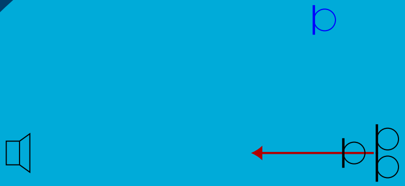
Steered Response Power-Based Direction-of-Arrival Estimation Exploiting an Auxiliary Microphone

EUSIPCO 2024, Lyon

Klaus Brümann, Simon Doclo,

Carl von Ossietzky Universität Oldenburg

27.08.2024



Introduction

- Steered response power (SRP) method commonly used for estimating direction-of-arrival (DOA) of speech source relative to compact microphone array

Introduction

- Steered response power (SRP) method commonly used for estimating direction-of-arrival (DOA) of speech source relative to compact microphone array
- DOA estimation depends on cross-power spectral densities (CPSDs), whose reliability is affected by noise and reverberation

Introduction

- Steered response power (SRP) method commonly used for estimating direction-of-arrival (DOA) of speech source relative to compact microphone array
- DOA estimation depends on cross-power spectral densities (CPSDs), whose reliability is affected by noise and reverberation
- Phase transform (PHAT) weighting improves DOA estimation accuracy

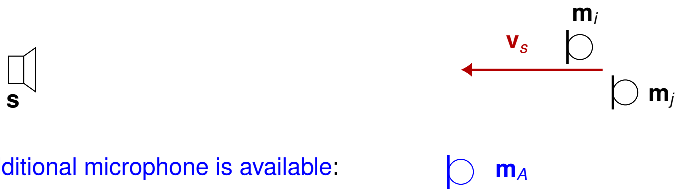
Introduction

- Steered response power (SRP) method commonly used for estimating direction-of-arrival (DOA) of speech source relative to compact microphone array
- DOA estimation depends on cross-power spectral densities (CPSDs), whose reliability is affected by noise and reverberation
- Phase transform (PHAT) weighting improves DOA estimation accuracy

Problem

- For closely spaced microphones, spatial coherence of noise & reverberation typically high
- CPSDs negatively affected **over extended frequency range**
- DOA estimation errors can occur despite PHAT weighting

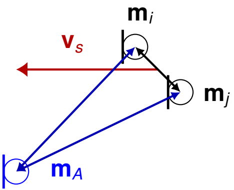
Introduction



Assuming additional microphone is available:

- position is unknown
- spatially separated from compact microphone array

Introduction



Assuming additional microphone is available:

- position is unknown
- spatially separated from compact microphone array

Proposed Idea

Compute more reliable CPSDs between $\mathbf{m}_i$ and $\mathbf{m}_j$ based on:

- CPSDs between $\mathbf{m}_i$ and $\mathbf{m}_A$
- CPSDs between $\mathbf{m}_A$ and $\mathbf{m}_j$

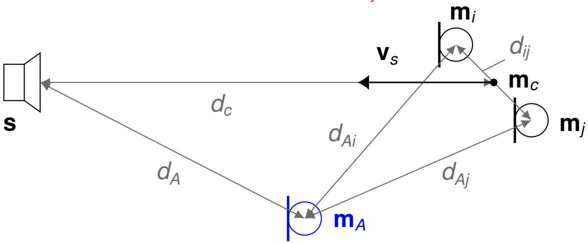
$\Rightarrow$ exploit **spatial separation** between $\mathbf{m}_A$ and compact microphone array

Carl von Ossietzky
Universität
Oldenburg

Signal Model and Acoustic Scenario

Signal Model

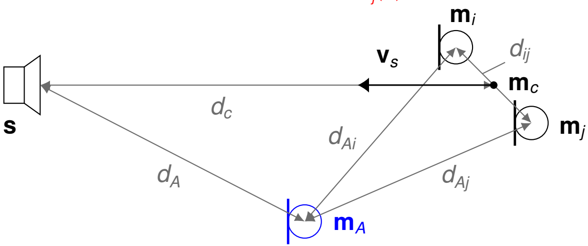
$$Y_j(\omega) = X_j^D(\omega) + \underbrace{X_j^R(\omega) + N_j(\omega)}_{U_j(\omega)}$$



