Cognitive-Driven Binaural Beamforming for Hearing Devices Using EEG-Based Auditory Attention Decoding

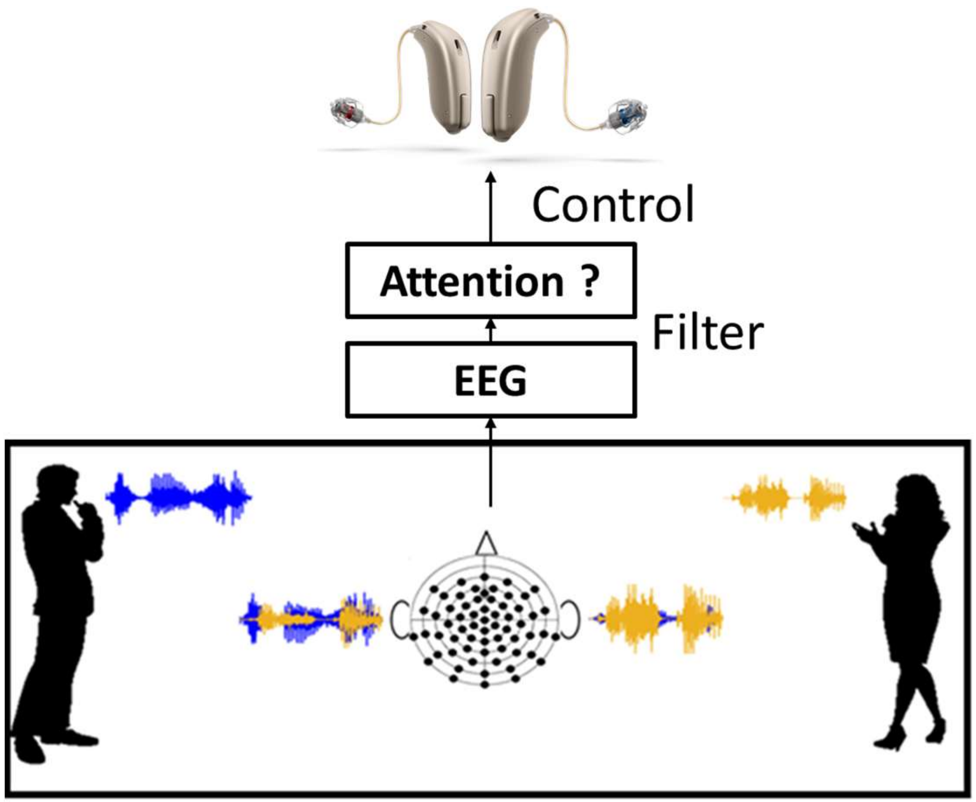
Simon Doclo, **Ali Aroudi**

Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4all
University of Oldenburg, Germany

CIAP – July 12, 2021

Problem statement

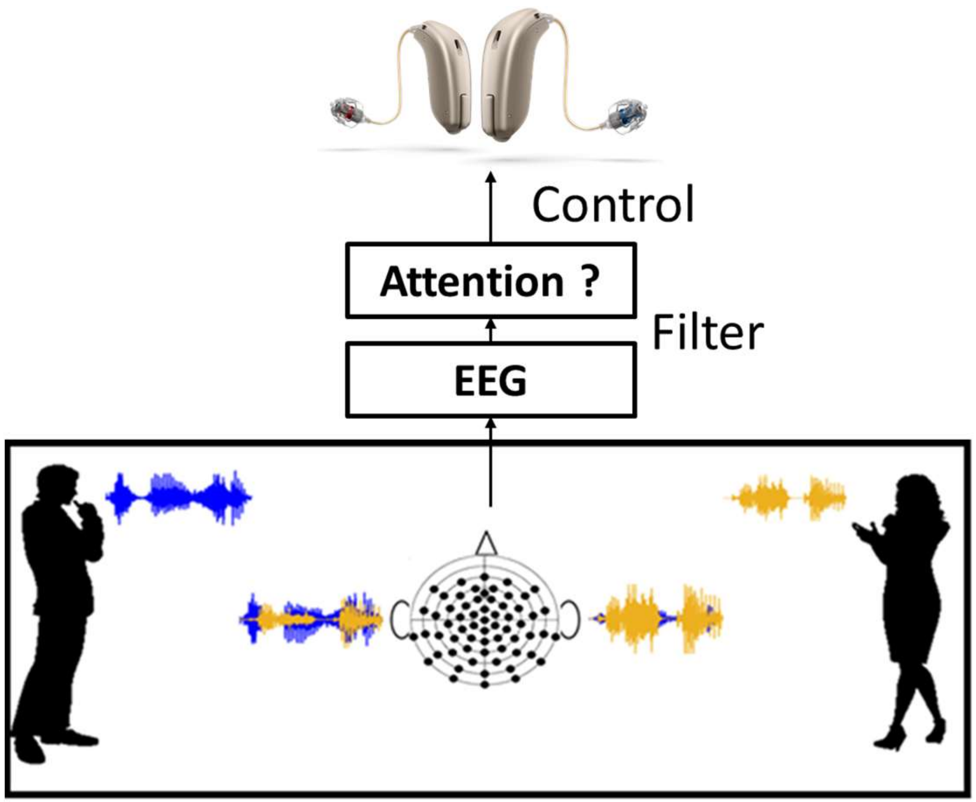
- Performance of speech enhancement and speaker separation algorithms depends on **correctly identifying target speaker** to be enhanced
- **Auditory attention decoding (AAD)** using single-trial EEG recordings
- **Cognitive-driven source separation and noise reduction algorithms**
- **This presentation: cognitive-driven binaural beamformer** based on AAD and acoustic scene analysis for realistic noisy and reverberant acoustic environments



[Aroudi & Doclo, *Cognitive-driven binaural beamforming using EEG-based auditory attention decoding*, IEEE TASLP, 2020.]

