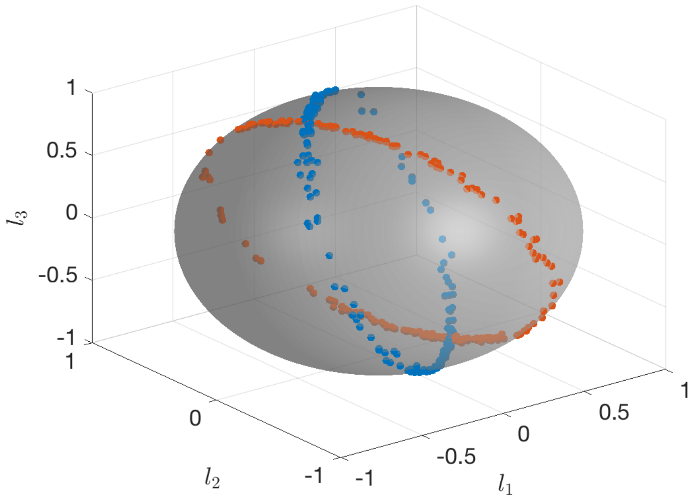
Source Localization (2/2)

Peak the **maxima**



Map to the **Projective Ray Space**

$$l_{1,n}^{(a)} = \gamma \sin(\hat{\alpha}_n^{(a)});$$
$$l_{2,n}^{(a)} = -\gamma \cos(\hat{\alpha}_n^{(a)});$$
$$l_{3,n}^{(a)} = \gamma[y^a \cos(\hat{\alpha}_n^{(a)}) - x^a \sin(\hat{\alpha}_n^{(a)})]$$

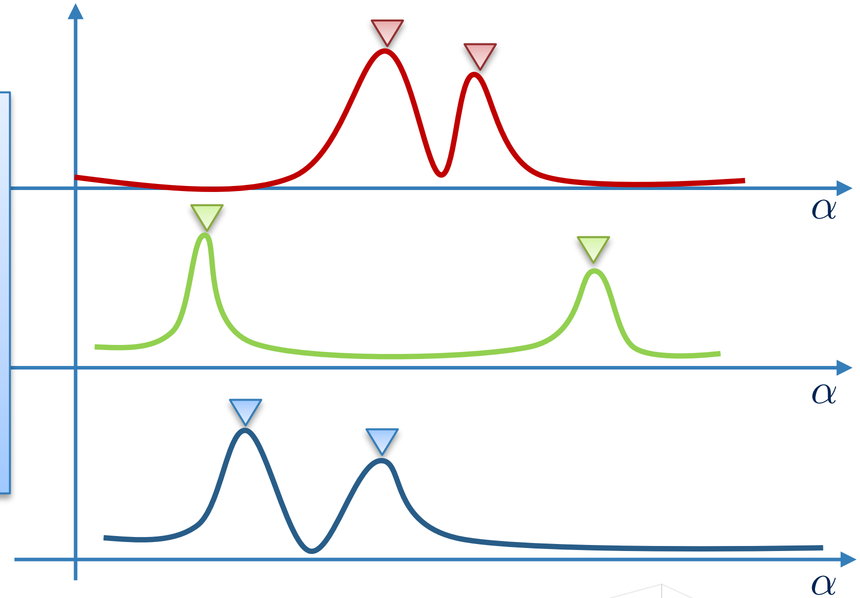


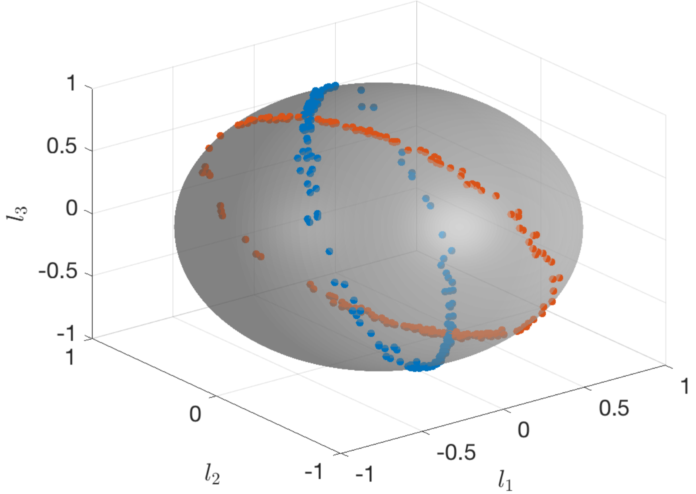
Source Localization (2/2)

