A Simulation Study on Binaural Dereverberation and Noise Reduction based on Diffuse Power Spectral Density Estimators

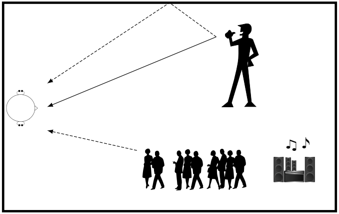
Ina Kodrasi, Daniel Marquardt, Simon Doclo

Department of Medical Physics and Acoustics
and Cluster of Excellence Hearing4All
Signal Processing Group
University of Oldenburg

19.08.2017

Binaural Speech Enhancement

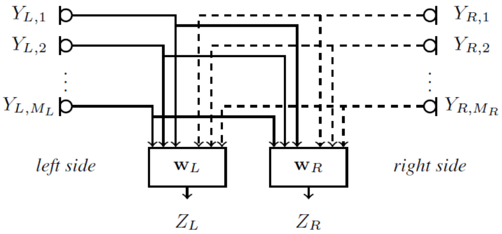
- Hearing impaired suffer from loss of speech understanding in **noisy and reverberant environments**
- Binaural dereverberation and noise reduction required
- Here: focus on **diffuse noise and late reverberation**



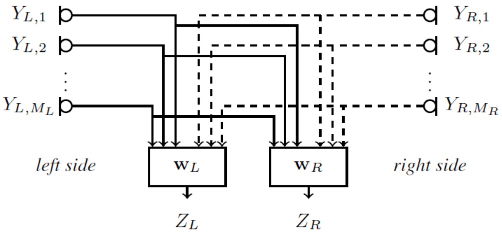
Objective of binaural dereverberation and noise reduction

- Improve speech quality and intelligibility
- Preserve spatial awareness (binaural cue preservation)

Binaural Hearing Aid Configuration



Binaural Hearing Aid Configuration



Frequency-domain signal model

$$\mathbf{y} = \mathbf{x} + \mathbf{d} = S_{\text{L,R}}\mathbf{a}_{\text{L,R}} + \mathbf{d}$$

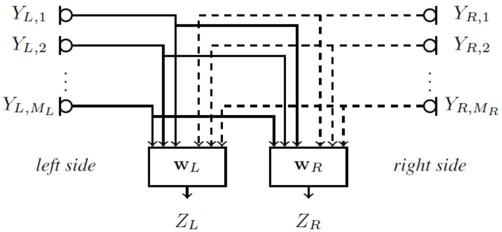
$\mathbf{x}$ $\rightarrow$ direct (and early reverberation) speech component

$S_{\text{L,R}}$ $\rightarrow$ target signal in the reference microphones (left/right)

$\mathbf{a}_{\text{L,R}}$ $\rightarrow$ relative early transfer functions (RETFs) of target signal

$\mathbf{d}$ $\rightarrow$ **late reverberation and background noise (diffuse)**

Binaural Hearing Aid Configuration



Frequency-domain signal model

$$\mathbf{y} = \mathbf{x} + \mathbf{d} = S_{\text{L,R}} \mathbf{a}_{\text{L,R}} + \mathbf{d}$$

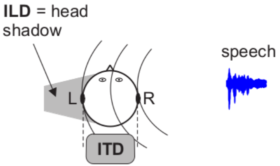
Uncorrelated signal components

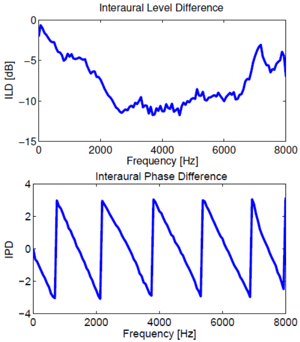
$$\mathbf{\Phi_y} = \mathbf{\Phi_x} + \mathbf{\Phi_d} = \Phi_{S_{\text{L,R}}} \mathbf{a}_{\text{L,R}} \mathbf{a}_{\text{L,R}}^H + \Phi_{\text{d}} \mathbf{\Gamma}$$

$\Phi_{S_{\text{L,R}}}$ $\rightarrow$ time-varying target signal PSD
Φ_{d} $\rightarrow$ **time-varying diffuse PSD**
$\mathbf{\Gamma}$ $\rightarrow$ time-invariant spatial coherence of diffuse sound field