Outline

- **Least-squares-based AAD method:** performance in noisy and reverberant environments
- **Cognitive-driven binaural beamforming system:**
  - Minimum variance distortionless response (MVDR) beamformer
  - Linearly constrained minimum variance (LCMV) beamformer
- Evaluation in **noisy and reverberant environment** with 2 competing speakers



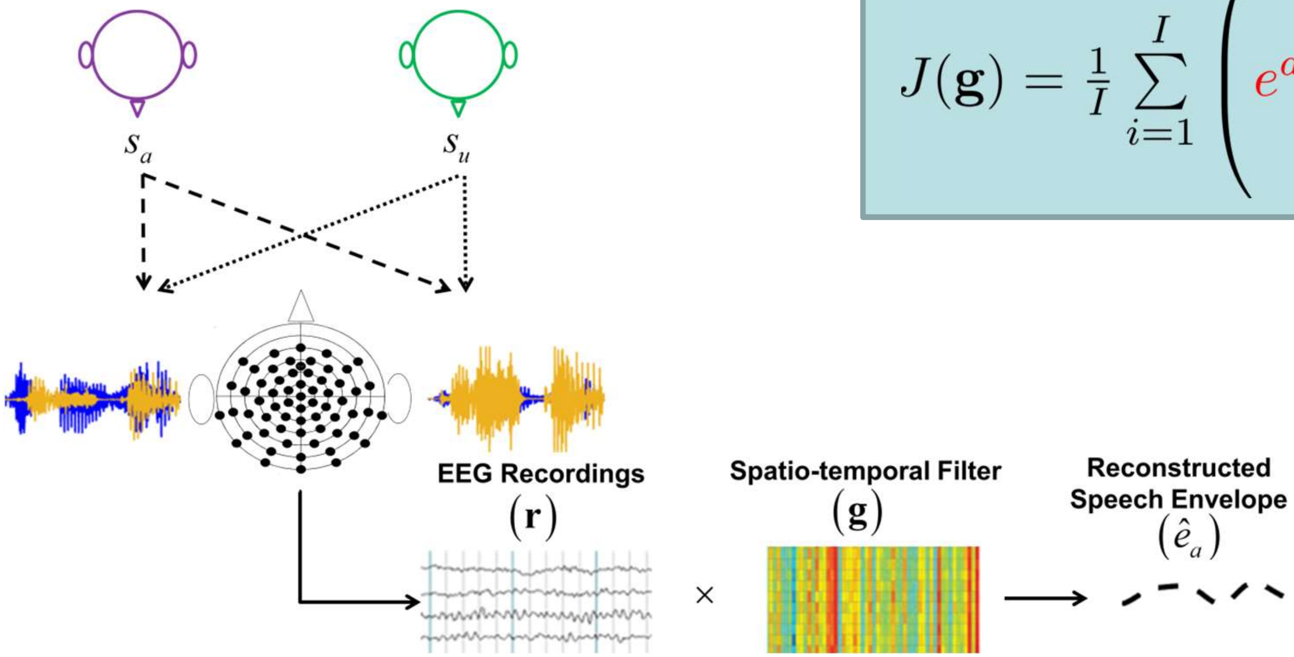
[O'Sullivan et al., *Attentional selection in a cocktail party environment can be decoded from single-trial EEG*, Cerebral Cortex, 2014.]
[Aroudi et al., *Impact of Different Acoustic Components on EEG-based Auditory Attention Decoding in Noisy and Reverberant Conditions*, IEEE Trans. Neural Systems and Rehabilitation Engineering, 2019.]

Least-squares-based AAD method

Auditory attention decoding method

- **Training step**: compute spatio-temporal filter **g**

$$J(\mathbf{g}) = \frac{1}{I}\sum_{i=1}^{I}\Bigg(e^a[i] - \underbrace{\mathbf{g}^T\mathbf{r}[i]}_{\hat{e}_a[i]}\Bigg)^2 + \beta\mathbf{g}^T\mathbf{D}\mathbf{g}$$

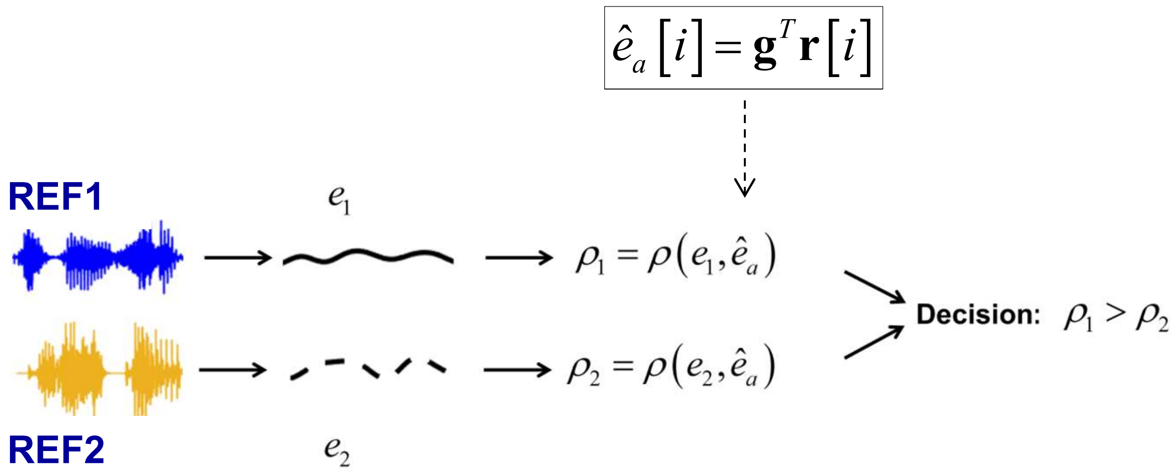


- Reconstruct envelope of attended speech signal by filtering and combining EEG signals **r**
- Regularization to avoid over-fitting

[O'Sullivan et al., *Cerebral Cortex*, 2014.]