Peak the **maxima**

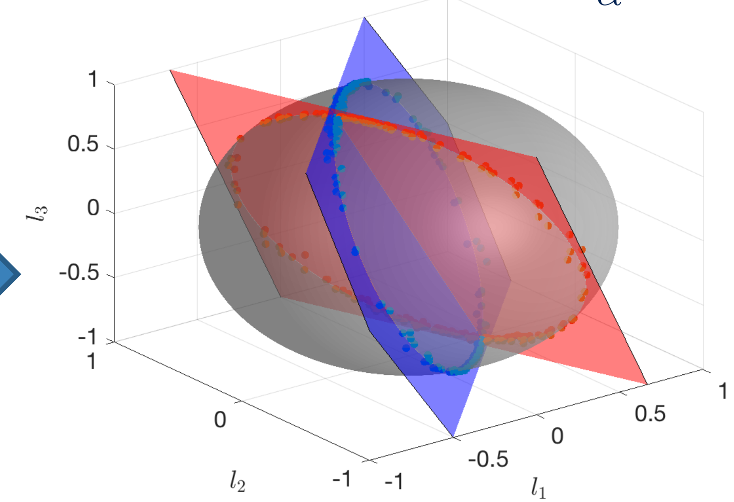
Map to the **Projective Ray Space**

$$l_{1,n}^{(a)} = \gamma \sin(\hat{\alpha}_n^{(a)});$$
$$l_{2,n}^{(a)} = -\gamma \cos(\hat{\alpha}_n^{(a)});$$
$$l_{3,n}^{(a)} = \gamma[y^a \cos(\hat{\alpha}_n^{(a)}) - x^a \sin(\hat{\alpha}_n^{(a)})]$$