Signal Model and Acoustic Scenario

Signal Model

$$Y_j(\omega) = X_j^D(\omega) + \underbrace{X_j^R(\omega) + N_j(\omega)}_{U_j(\omega)}$$



Direct component in *j*-th microphone

$$X_j^D(\omega) = \frac{1}{4\pi d_j} \exp(-\jmath\omega d_j/\nu) S(\omega)$$
$$= X_i^D(\omega) \exp(\jmath\omega\tau_{ij}(\mathbf{v}_s))$$

with (far field) TDOA $\tau_{ij}(\mathbf{v}_s) = \mathbf{v}_s^T(\mathbf{m}_j - \mathbf{m}_i)/\nu$

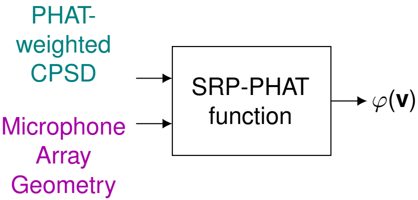
SRP-Based DOA Estimation

SRP-PHAT Function with candidate DOA vector **v**

$$\varphi(\mathbf{v}) = \sum_{(i,j)\,:\,i>j} \int_{-\infty}^{\infty} \underbrace{\psi_{ij}(\omega)}_{\text{Signal Component}} \; \underbrace{\exp(\jmath\omega\tau_{ij}(\mathbf{v}))}_{\text{Model Component}} \; d\omega$$

PHAT-weighted CPSD

$$\psi_{ij}(\omega) = \frac{\mathbb{E}\{Y_i(\omega)\,Y_j^*(\omega)\}}{\left|\underbrace{\mathbb{E}\{Y_i(\omega)\,Y_j^*(\omega)\}}_{\tilde{\psi}_{ij}(\omega)}\right|}$$



PHAT-weighted CPSD

Microphone Array Geometry

SRP-PHAT function

$\varphi(\mathbf{v})$

SRP-Based DOA Estimation

SRP-PHAT Function with candidate DOA vector **v**

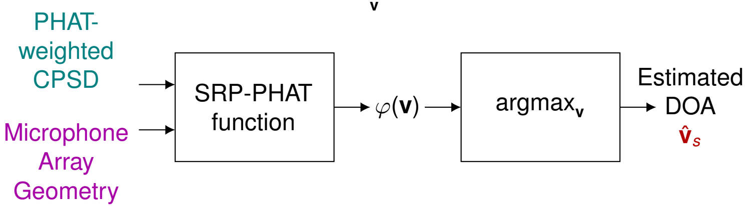
$$\varphi(\mathbf{v}) = \sum_{(i,j)\,:\,i>j} \int_{-\infty}^{\infty} \underbrace{\psi_{ij}(\omega)}_{\text{Signal Component}} \; \underbrace{\exp(\jmath\omega\tau_{ij}(\mathbf{v}))}_{\text{Model Component}} \, d\omega$$

PHAT-weighted CPSD

$$\psi_{ij}(\omega) = \frac{\mathbb{E}\{Y_i(\omega)\,Y_j^*(\omega)\}}{\left|\underbrace{\mathbb{E}\{Y_i(\omega)\,Y_j^*(\omega)\}}_{\tilde{\psi}_{ij}(\omega)}\right|}$$

Estimate DOA

$$\hat{\mathbf{v}}_s = \underset{\mathbf{v}}{\operatorname{argmax}} \; \varphi(\mathbf{v})$$



PHAT-weighted CPSD

Microphone Array Geometry

SRP-PHAT function

$\varphi(\mathbf{v})$

$\operatorname{argmax}_{\mathbf{v}}$

Estimated DOA $\hat{\mathbf{v}}_s$

PHAT-Weighted CPSD

Assuming

- mutually uncorrelated direct and undesired (noise and reverberation) components