Auditory attention decoding method

- **Decoding step:** correlate envelope of estimated attended speech signal with envelopes of *reference signals*



speaker 1 = attended
speaker 2 = unattended

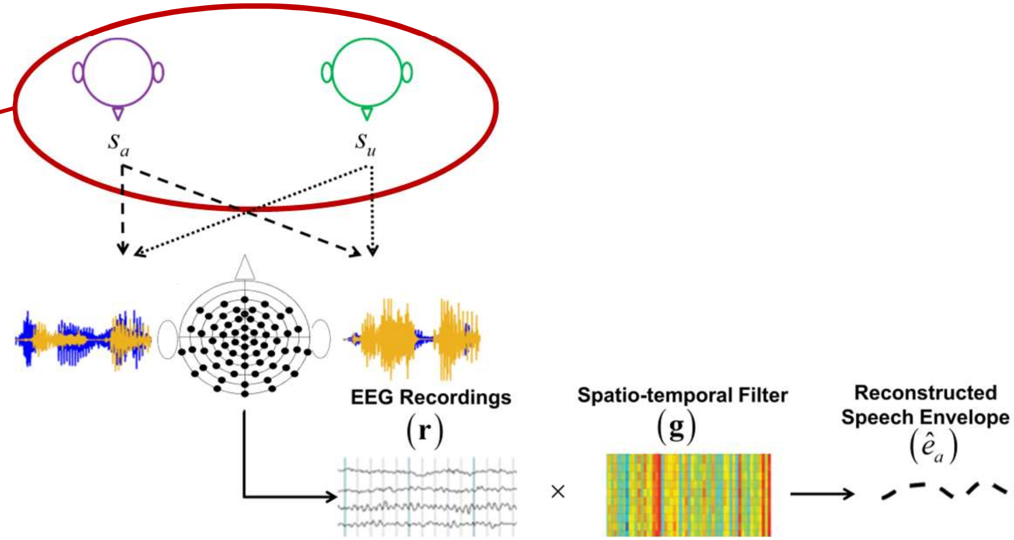
[O'Sullivan et al., *Cerebral Cortex*, 2014.]

Acoustic setup and simulation

Left and right speaker simulated at -45° and 45°

Two audio stories by two different male speakers (German)

Acoustic stimuli presented to participants using insert earphones



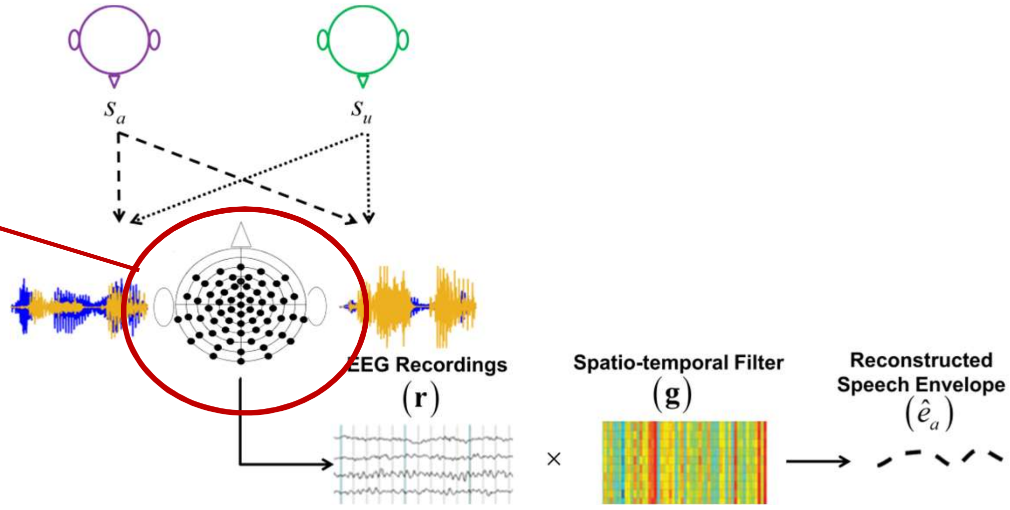
Experimental Analysis Condition	Stimuli Presentation	SNR[dB]	T_{60}[s]
Noiseless	Noiseless	∞	<0.05
Reverberant	Reverberant I	∞	0.50
	Reverberant II	∞	1.00
Noisy	Noisy I	9.0	<0.05
	Noisy II	4.0	<0.05
Reverberant-noisy	Reverberant-noisy I	9.0	0.50
	Reverberant-noisy II	4.0	0.50
	Reverberant-noisy III	9.0	1.00

diffuse babble noise

[Aroudi, Mirkovic, De Vos, Doclo, *IEEE Trans. Neural Systems and Rehabilitation Engineering,* 2019.]

EEG setup, training and decoding

- **Subjects:**
  - *N*=18 German-speaking participants
  - 8 instructed to attend to left speaker, 10 instructed to attend to right speaker
- **EEG signals:**
  - 64 channels (Easycap GmbH)
  - band-pass filtered (2-8 Hz), f_s = 64 Hz
- **Training and decoding:**
  - trial length: **60 seconds**
  - each participant's own data
- **Decoding performance:**
  - percentage of correctly decoded trials over all considered trials and participants
  - leave-one-out cross-validation approach

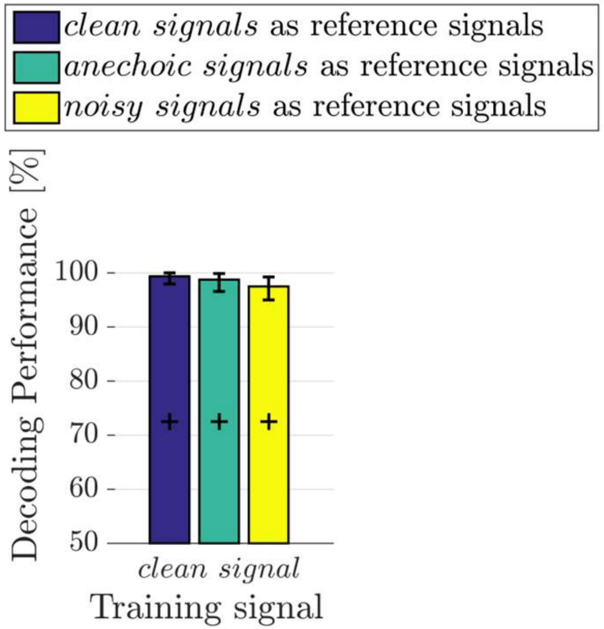


[Aroudi, Mirkovic, De Vos, Doclo, *IEEE Trans. Neural Systems and Rehabilitation Engineering*, 2019.]

