

Introduction

- Reverberant speech contains a direct and **reverberation** component
- Coherence** of the reverberation encodes **distances** between microphones and can be estimated using expectation conditional-maximization (ECM) [1]
- ECM requires **initial estimates** of direct and reverberation power spectral densities (PSDs), which can be estimated using a **data-independent coherence** matrix and a **matched beamformer** [2]

MAIN IDEA

Replace with a **data-dependent** coherence matrix estimate and a **minimum-power distortionless response** (MPDR) beamformer to reduce the MAG estimation error

Geometry Estimation

Reverberant Speech Signal Model (STFT-Domain)

$$\underbrace{\mathbf{x}[k, l]}_{\text{reverberant speech}} = \underbrace{\mathbf{g}[k]X_{d,1}[k, l]}_{\text{direct}} + \underbrace{\mathbf{x}_r[k, l]}_{\text{reverberation}} \quad \begin{array}{l} k: \text{freq. index} \\ l: \text{frame index} \end{array}$$

- $\mathbf{g}[k] = \exp(-j2\pi\tau k/K)$: Relative direct transfer function (RDTF) vector

Covariance Matrices

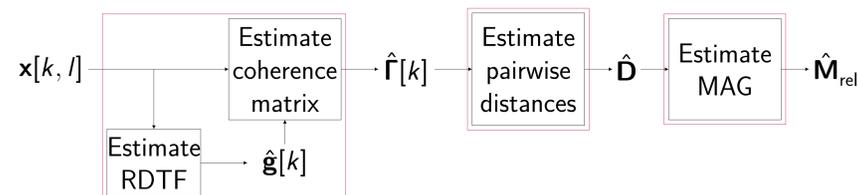
$$\underbrace{\Phi_x[k, l]}_{\text{cov. of recorded speech}} = \underbrace{\phi_d[k, l]\mathbf{g}[k]\mathbf{g}^H[k]}_{\text{cov. of direct speech } \Phi_d[k, l]} + \underbrace{\phi_r[k, l]\Gamma[k]}_{\text{cov. of reverberation } \Phi_r[k, l]}$$

- $\Gamma[k]$: coherence matrix of the reverberation
- $\phi_d[k, l]$: direct speech PSD
- $\phi_r[k, l]$: reverberation PSD

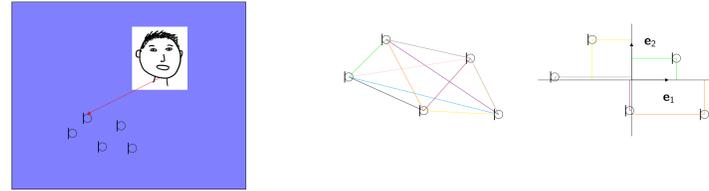
Assumptions:

- Direct and reverberation components uncorrelated
- Reverberation spatially homogeneous sound field

Overview:



- \mathbf{D} : matrix of squared pairwise distances between the microphones
- $\mathbf{M}_{\text{rel}} = [\mathbf{m}_{\text{rel},1}, \mathbf{m}_{\text{rel},2}, \dots, \mathbf{m}_{\text{rel},M}]$: MAG, i.e., relative microphone coordinates



- Estimate the coherence matrix of the reverberation $\hat{\Gamma}[k]$ using ECM [1, 2]
- Estimate the pairwise distances (PDs) between the microphones $d_{a,b}$ by comparing estimated coherence $\hat{\Gamma}_{a,b}[k]$ with a coherence model $\gamma(k, d_{a,b}) = \text{sinc}(\varepsilon k d_{a,b})$, for a specific range of frequencies [3, 2]
- Estimate the MAG $\hat{\mathbf{M}}_{\text{rel}}$ using multidimensional scaling [3, 4]

ECM Initialization

ECM requires **initial estimates** of the direct speech PSD $\hat{\phi}_d[k, l]$ and Reverberation PSD $\hat{\phi}_r[k, l]$. These are estimated using a beamformer $\mathbf{h}[k, l]$ and an initial coherence matrix estimate $\hat{\Gamma}^{(0)}[k]$ [5]

State-of-the-art [2]

$$\hat{\Gamma}_i^{(0)}[k] = \mathbf{I}, \quad \forall k$$

$$\mathbf{h}_{\text{MVDR}}[k] = \frac{\hat{\mathbf{g}}[k]}{\|\hat{\mathbf{g}}[k]\|^2}$$

- $\hat{\Gamma}_i^{(0)}[k]$ is **data-independent**
- Time-invariant
- matched beamformer**