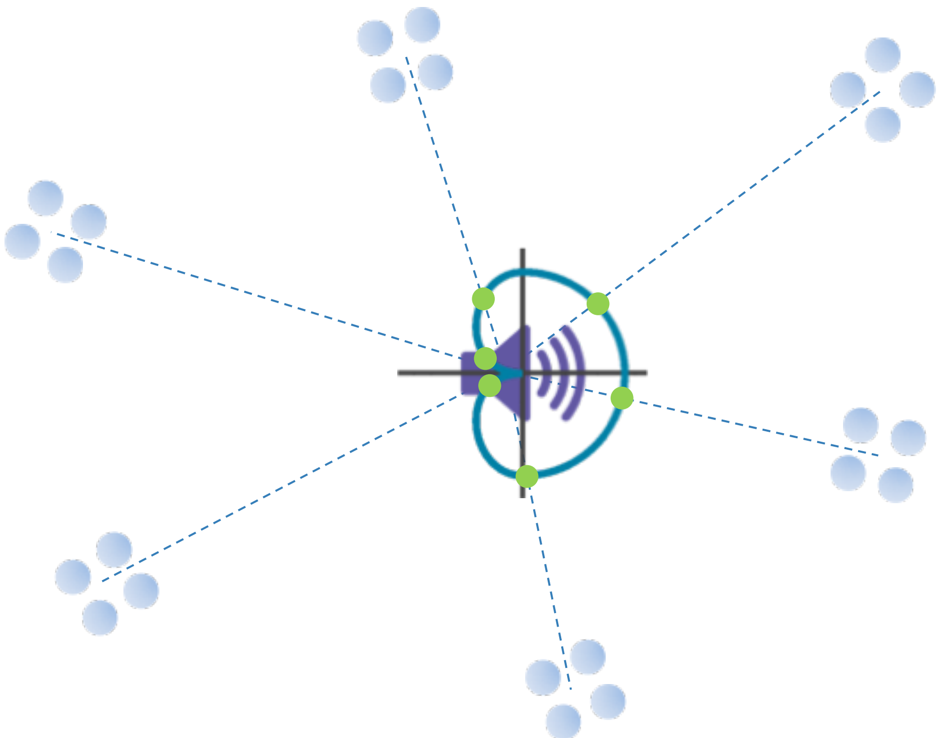


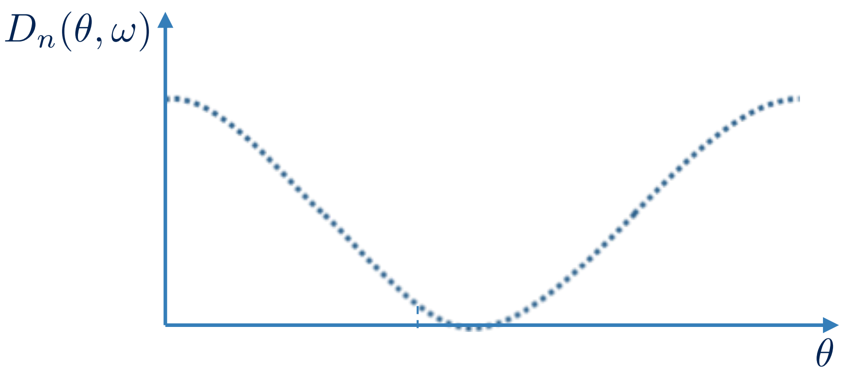


Clustering and linear regression (RANSAC)

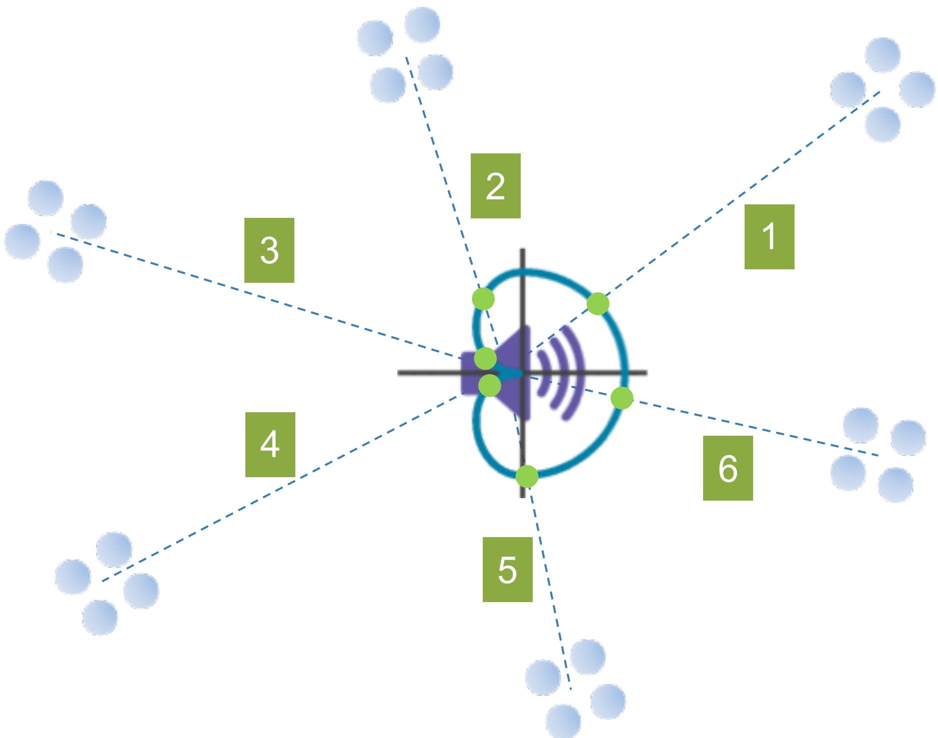


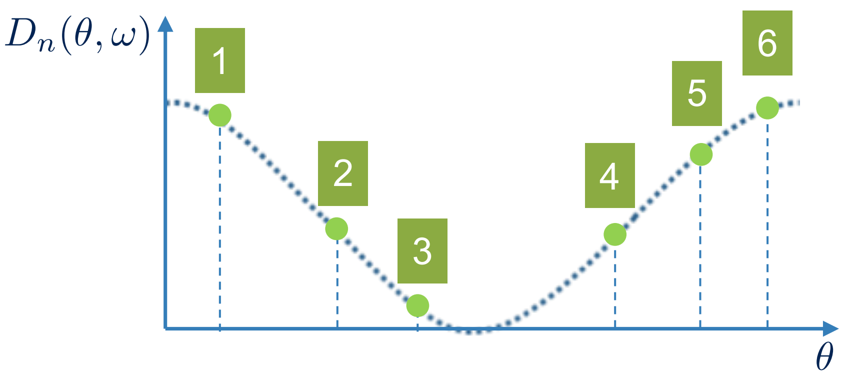
Radiance Pattern Estimation (1/3)