PHAT-weighted CPSD:

$$\psi_{ij}(\omega) = \frac{\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)}{\left|\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)\right|}$$

PHAT-Weighted CPSD

Assuming

- mutually uncorrelated direct and undesired (noise and reverberation) components
- spherically diffuse undesired component

PHAT-weighted CPSD:

$$\psi_{ij}(\omega) = \frac{\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)}{|\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)|}$$

$$= \frac{\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\,\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_S)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})}{|\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\,\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_S)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})|}$$

PHAT-Weighted CPSD

Assuming

- mutually uncorrelated direct and undesired (noise and reverberation) components
- spherically diffuse undesired component

PHAT-weighted CPSD:

$$\psi_{ij}(\omega) = \frac{\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)}{|\tilde{\psi}_{ij}^{D}(\omega) + \tilde{\psi}_{ij}^{U}(\omega)|}$$

$$= \frac{\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\,\mathrm{exp}(-\jmath\omega\tau_{ij}(\mathbf{v}_S)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})}{|\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\,\mathrm{exp}(-\jmath\omega\tau_{ij}(\mathbf{v}_S)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})|}$$

For small inter-microphone distances $d_{ij} \Rightarrow \mathrm{sinc}(\frac{\omega d_{ij}}{\nu})$ affects many frequencies

PHAT-Weighted CPSD Using Auxiliary Microphone

Assuming

- mutually uncorrelated direct and undesired (noise and reverberation) components
- spherically diffuse undesired component

PHAT-weighted CPSD:

$$\psi_{ij}(\omega) = \frac{\tilde{\psi}^{D}_{ij}(\omega) + \tilde{\psi}^{U}_{ij}(\omega)}{|\tilde{\psi}^{D}_{ij}(\omega) + \tilde{\psi}^{U}_{ij}(\omega)|}$$

$$= \frac{\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})}{|\frac{\phi_S(\omega)}{16\pi^2 d_c^2}\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \phi_U(\omega)\,\mathrm{sinc}(\frac{\omega d_{ij}}{\nu})|}$$

For small inter-microphone distances $d_{ij} \Rightarrow \mathrm{sinc}(\frac{\omega d_{ij}}{\nu})$ affects many frequencies

Proposed Method

Since d_{Ai} and d_{Aj} typically larger than d_{ij}, we propose to compute CPSD based on auxiliary microphone

$$\boxed{\psi^{A}_{ij}(\omega) = \psi_{iA}(\omega)\,\psi_{Aj}(\omega)}$$

Distortion of CPSD

Without Auxiliary Microphone

$$\psi_{ij}(\omega) = \frac{\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \overbrace{\frac{16\pi^2 d_c^2}{\mathrm{SUR}(\omega)}\,\mathrm{sinc}\left(\frac{\omega d_{ij}}{\nu}\right)}^{D_{ij}(\omega)}}{\left|\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \frac{16\pi^2 d_c^2}{\mathrm{SUR}(\omega)}\,\mathrm{sinc}\left(\frac{\omega d_{ij}}{\nu}\right)\right|}$$

- source-to-undesired ratio $\mathrm{SUR}(\omega) = \frac{\phi_S(\omega)}{\phi_U(\omega)}$ (defined at source position).
- Distortion term $D_{ij}(\omega)$

Distortion of CPSD

Without Auxiliary Microphone

$$\psi_{ij}(\omega) = \frac{\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \overbrace{\frac{16\pi^2 d_c^2}{\mathrm{SUR}(\omega)}\,\mathrm{sinc}\left(\frac{\omega d_{ij}}{\nu}\right)}^{D_{ij}(\omega)}}{\left|\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + \frac{16\pi^2 d_c^2}{\mathrm{SUR}(\omega)}\,\mathrm{sinc}\left(\frac{\omega d_{ij}}{\nu}\right)\right|}$$