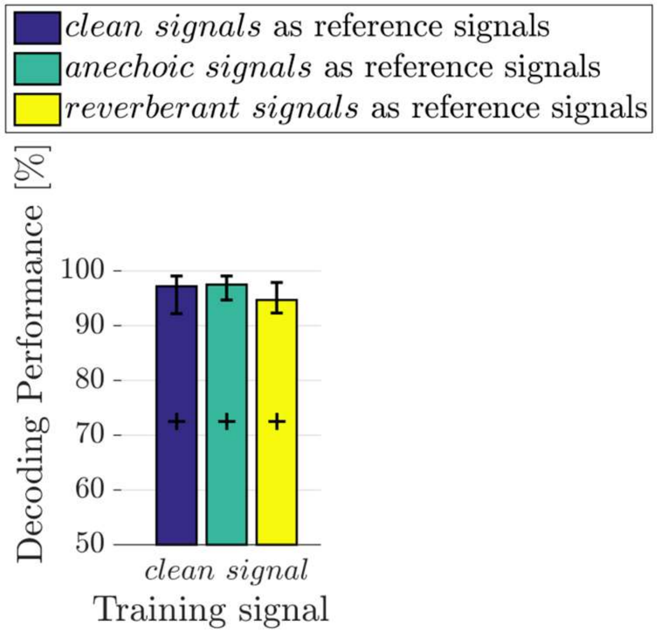
Experimental results: decoding performance

Reference signals: influence of noise, reverberation and interfering speaker

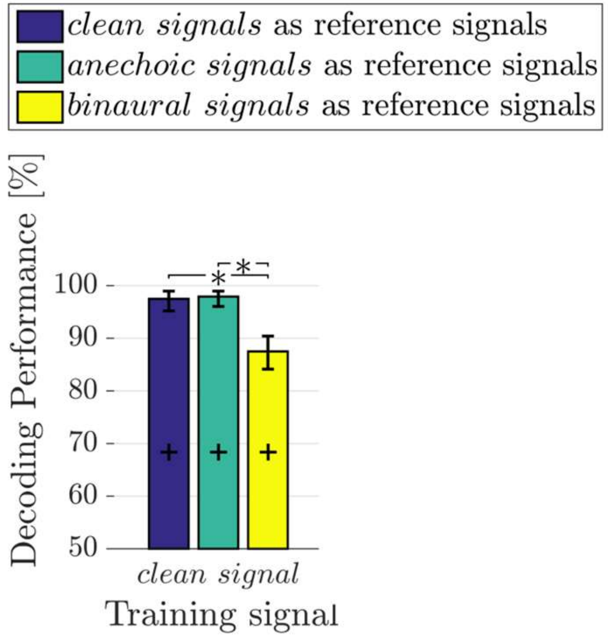
Anechoic - Noisy



Reverberant - Noiseless



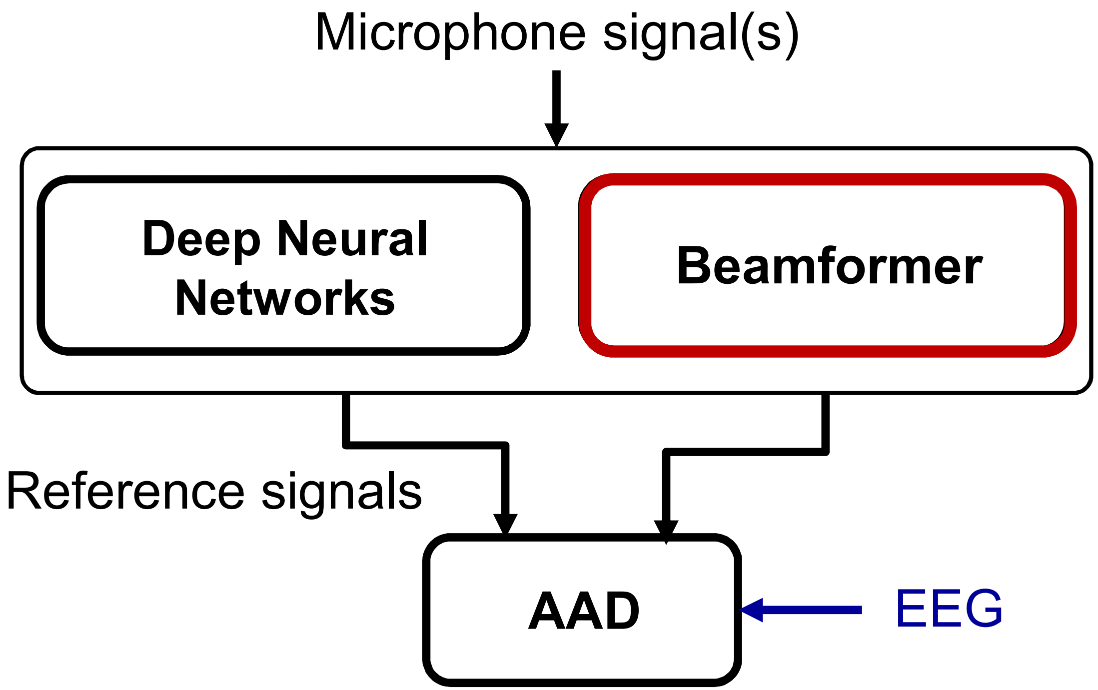
Reverberant - Noisy



- Reference signals affected by **reverberation or noise** → comparable decoding performance as when using clean reference signals
- Reference signals affected by **interfering speaker** → decoding performance significantly decreases