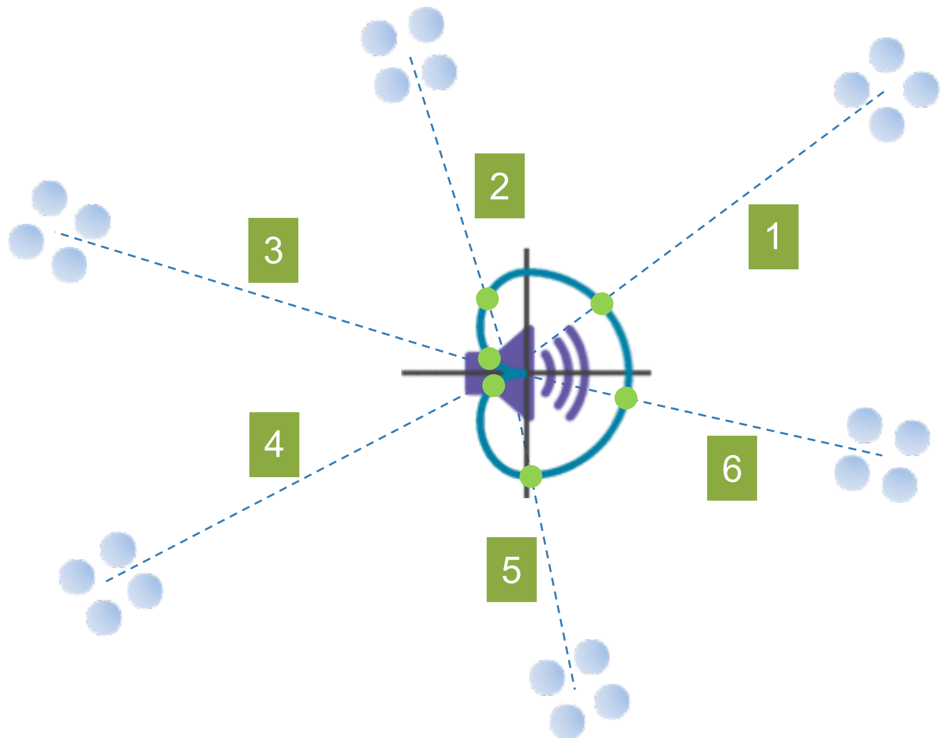


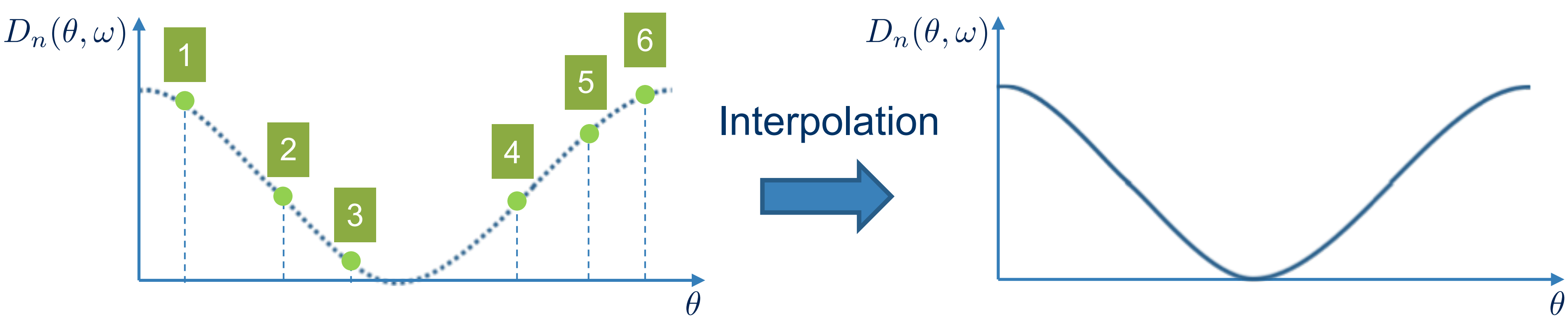
Radiance Pattern Estimation (1/3)





Radiance Pattern Estimation (1/3)





Radiance Pattern Estimation (2/3)

Estimate of the Radiance Pattern at the microphone arrays

Assumption: sources in the far field w.r.t to each array

⇩

Microphones within each array "see" the *n*th source under the same angle

LCMV beamformer in order to estimate the radiance pattern value

$$\mathbf{h}_n^{(a)} = \underset{\mathbf{h}}{\operatorname{argmin}} \ \mathbf{h}^H \mathbf{h} \qquad \text{s.t.} \ \mathbf{h}^H \hat{\mathbf{G}}^{(a)} = \mathbf{c}$$

Radiance Pattern Estimation (2/3)