- source-to-undesired ratio $\mathrm{SUR}(\omega) = \frac{\phi_S(\omega)}{\phi_U(\omega)}$ (defined at source position).
- Distortion term $D_{ij}(\omega)$

With Auxiliary Microphone

$$\psi_{ij}^{\mathrm{A}}(\omega) = \frac{\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + D_{ij}^{\mathrm{A}}(\omega)}{\left|\exp(-\jmath\omega\tau_{ij}(\mathbf{v}_s)) + D_{ij}^{\mathrm{A}}(\omega)\right|}$$

$$D_{ij}^{\mathrm{A}}(\omega) = \frac{(4\pi)^2 d_c d_A}{\mathrm{SUR}(\omega)}\left[\mathrm{sinc}(\omega d_{Ai}/\nu)\exp(-\jmath\omega\tau_{Aj}(\mathbf{v}_s)) + \mathrm{sinc}(\omega d_{Aj}/\nu)\exp(-\jmath\omega\tau_{iA}(\mathbf{v}_s))\right] + \frac{(4\pi)^4 d_c^2 d_A^2}{\mathrm{SUR}^2(\omega)}\ \mathrm{sinc}(\omega d_{Ai}/\nu)\ \mathrm{sinc}(\omega d_{Aj}/\nu)$$

Analysis and Evaluation

Numerical Analysis - Based on Theoretical Equations

Proportion of frequencies for which $D^{A}_{ij}(\omega) \leq D_{ij}(\omega)$:

$$P = \frac{1}{\omega_0} \int_{0}^{\omega_0} \mathcal{H}\left(|D_{ij}(\omega)| - |D^{A}_{ij}(\omega)| \right) d\omega$$

($\mathcal{H}(x)$: Heaviside step function)

- average P over 18 array orientations ($M = 2$, $d_{12} = 5$ cm)
- SURs: 15, 10, 5, 0 dB

Analysis and Evaluation

Numerical Analysis - Based on Theoretical Equations

Proportion of frequencies for which $D_{ij}^{A}(\omega) \leq D_{ij}(\omega)$:

$$P = \frac{1}{\omega_0} \int_0^{\omega_0} \mathcal{H}\left(|D_{ij}(\omega)| - |D_{ij}^{A}(\omega)| \right) d\omega$$

($\mathcal{H}(x)$: Heaviside step function)

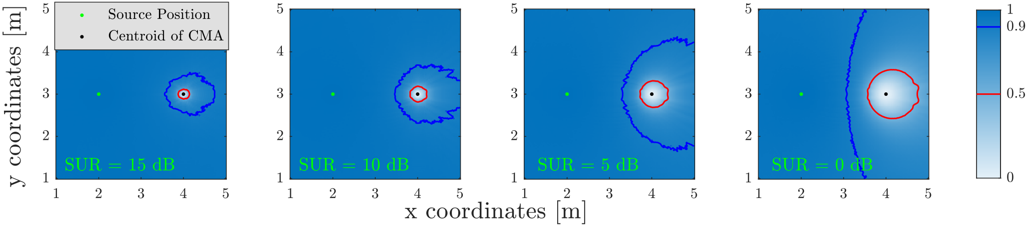
- average P over 18 array orientations ($M = 2$, $d_{12} = 5$ cm)
- SURs: 15, 10, 5, 0 dB

Evaluation - Based on Simulated Signals

DOA estimation error: $\varepsilon = \cos^{-1}\left(\frac{\hat{\mathbf{v}}_s^T \mathbf{v}_s}{||\hat{\mathbf{v}}_s||_2 \cdot ||\mathbf{v}_s||_2}\right)$

- average ε over 12 array orientations & 10 speech signals ($M = 3$, $d_{ij} = 5$ cm)
- Two reverberation and noise conditions (based on simulated RIRs and diffuse babble noise):
  1. DRR = RSNR = -1.4 dB (SUR $\approx$ 15 dB)
  2. DRR = RSNR = -7.2 dB (SUR $\approx$ 0 dB)
- 3 s signals, f_s = 16 kHz, STFT: 512 sample frame length (32 ms), 50% overlap