Binaural Cues

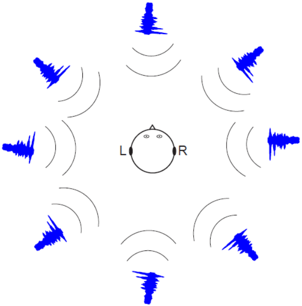
Directional sources: described by **Interaural Level Difference** (ILD) and **Interaural Phase/Time Difference** (IPD/ITD)

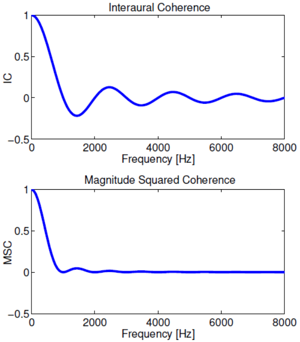




Binaural Cues

Diffuse sound fields: described by **Interaural Coherence** (IC) and **Magnitude Squared Coherence** (MSC)

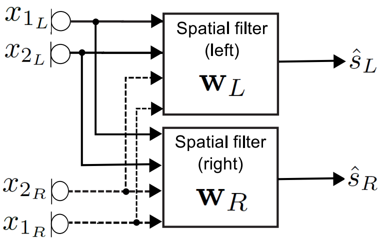




Binaural Speech Enhancement Techniques

1. Binaural Minimum Variance Distorionless Response (MVDR) beamformer

Objective: minimize output PSD of interference (reverberation and noise) while preserving target signal



Binaural Speech Enhancement Techniques

1. Binaural Minimum Variance Distorionless Response (MVDR) beamformer

Objective: minimize output PSD of interference (reverberation and noise) while preserving target signal

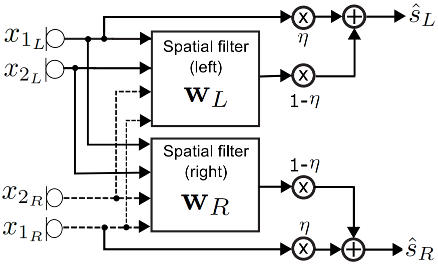
→ Perfect preservation of binaural cues of speech source

→ Distortion of output MSC of interference

Binaural Speech Enhancement Techniques

1. Binaural Minimum Variance Distorionless Response (MVDR) beamformer
2. Binaural MVDR with **partial noise estimation** (MVDR-N)

Objective: minimize output PSD of interference while preserving target signal and a scaled version of the interference



Binaural Speech Enhancement Techniques

1. Binaural Minimum Variance Distorionless Response (MVDR) beamformer
2. Binaural MVDR with **partial noise estimation** (MVDR-N)

Objective: minimize output PSD of interference while preserving target signal and a scaled version of the interference

→ Perfect preservation of binaural cues of speech source

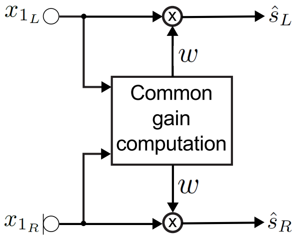
→ Output MSC of interference can be set to a desired MSC

Binaural Speech Enhancement Techniques

1. Binaural MVDR + **common spectro-temporal postfilter**
2. Binaural MVDR-N + **common spectro-temporal postfilter**

Objective: increase interference (reverberation and noise) reduction while allowing some speech distortion

→ Same binaural cues as at beamformer output



Binaural Speech Enhancement Techniques

1. Binaural MVDR + common spectro-temporal postfilter
2. Binaural MVDR-N + common spectro-temporal postfilter

These techniques require (among other quantities) **an estimate of the time-varying diffuse PSD Φ_d**

Diffuse Power Spectral Density Estimators

1. Blocking-based estimators

- Diffuse PSD estimated at output of blocking matrix aiming to block the target signal
- Require knowledge of the RETF vector $\mathbf{a}_{\mathrm{L,R}}$ and diffuse coherence matrix $\mathbf{\Gamma}$

Diffuse Power Spectral Density Estimators

1. Blocking-based estimators

- Diffuse PSD estimated at output of blocking matrix aiming to block the target signal
- Require knowledge of the RETF vector $\mathbf{a}_{L,R}$ and diffuse coherence matrix $\mathbf{\Gamma}$