Auditory attention decoding

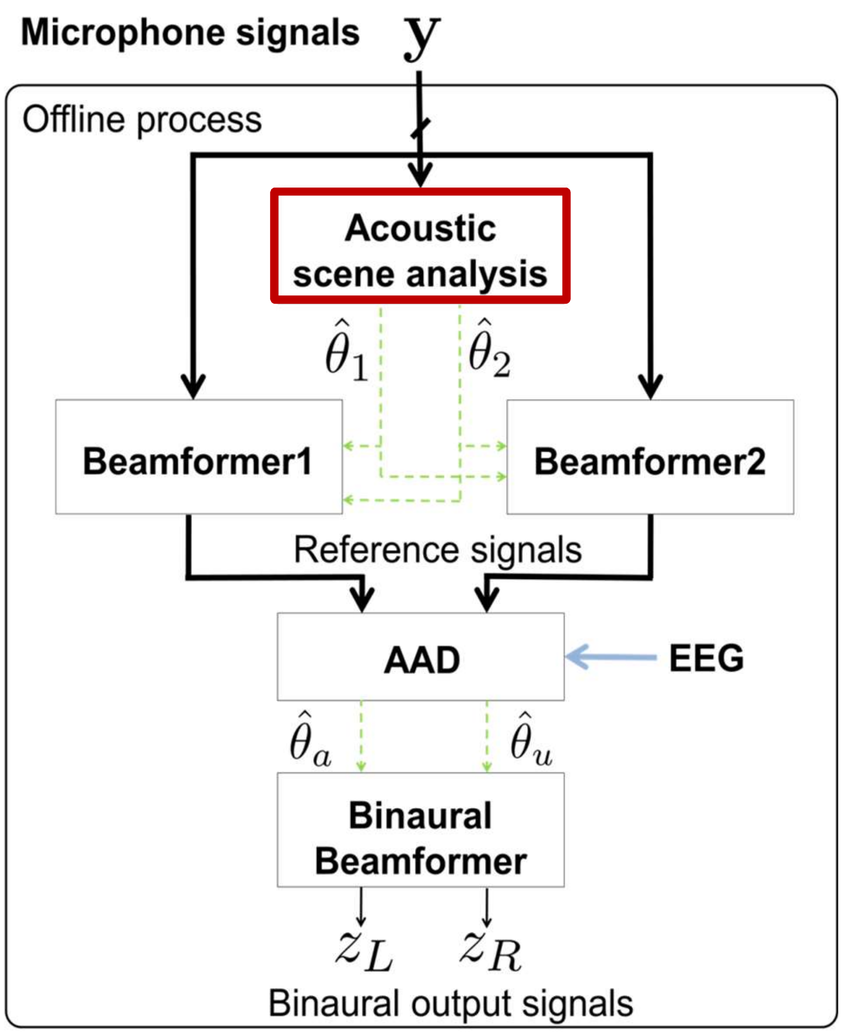
- Feasible to perform decoding in reverberant-noisy conditions ✓
- Best performance using clean speech signals as reference signals, but are not available in practice ✗
- **Generate** reference signals for decoding from microphone signals



[O'Sullivan 2017] [Van Eyndhoven 2017] [Das 2017] [Han 2019] [Aroudi 2020] [Borgström 2021]

Cognitive-driven binaural beamformer

Cognitive-driven beamformer

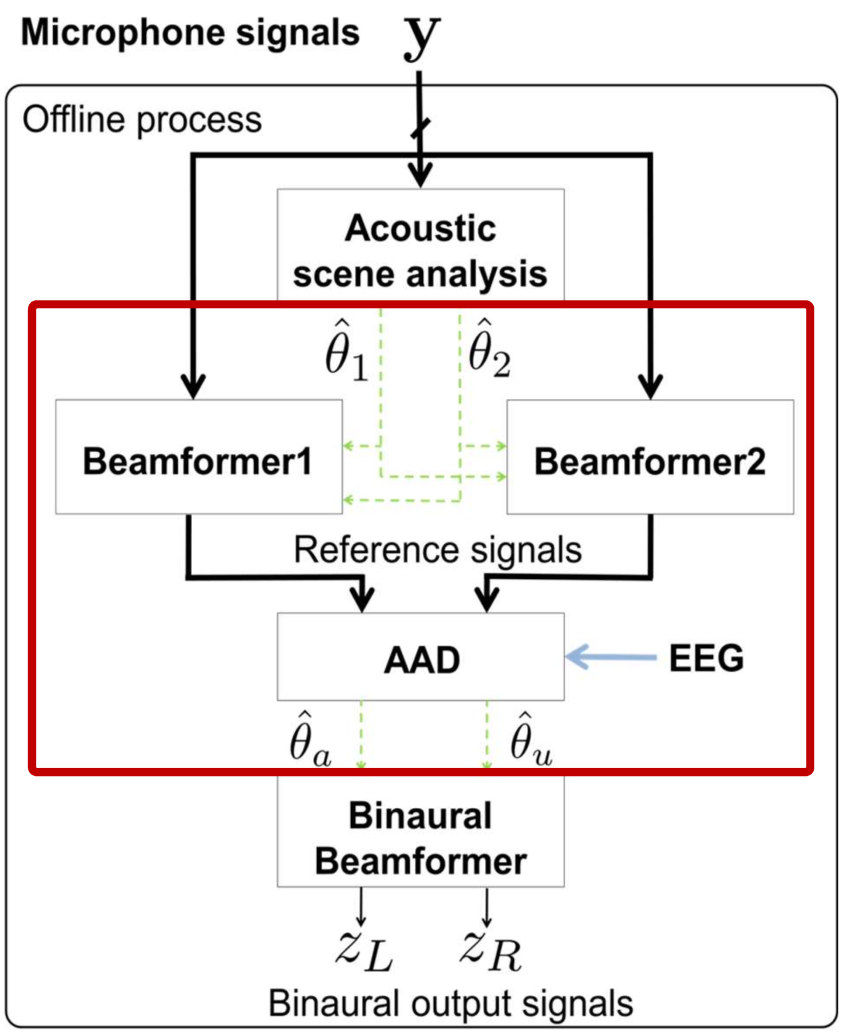


Process flow

1. **Acoustic scene analysis:** estimate direction-of-arrival (DOA) of speakers

[Aroudi, Doclo, *IEEE/ACM Trans. Audio, Speech and Language Processing,* 2020.]