Proposed

$$\hat{\Gamma}_x^{(0)}[k] = \frac{1}{L} \sum_{l=1}^L \frac{\mathbf{x}[k, l]\mathbf{x}^H[k, l]}{\frac{1}{M} \text{Tr}(\mathbf{x}[k, l]\mathbf{x}^H[k, l])}$$

$$\mathbf{h}_{\text{MPDR}}[k, l] = \frac{(\hat{\Phi}_x^{(0)})^{-1}[k, l]\hat{\mathbf{g}}[k]}{\hat{\mathbf{g}}^H[k](\hat{\Phi}_x^{(0)})^{-1}[k, l]\hat{\mathbf{g}}[k]}$$

- $\hat{\Gamma}_x^{(0)}[k]$ is **data-dependent**
- Time-varying
- MPDR beamformer**

Experimental Evaluation

Framework and Acoustical Parameters

- Simulated acoustic scenarios using RIR generator [6]
- 50 acoustic scenarios (random 5 s speech signal, array location & geometry, and speech source location) simulated per experiment
- $6 \times 6 \times 2.4$ m room with equally reflective walls and DRR = 0 dB (at ref. mic.)
- 32 ms frame length with 50% overlap between frames

Experiment 1:

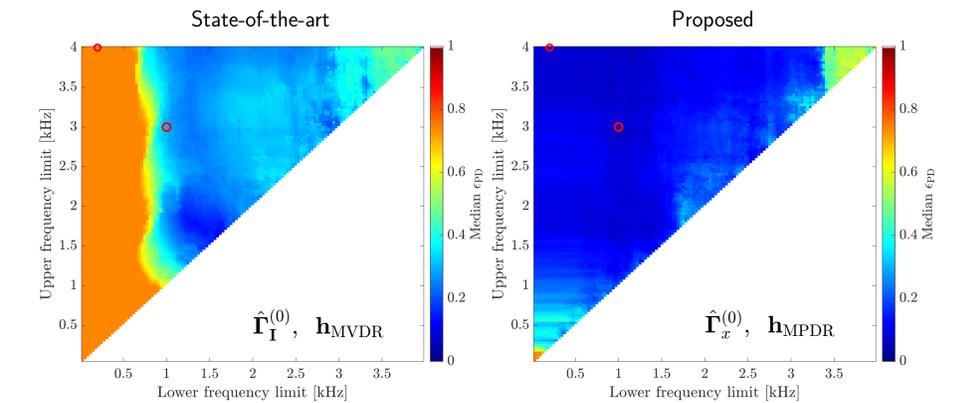
Analysis of the median PD error $\epsilon_{\text{PD}} = \frac{|\hat{d}_{1,2} - d_{1,2}|}{d_{1,2}}$
 $M = 2$ microphones $d_{1,2} = 10$ cm apart and frequency range varied between (0, 4] kHz

Experiment 2:

Analysis of the normalized MAG error $\epsilon_{\mathbf{m}_m} = \frac{\|\hat{\mathbf{m}}_m - \mathbf{m}_m\|_2}{\text{CL}}$
 $M = 6$ microphones randomly located within cubes with cube lengths (CLs) 10, 20, ... 50 cm

Results

Experiment 1:

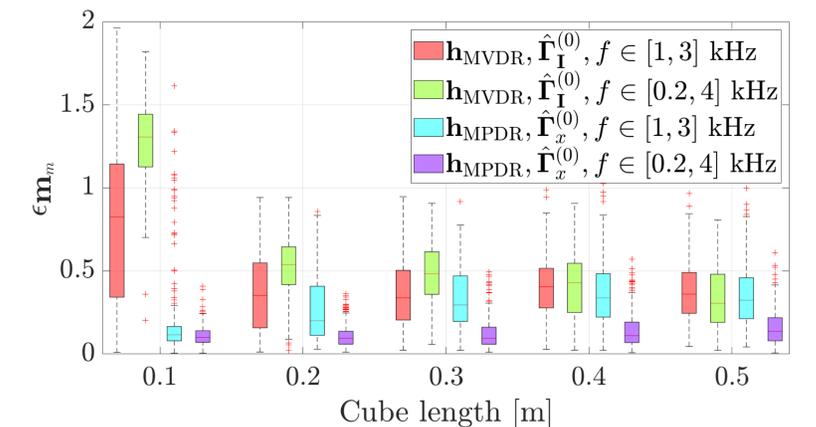


- Red circles indicate the considered frequency ranges in experiment 2

Conclusions

- Proposed initialization greatly reduces PD errors at low frequencies
- Proposed initialization allows a fixed lower frequency limit to be used which is **independent of the distances** between microphones

Experiment 2:



Conclusions

- Proposed initialization significantly reduces MAG errors for all array sizes
- MAG error is relatively **constant** for all considered array sizes when using proposed initialization

Future Work

- Comparison of estimation methods of the coherence matrix of the reverberation
- Integration/combination with blind MAG estimation approaches based on time-difference of arrivals (TDoAs)