2. Eigenvalue decomposition (EVD)-based estimators

- Diffuse PSD estimated using the eigenvalues of the prewhitened signal PSD matrix
- Require only knowledge of the diffuse coherence matrix $\mathbf{\Gamma}$

1. Blocking-based Estimators

- Construct a blocking matrix $\mathbf{B}$ such that $\mathbf{B}^H \mathbf{a}_{\mathrm{L,R}} = \mathbf{0}$
- Block the target signal from the received microphone signals

$$\tilde{\mathbf{u}} = \mathbf{B}^H \mathbf{y} \qquad \boldsymbol{\Phi}_{\tilde{\mathbf{u}}} = \Phi_{\mathrm{d}} \underbrace{\mathbf{B}^H \boldsymbol{\Gamma} \mathbf{B}}_{\tilde{\boldsymbol{\Gamma}}}$$

1. Blocking-based Estimators

- Construct a blocking matrix $\mathbf{B}$ such that $\mathbf{B}^H \mathbf{a}_{\mathrm{L,R}} = \mathbf{0}$
- Block the target signal from the received microphone signals

$$\tilde{\mathbf{u}} = \mathbf{B}^H \mathbf{y} \qquad \mathbf{\Phi}_{\tilde{\mathbf{u}}} = \Phi_{\mathrm{d}} \underbrace{\mathbf{B}^H \mathbf{\Gamma} \mathbf{B}}_{\tilde{\mathbf{\Gamma}}}$$

- Least-squares cost function to estimate Φ_{d}

$$\min_{\Phi_{\mathrm{d}}} \left\| \mathbf{\Phi}_{\tilde{\mathbf{u}}} - \Phi_{\mathrm{d}} \tilde{\mathbf{\Gamma}} \right\|_F^2$$

- Blocking-based diffuse PSD estimate

$$\hat{\Phi}_{\mathrm{d}}^{\mathrm{BM}} = \frac{\mathrm{trace}\{\mathbf{\Phi}_{\tilde{\mathbf{u}}}^H \tilde{\mathbf{\Gamma}}\}}{\mathrm{trace}\{\tilde{\mathbf{\Gamma}}^H \tilde{\mathbf{\Gamma}}\}}$$

1. Blocking-based Estimators

- Construct a blocking matrix $\mathbf{B}$ such that $\mathbf{B}^H \mathbf{a}_{\mathrm{L,R}} = \mathbf{0}$
- Block the target signal from the received microphone signals

$$\tilde{\mathbf{u}} = \mathbf{B}^H \mathbf{y} \qquad \boldsymbol{\Phi}_{\tilde{\mathbf{u}}} = \Phi_{\mathrm{d}} \underbrace{\mathbf{B}^H \boldsymbol{\Gamma} \mathbf{B}}_{\tilde{\boldsymbol{\Gamma}}}$$

- Least-squares cost function to estimate Φ_{d}

$$\min_{\Phi_{\mathrm{d}}} \left\| \boldsymbol{\Phi}_{\tilde{\mathbf{u}}} - \Phi_{\mathrm{d}} \tilde{\boldsymbol{\Gamma}} \right\|_F^2$$

- Blocking-based diffuse PSD estimate

$$\hat{\Phi}_{\mathrm{d}}^{\mathrm{BM}} = \frac{\mathrm{trace}\{\boldsymbol{\Phi}_{\tilde{\mathbf{u}}}^H \tilde{\boldsymbol{\Gamma}}\}}{\mathrm{trace}\{\tilde{\boldsymbol{\Gamma}}^H \tilde{\boldsymbol{\Gamma}}\}}$$

Considered blocking-based estimates

- $\hat{\Phi}_{\mathrm{d,2}}^{\mathrm{BM}}$ obtained using $M = 2$ [Marquardt, WASPAA 2017]
- $\hat{\Phi}_{\mathrm{d,M}}^{\mathrm{BM}}$ obtained using $M > 2$ [Braun, EUSIPCO 2013]