Cognitive-driven beamformer



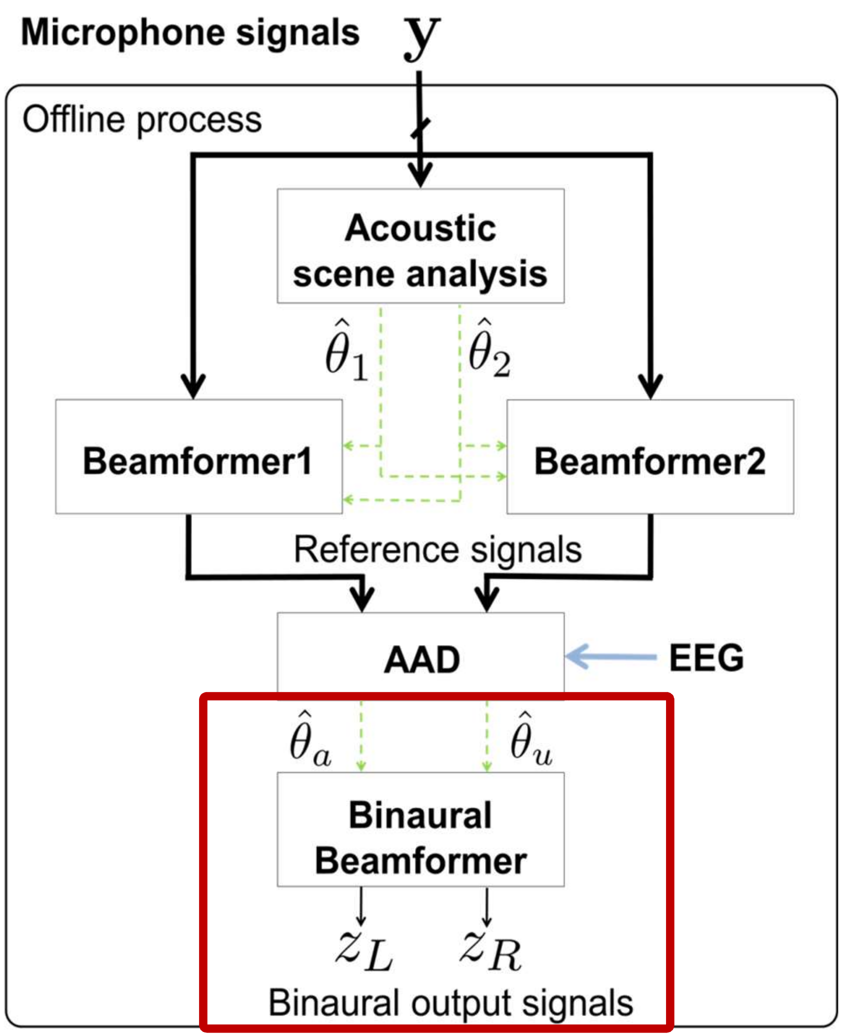
Process flow

1. Acoustic scene analysis: estimate direction-of-arrival (DOA) of speakers

2. **AAD using beamformer output signals** (steered to speakers) decides which speaker is attended/unattended

[Aroudi, Doclo, *IEEE/ACM Trans. Audio, Speech and Language Processing*, 2020.]

Cognitive-driven beamformer



Process flow

1. Acoustic scene analysis: estimate direction-of-arrival (DOA) of speakers

2. AAD using beamformer output signals (steered to speakers) decides which speaker is attended/unattended

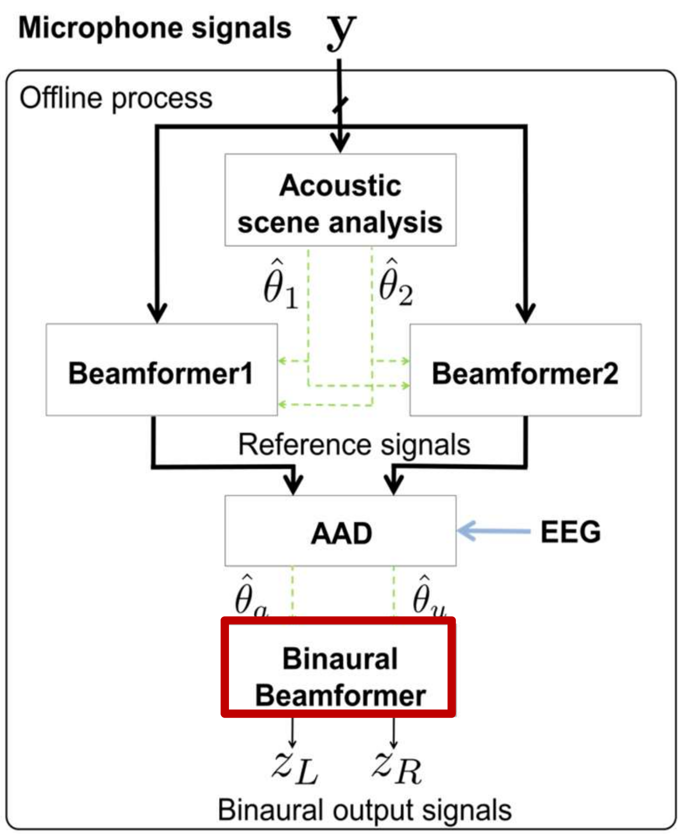
3. AAD information is used in **binaural beamformer** to:
   - **Pass** (estimated) attended speaker
   - **Suppress** (estimated) unattended speaker
   - **Preserve spatial impression** of acoustic scene (binaural cues)

[Aroudi, Doclo, *IEEE/ACM Trans. Audio, Speech and Language Processing*, 2020.]