Estimate of the Radiance Pattern at the microphone arrays

Assumption: sources in the far field w.r.t to each array

⇩

Microphones within each array "see" the *n*th source under the same angle

LCMV beamformer in order to estimate the radiance pattern value

$$\mathbf{h}_n^{(a)} = \underset{\mathbf{h}}{\operatorname{argmin}} \ \mathbf{h}^H \mathbf{h} \qquad \text{s.t.} \ \ \mathbf{h}^H \boxed{\hat{\mathbf{G}}^{(a)}} = \boxed{\mathbf{c}} \rightarrow \begin{cases} 1 & \text{For the nth source} \\ 0 & \text{Otherwise} \end{cases}$$

Green's functions of the sources obtained after the localization

$$\hat{D}_n\left(\theta_n^{(a)}\right) = \mathbf{h}_n^{(a)H} \mathbf{p}^{(a)}$$

Radiance Pattern Estimation (3/3)

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$D_n(\theta_n^{(a)}) = \sum_{l=0}^{L-1} w_{n,l} \cos(l\theta_n^{(a)}) + r_{n,l} \sin(l\theta_n^{(a)})$$

Radiance Pattern Estimation (3/3)

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$D_n(\theta_n^{(a)}) = \sum_{l=0}^{L-1} w_{n,l} \cos(l\theta_n^{(a)}) + r_{n,l} \sin(l\theta_n^{(a)})$$

$$\hat{\mathbf{y}}_n = \underset{\mathbf{y}_n}{\operatorname{argmin}} \; \|\mathbf{q}_n - \mathbf{A}_n \mathbf{y}_n\|^2 \qquad \text{s.t. } \mathbf{F}\,\mathbf{y}_n \geq \mathbf{0}$$

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$\boxed{D_n(\theta_n^{(a)})} = \sum_{l=0}^{L-1} w_{n,l} \cos(l\theta_n^{(a)}) + r_{n,l} \sin(l\theta_n^{(a)})$$

$$\hat{\mathbf{y}}_n = \underset{\mathbf{y}_n}{\operatorname{argmin}} \; \|\boxed{\mathbf{q}_n} - \mathbf{A}_n \, \mathbf{y}_n\|^2 \qquad \text{s.t. } \; \mathbf{F} \, \mathbf{y}_n \geq \mathbf{0}$$

Radiance Pattern Estimation (3/3)

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$\boxed{D_n(\theta_n^{(a)})} = \sum_{l=0}^{L-1} \boxed{w_{n,l}} \cos(l\theta_n^{(a)}) + \boxed{r_{n,l}} \sin(l\theta_n^{(a)})$$

$$\hat{\mathbf{y}}_n = \underset{\mathbf{y}_n}{\operatorname{argmin}} \ \|\boxed{\mathbf{q}_n} - \mathbf{A}_n \boxed{\mathbf{y}_n}\|^2 \qquad \text{s.t.} \ \ \mathbf{F}\boxed{\mathbf{y}_n} \geq \mathbf{0}$$

Radiance Pattern Estimation (3/3)

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$\boxed{D_n(\theta_n^{(a)})} = \sum_{l=0}^{L-1} \boxed{w_{n,l}} \boxed{\cos(l\theta_n^{(a)})} + \boxed{r_{n,l}} \boxed{\sin(l\theta_n^{(a)})}$$

$$\hat{\mathbf{y}}_n = \underset{\mathbf{y}_n}{\operatorname{argmin}} \; \|\boxed{\mathbf{q}_n} - \boxed{\mathbf{A}_n}\boxed{\mathbf{y}_n}\|^2 \qquad \text{s.t.} \;\; \mathbf{F}\boxed{\mathbf{y}_n} \geq \mathbf{0}$$

Radiance Pattern Estimation (3/3)

Interpolation of the Radiance Pattern

Idea: adopt a **model** to reconstruct the whole radiance pattern from the estimates at a **limited set of angles**

$$D_n(\theta_n^{(a)}) = \sum_{l=0}^{L-1} w_{n,l} \cos(l\theta_n^{(a)}) + r_{n,l} \sin(l\theta_n^{(a)})$$

$$\boxed{\hat{\mathbf{y}}_n} = \underset{\mathbf{y}_n}{\operatorname{argmin}} \; \|\mathbf{q}_n - \mathbf{A}_n \mathbf{y}_n\|^2 \quad \text{s.t.} \; \boxed{\mathbf{F}\mathbf{y}_n \geq \mathbf{0}}$$

Estimate of the nth source Radiation Pattern Coefficients

Constrain the Radiance Pattern to be positive

Signal Reconstruction

We exploit the estimate of the **radiance patterns** and the **source positions** to **design** an **informed spatial filter**