2. EVD-based Estimators

- Prewhiten the signal PSD matrix using $\mathbf{\Gamma}^{-1}$

$$\mathbf{\Phi}_{\mathbf{y}}^{\mathrm{w}} = \mathbf{\Gamma}^{-1}\mathbf{\Phi}_{\mathbf{y}} = \underbrace{\mathbf{\Gamma}^{-1}\Phi_{S_{\mathrm{L,R}}}\mathbf{a}_{\mathrm{L,R}}\mathbf{a}_{\mathrm{L,R}}^{H}}_{\text{rank-1}} + \Phi_{\mathrm{d}}\mathbf{I}$$

2. EVD-based Estimators

- Prewhiten the signal PSD matrix using $\mathbf{\Gamma}^{-1}$

$$\mathbf{\Phi}_{\mathbf{y}}^{\mathrm{w}} = \mathbf{\Gamma}^{-1}\mathbf{\Phi}_{\mathbf{y}} = \underbrace{\mathbf{\Gamma}^{-1}\Phi_{S_{\mathrm{L,R}}}\mathbf{a}_{\mathrm{L,R}}\mathbf{a}_{\mathrm{L,R}}^{H}}_{\text{rank-1}} + \Phi_{\mathrm{d}}\mathbf{I}$$

- Estimate Φ_{d} using the eigenvalues of $\mathbf{\Phi}_{\mathbf{y}}^{\mathrm{w}}$

$$\mathbf{\Lambda} = \begin{bmatrix} \sigma + \Phi_{\mathrm{d}} & 0 & \dots & 0 \\ 0 & \Phi_{\mathrm{d}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \Phi_{\mathrm{d}} \end{bmatrix}$$

2. EVD-based Estimators

- Prewhiten the signal PSD matrix using $\boldsymbol{\Gamma}^{-1}$

$$\boldsymbol{\Phi}_{\mathbf{y}}^{\mathrm{w}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Phi}_{\mathbf{y}} = \underbrace{\boldsymbol{\Gamma}^{-1}\Phi_{S_{\mathrm{L,R}}}\mathbf{a}_{\mathrm{L,R}}\mathbf{a}_{\mathrm{L,R}}^{H}}_{\text{rank-1}} + \Phi_{\mathrm{d}}\mathbf{I}$$

- Estimate Φ_{d} using the eigenvalues of $\boldsymbol{\Phi}_{\mathbf{y}}^{\mathrm{w}}$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sigma + \Phi_{\mathrm{d}} & 0 & \dots & 0 \\ 0 & \Phi_{\mathrm{d}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \Phi_{\mathrm{d}} \end{bmatrix}$$

Considered EVD-based estimates [Kodrasi, ICASSP 2017]

- $\hat{\Phi}_{\mathrm{d},\lambda_1}^{\mathrm{EVD}} = \frac{\mathrm{trace}\{\boldsymbol{\Phi}_{\mathbf{y}}^{\mathrm{w}}\} - \lambda_1\{\boldsymbol{\Phi}_{\mathbf{y}}^{\mathrm{w}}\}}{M-1}$
- $\hat{\Phi}_{\mathrm{d},\lambda_2}^{\mathrm{EVD}} = \lambda_2\{\boldsymbol{\Phi}_{\mathbf{y}}^{\mathrm{w}}\}$

Objective of Simulation Study

Compare the performance of

1. MVDR beamformer + common postfilter
2. MVDR-N beamformer + common postfilter

using

1. blocking-based diffuse PSD estimate $\hat{\Phi}_{\mathrm{d},2}^{\mathsf{BM}}$
2. blocking-based diffuse PSD estimate $\hat{\Phi}_{\mathrm{d},4}^{\mathsf{BM}}$
3. EVD-based diffuse PSD estimate $\hat{\Phi}_{\mathrm{d},\lambda_1}^{\mathsf{EVD}}$
4. EVD-based diffuse PSD estimate $\hat{\Phi}_{\mathrm{d},\lambda_2}^{\mathsf{EVD}}$

Acoustic Scenarios

Recordings in a variable acoustics lab using hearing aid dummies on a HATS with $M_\mathrm{L} = M_\mathrm{R} = 2$

Acoustic Scenarios

Recordings in a variable acoustics lab using hearing aid dummies on a HATS with $M_\mathsf{L} = M_\mathsf{R} = 2$

Stationary speaker scenarios: Loudspeaker placed at 35° and −35°, $T_{60} \in \{0.5\ \mathrm{s}, 0.75\ \mathrm{s}, 1\ \mathrm{s}\}$