MVDR/LCMV Beamformer

- **Minimum Variance Distortionless Response (MVDR) beamformer** aims at
  1. minimizing noise output PSD
  2. passing *attended direction* $\hat{\theta}_a$ without distortion

$$\min_{\mathbf{w}} \underbrace{\mathbf{w}^H \boldsymbol{\Phi}_v \mathbf{w}}_{\text{noise PSD}} \quad \text{subject to} \quad \underbrace{\mathbf{w}^H \mathbf{a}(\hat{\theta}_a) = 1}_{\text{target}}$$

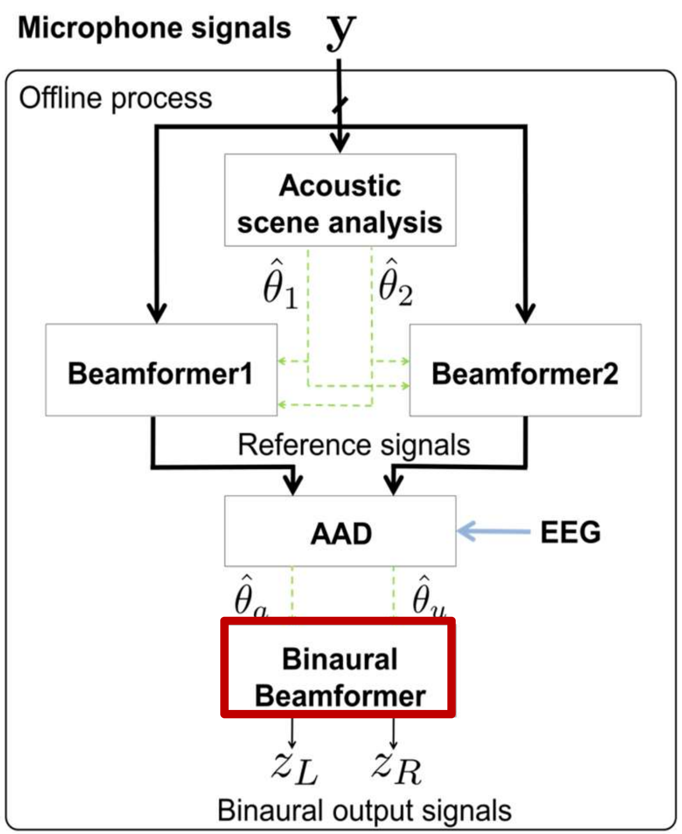


[Doclo, Kellermann, Makino, Nordholm, *IEEE Signal Processing Magazine*, 2015.]

MVDR/LCMV Beamformer

- **Linearly Constrained Minimum Variance (LCMV) beamformer** aims at
  1. minimizing noise output PSD
  2. passing *attended direction* $\hat{\theta}_a$ without distortion
  3. suppressing *unattended direction* $\hat{\theta}_u$ with factor $\delta < 1$
     → enables to control suppression

$$\min_{\mathbf{w}} \underbrace{\mathbf{w}^H \mathbf{\Phi}_v \mathbf{w}}_{\text{noise PSD}} \quad \text{subject to} \quad \underbrace{\mathbf{w}^H \mathbf{a}(\hat{\theta}_a) = 1}_{\text{target}}, \underbrace{\mathbf{w}^H \mathbf{a}(\hat{\theta}_u) = \delta}_{\text{interference}}$$



- **Requires**
  - *Noise covariance matrix*, e.g., diffuse noise assumption
  - Relative transfer functions (RTFs) of sources:
    - *Anechoic RTFs,* based on measured head-related transfer functions (HRTFs) and DOAs
    - *Reverberant RTFs*

[Hadad, Doclo, Gannot, *IEEE/ACM Trans. Audio, Speech and Language Processing,* 2016.]

Cognitive-driven binaural beamformer

Experimental evaluation

Acoustic setup and simulation

Experimental Analysis Condition	Stimuli Presentation	SINR [dB]	T_{60} [s]
Reverberant + Noisy	Reverberant-noisy I	-1.0	0.5
	Reverberant-noisy II	-2.5	0.5

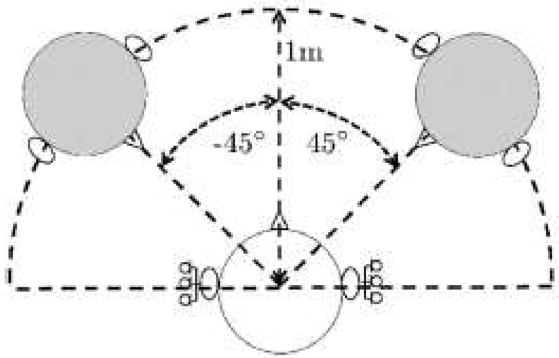
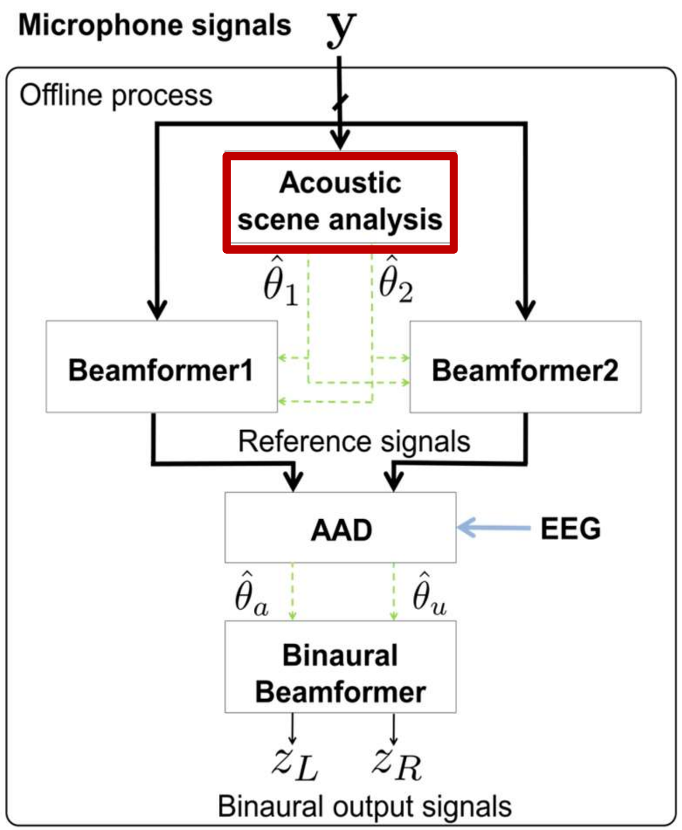