Signal Reconstruction

We exploit the estimate of the **radiance patterns** and the **source positions** to **design** an **informed spatial filter**

$$\mathbf{u}_n^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \ \mathbf{u}^H \mathbf{u} \qquad \text{s.t.} \ \mathbf{u}^H \mathbf{Y} = \mathbf{d}$$

Signal Reconstruction

We exploit the estimate of the **radiance patterns** and the **source positions** to **design** an **informed spatial filter**

$$\mathbf{u}_n^\star = \underset{\mathbf{u}}{\operatorname{argmin}}\ \mathbf{u}^H\mathbf{u} \qquad \text{s.t.}\ \mathbf{u}^H \boxed{\mathbf{Y}} = \boxed{\mathbf{d}}$$

$\mathbf{d}$: 1 For the nth source; 0 Otherwise

$$\mathbf{Y} = \left[\underbrace{\left(\hat{\mathbf{G}}^{(0)} \otimes \hat{\mathbf{D}}^{(0)}\right)^T, \dots, \left(\hat{\mathbf{G}}^{(A-1)} \otimes \hat{\mathbf{D}}^{(A-1)}\right)^T}_{\text{Estimates of source Radiation Patterns and Positions}}\right]^T$$

Estimates of source Radiation Patterns and Positions

Signal Reconstruction

We exploit the estimate of the **radiance patterns** and the **source positions** to **design** an **informed spatial filter**

$$\mathbf{u}_n^\star = \underset{\mathbf{u}}{\operatorname{argmin}} \ \mathbf{u}^H\mathbf{u} \qquad \text{s.t. } \mathbf{u}^H \boxed{\mathbf{Y}} = \boxed{\mathbf{d}}$$

$\mathbf{d}$: 1 — For the nth source; 0 — Otherwise

$$\mathbf{Y} = \Big[\underbrace{\left(\hat{\mathbf{G}}^{(0)} \otimes \hat{\mathbf{D}}^{(0)}\right)^T, \ldots, \left(\hat{\mathbf{G}}^{(A-1)} \otimes \hat{\mathbf{D}}^{(A-1)}\right)^T}_{\text{Estimates of source Radiation Patterns and Positions}} \Big]^T$$

$$\hat{S}_n(\omega) = \left(\mathbf{u}_n^\star\right)^H \mathbf{p}$$

Sound Synthesis

Once the **parameters are estimated** as described in the previous sections, the **sound field at the virtual microphone** location can **be computed through the parametric model**

$$\hat{P}(\mathbf{r}_{VM},\omega) = \sum_{n=0}^{N-1} \hat{D}_n(\theta_{VM,n},\omega)\, g(\mathbf{r}_{VM},\hat{\mathbf{r}'}_n,\omega)\, \hat{S}_n(\omega)$$

Sound Synthesis

Once the **parameters are estimated** as described in the previous sections, the **sound field at the virtual microphone** location can **be computed through the parametric model**

$$\underbrace{\hat{P}(\mathbf{r}_{VM},\omega)}_{\text{Virtual Miking signal estimate}} = \sum_{n=0}^{N-1} \underbrace{\hat{D}_n(\theta_{VM,n},\omega)}_{\text{Radiation Pattern estimate}} \underbrace{g(\mathbf{r}_{VM},\hat{\mathbf{r}}'_n,\omega)}_{\text{Source Position estimate}} \underbrace{\hat{S}_n(\omega)}_{\text{Source Signal estimate}}$$

Simulations and Experiments Results (1/3)

Metrics

Localization Error

$$\mathrm{LE}(\mathbf{r}') = \frac{1}{N}\sum_{n=0}^{N-1} \|\hat{\mathbf{r}}'_n - \mathbf{r}'_n\|$$