Moving speaker scenario: Human speaker walking in the frontal hemisphere of the HATS, $T_{60} \approx 1\ \mathrm{s}$

Background noise: 4 loudspeakers playing back uncorrelated multi-talker noise, iSNR $\in \{0\ \mathrm{dB}, 5\ \mathrm{dB}, \ldots, 20\ \mathrm{dB}, \infty\ \mathrm{dB}\}$

Algorithmic Settings and Measures

- WOLA, frame size of 512 samples, 50% overlap, $f_s = 16$ kHz

Algorithmic Settings and Measures

- WOLA, frame size of 512 samples, 50% overlap, $f_s = 16$ kHz
- DOA estimated using binaural DOA estimator from [Marquardt, WASPAA 2017]
- RETFs constructed based on the estimated DOA and measured anechoic HRTFs
- $\mathbf{\Gamma}$ constructed using measured anechoic HRTFs

Algorithmic Settings and Measures

- WOLA, frame size of 512 samples, 50% overlap, $f_s = 16$ kHz
- DOA estimated using binaural DOA estimator from [Marquardt, WASPAA 2017]
- RETFs constructed based on the estimated DOA and measured anechoic HRTFs
- $\mathbf{\Gamma}$ constructed using measured anechoic HRTFs
- Desired output MSC for the MVDR-N beamformer defined using psycho-acoustically motivated boundaries [Walther, JASA 2013]

Algorithmic Settings and Measures

- WOLA, frame size of 512 samples, 50% overlap, $f_s = 16$ kHz
- DOA estimated using binaural DOA estimator from [Marquardt, WASPAA 2017]
- RETFs constructed based on the estimated DOA and measured anechoic HRTFs
- $\mathbf{\Gamma}$ constructed using measured anechoic HRTFs
- Desired output MSC for the MVDR-N beamformer defined using psycho-acoustically motivated boundaries [Walther, JASA 2013]
- Minimum postfilter gain -20 dB

Algorithmic Settings and Measures

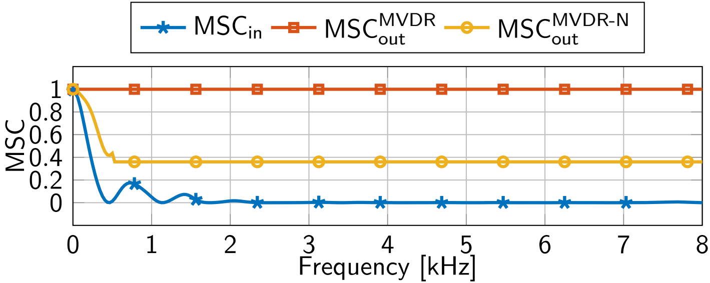
- WOLA, frame size of 512 samples, 50% overlap, $f_s = 16$ kHz
- DOA estimated using binaural DOA estimator from [Marquardt, WASPAA 2017]
- RETFs constructed based on the estimated DOA and measured anechoic HRTFs
- $\mathbf{\Gamma}$ constructed using measured anechoic HRTFs
- Desired output MSC for the MVDR-N beamformer defined using psycho-acoustically motivated boundaries [Walther, JASA 2013]
- Minimum postfilter gain -20 dB

Spatial awareness preservation: MSC

Dereverberation and noise reduction: ΔPESQ, ΔfSSNR

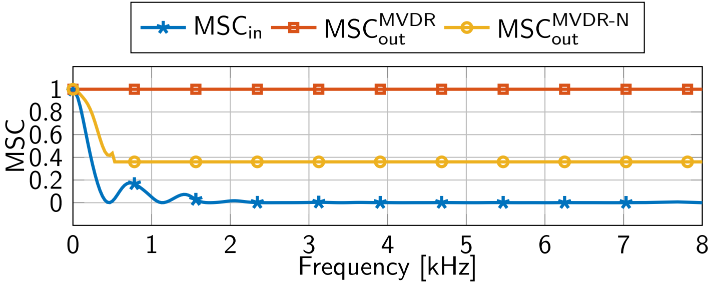
Spatial Awareness

Interference MSC at input and beamformer output



Spatial Awareness

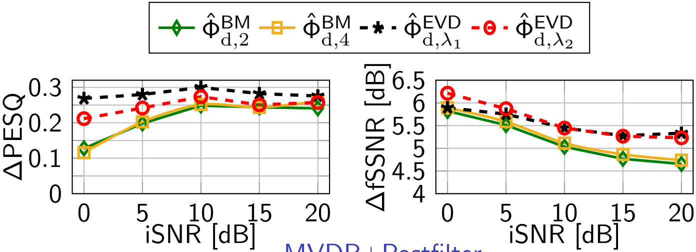
Interference MSC at input and beamformer output