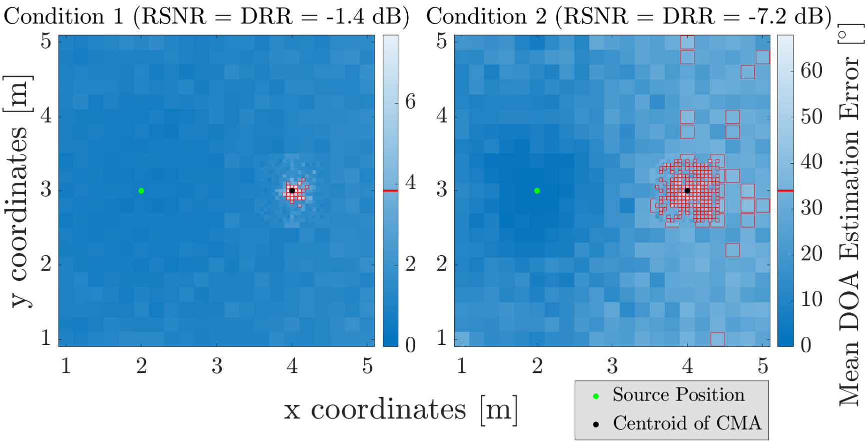
Analysis of Distortion



- Each point in 2D plane corresponds to a different auxiliary microphone position
- Auxiliary microphone positions outside of red circle ⇒ $D_{ij}^{A}(\omega) < D_{ij}(\omega)$ for more than 50% of frequencies

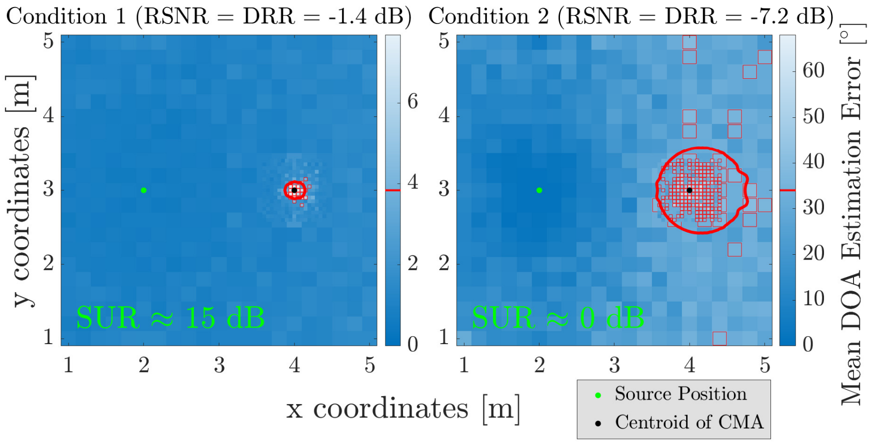
Auxiliary microphone is beneficial when spatially separated from compact microphone array.
For lower SUR, spatial separation needs to be larger

Signal-Based Evaluation



Auxiliary microphone brings benefit when spatially separated from compact microphone array
In more challenging condition, spatial separation needs to be larger

Signal-Based Evaluation



Auxiliary microphone brings benefit when spatially separated from compact microphone array
In more challenging condition, spatial separation needs to be larger
Simulation results correspond very well with theoretical analysis of distortion

Conclusions and Outlook

Conclusions:

- ✓ SRP-based DOA estimation can be significantly improved by incorporating auxiliary microphone which is spatially separated from compact microphone array
- ✓ Benefit of using auxiliary microphone can be predicted by theoretical analysis of distortion
  → required spatial separation mainly depends on source-to-undesired ratio

Outlook:

- Multiple auxiliary microphones
- Multiple sources