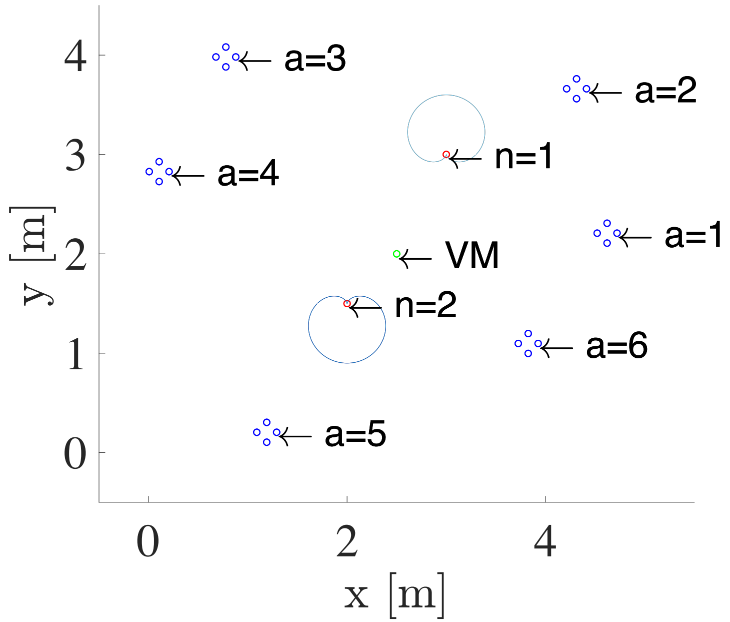
Directivity Error

$$\mathrm{DE}_n = \frac{1}{F}\sum_{f=0}^{F-1}\frac{1}{I}\sum_{i=0}^{I-1}(\hat{D}_n(\theta_i,\omega_f) - D_n(\theta_i,\omega_f))^2$$

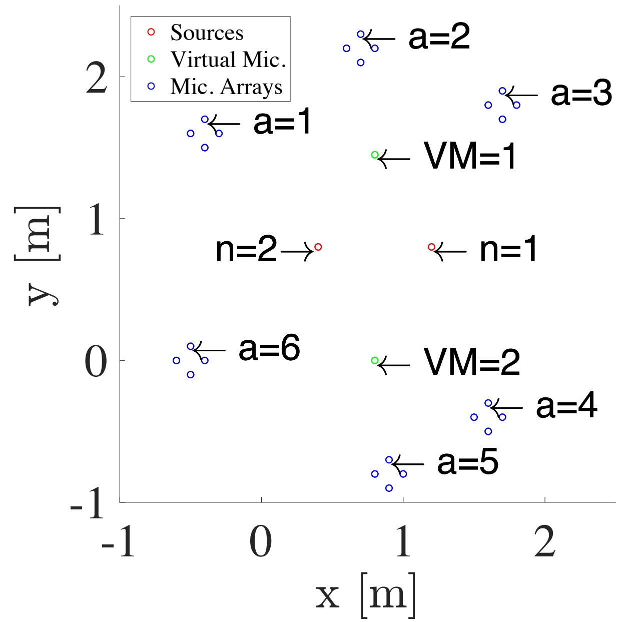
Synthesized Signal Error

$$\mathrm{SSE}_{VM} = \sqrt{\frac{1}{T}\sum_{t=0}^{T}(x_{VM}(t) - x_{ref}(t))^2}$$

Simulations Setup



Experiments Setup



Simulations and Experiments Results (2/3)

Simulations Results

Simulations					
Sources	Orientation	Source Signal	LE [m]	DE	SSE
1	$0°$	White Noise	1.12×10^{-2}	6.61×10^{-3}	6.74×10^{-4}
2	$0°$		7.40×10^{-3}	2.66×10^{-2}	5.41×10^{-4}
1, 2	$0°, 0°$		3.50×10^{-2}	$8.56 \times 10^{-2}, 5.24 \times 10^{-2}$	2.42×10^{-2}
1	$-90°$	Speech	1.10×10^{-2}	3.57×10^{-3}	8.18×10^{-4}
2	$90°$		8.00×10^{-3}	1.12×10^{-2}	7.72×10^{-4}
1, 2	$-90°, 0°$		4.39×10^{-2}	$8.97 \times 10^{-2}, 9.94 \times 10^{-2}$	5.75×10^{-3}