Using

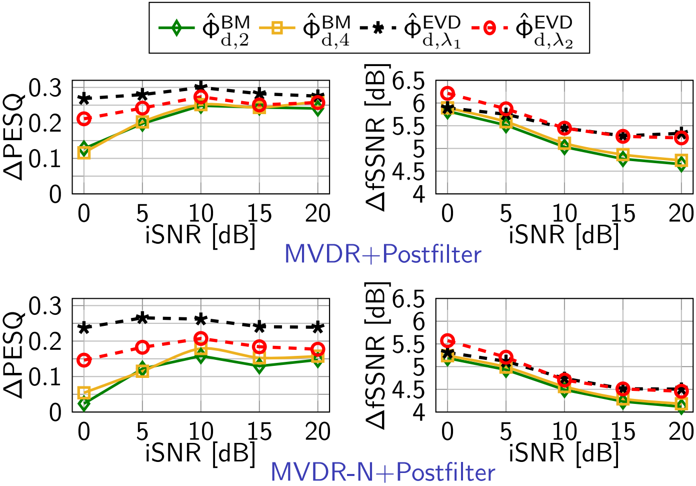
- Binaural MVDR beamformer distorts interference MSC
- Binaural MVDR-N beamformer sets interference MSC to desired MSC

Stationary Speaker



MVDR+Postfilter

Stationary Speaker



Stationary Speaker

When dereverberating and denoising a stationary speaker

- As expected, interference reduction performance of MVDR-N is lower than MVDR, but binaural cues of interference are preserved
- **EVD-based estimators outperform blocking-based estimators** (best performance with $\hat{\Phi}^{\mathrm{EVD}}_{\mathrm{d},\lambda_1}$)
- Increasing the number of microphones for blocking-based estimators does not significantly increase the performance

Moving Speaker

ΔfSSNR [dB]	$\hat{\Phi}^{\text{BM}}_{\text{d},2}$	$\hat{\Phi}^{\text{BM}}_{\text{d},4}$	$\hat{\Phi}^{\text{EVD}}_{\text{d},\lambda_1}$	$\hat{\Phi}^{\text{EVD}}_{\text{d},\lambda_2}$
MVDR+Postfilter	7.42	7.55	6.78	**7.86**
MVDR-N+Postfilter	6.83	6.89	6.66	**7.63**

Moving Speaker

ΔfSSNR [dB]	$\hat{\Phi}^{\mathrm{BM}}_{\mathrm{d},2}$	$\hat{\Phi}^{\mathrm{BM}}_{\mathrm{d},4}$	$\hat{\Phi}^{\mathrm{EVD}}_{\mathrm{d},\lambda_1}$	$\hat{\Phi}^{\mathrm{EVD}}_{\mathrm{d},\lambda_2}$
MVDR+Postfilter	7.42	7.55	6.78	**7.86**
MVDR-N+Postfilter	6.83	6.89	6.66	**7.63**

When dereverberating and denoising a moving speaker

- EVD-based estimators (in particular $\hat{\Phi}^{\mathrm{EVD}}_{\mathrm{d},\lambda_2}$) outperform blocking-based estimators
- Increasing the number of microphones for blocking-based estimators does not significantly increase the performance

Summary and Outlook

Simulations show that

- State-of-the-art diffuse PSD estimators yield a high performance also when used for binaural dereverberation and noise reduction
- **EVD-based PSD estimators, not requiring DOA/RETF, yield the best performance**

Summary and Outlook

Simulations show that

- State-of-the-art diffuse PSD estimators yield a high performance also when used for binaural dereverberation and noise reduction
- **EVD-based PSD estimators, not requiring DOA/RETF, yield the best performance**

In the future

- Analyze performance in the presence of non-diffuse background noise
- Subjective listening tests to truly evaluate quality and spatial awareness