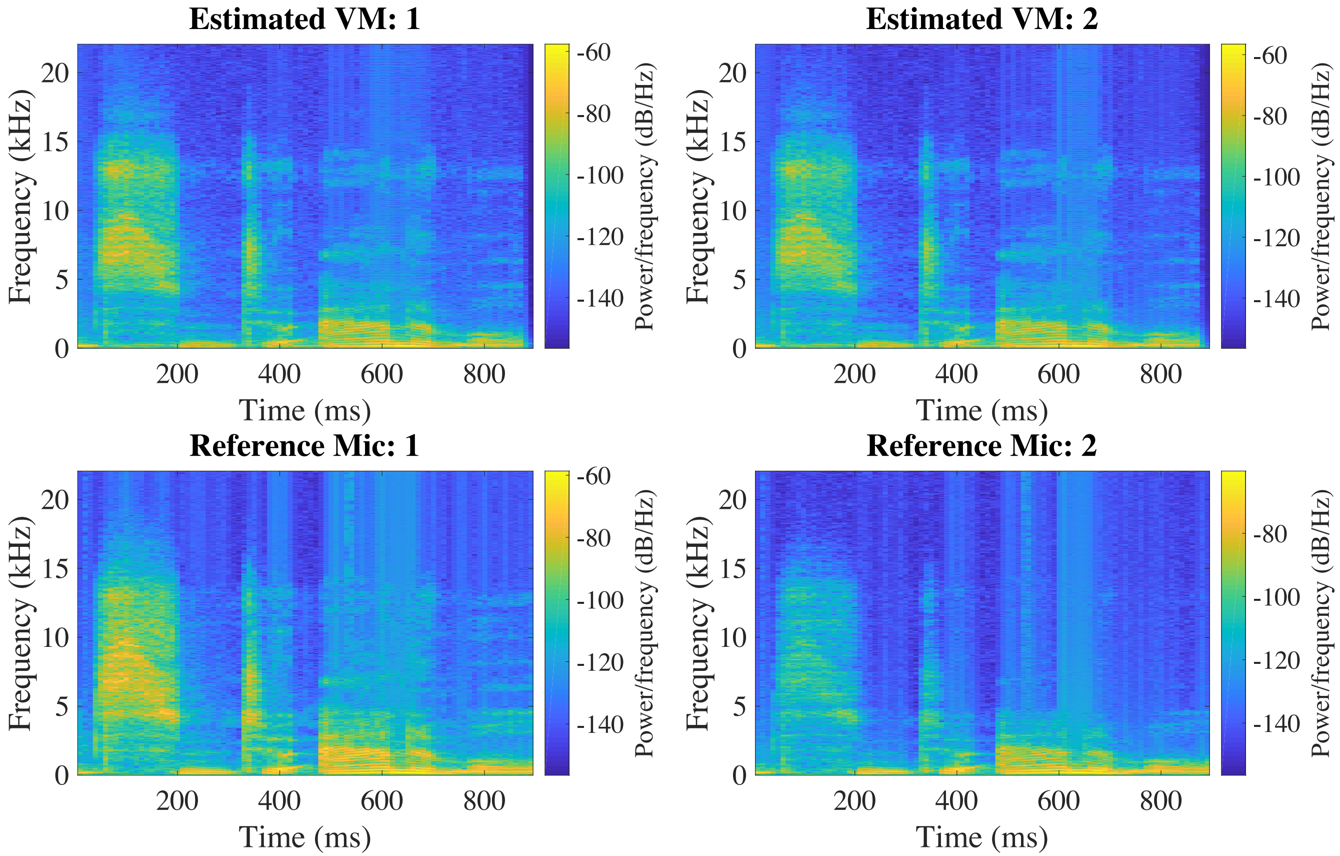
Experiments Results

Experiments				
Sources	Orientation	Source Signal	LE [m]	SSE
1	$90°$	White Noise	3.54×10^{-2}	$1.00 \times 10^{-2}, 2.84 \times 10^{-3}$
2	$-90°$		5.33×10^{-2}	$3.44 \times 10^{-3}, 9.30 \times 10^{-3}$
1, 2	$90°, -90°$		6.40×10^{-2}	$1.32 \times 10^{-2}, 1.29 \times 10^{-2}$
1	$90°$	Speech	6.03×10^{-2}	$5.83 \times 10^{-3}, 4.26 \times 10^{-3}$
2	$90°$		3.18×10^{-2}	$9.72 \times 10^{-3}, 6.98 \times 10^{-3}$
1, 2	$90°, 90°$		5.53×10^{-2}	$1.34 \times 10^{-2}, 1.00 \times 10^{-2}$

Simulations and Experiments Results (3/3)

Experiments Results

Example of a **spectrogram** of the **VM** signal w.r.t to the two **reference microphones**



Conclusions

- We proposed a parametric technique for sound field reconstruction in the presence of multiple sources. The parameters are
- Sources locations;
- Radiance pattern of the sources;
- Sources dry signals

and they are estimated directly from the audio data.

Results demonstrate that the approach is promising even with some errors in the estimation of the parameters.

Future Works

- Include a directional sensitivity pattern of the virtual microphone.
- Extend the presented technique also to reverberant environments:
  - Estimate the parameters robustly with respect to the reverberation
  - Synthesize the sound field including the direct and diffuse components of the sound field.