Fig. 2. Acoustic simulation setup for the reverberant condition. Two competing speakers were located at DOAs $\theta_1 = -45°$ and $\theta_2 = 45°$ and a distance of 1 m from the listener with two hearing aids, each equipped with 3 microphones.

Cognitive-driven beamformer: implementation

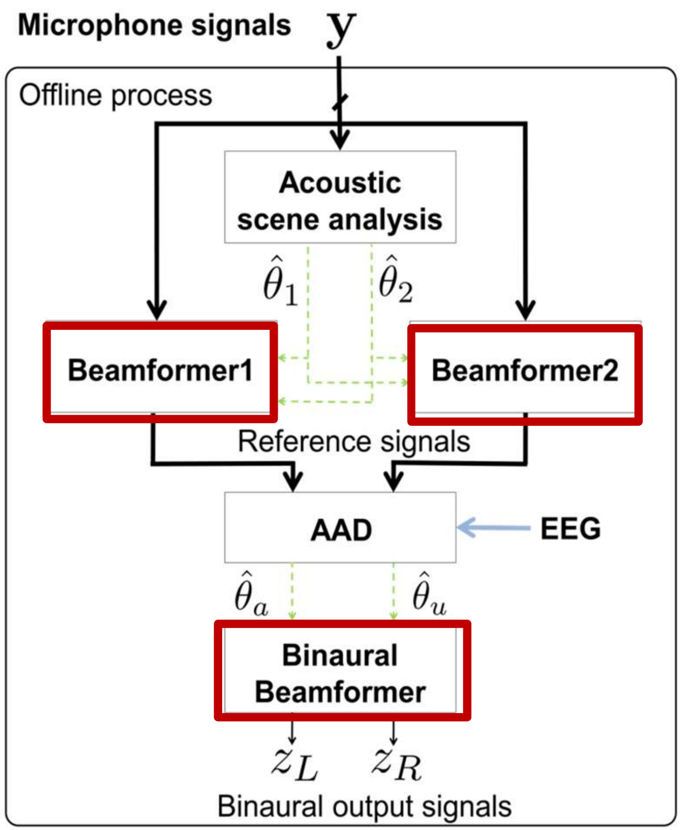
- **DOA estimation of speakers**
  - oracle DOA (ODOA)
  - estimated DOA (EDOA) from binaural microphone signals with SVM-based multi-source localization method using GCC-PHAT features



[Kayser et al., *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014.]

Cognitive-driven beamformer: implementation

- **DOA estimation of speakers**
  - oracle DOA (ODOA)
  - estimated DOA (EDOA) from binaural microphone signals with SVM-based multi-source localization method using GCC-PHAT features
- **MVDR/LCMV beamformer**
  - *Noise covariance matrix*: diffuse noise assumption
  - *Relative transfer functions*:
    - oracle reverberant RTFs (ORTF)
    - estimated reverberant RTFs (ERTF)
    - anechoic RTFs using oracle DOAs (ODOA)
    - anechoic RTFs using estimated DOAs (EDOA)

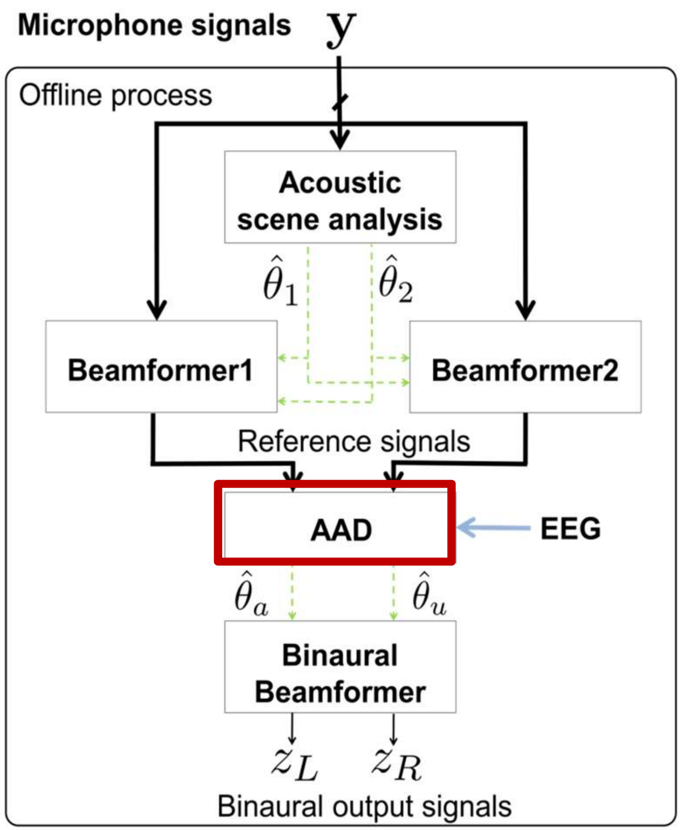


Cognitive-driven beamformer: implementation

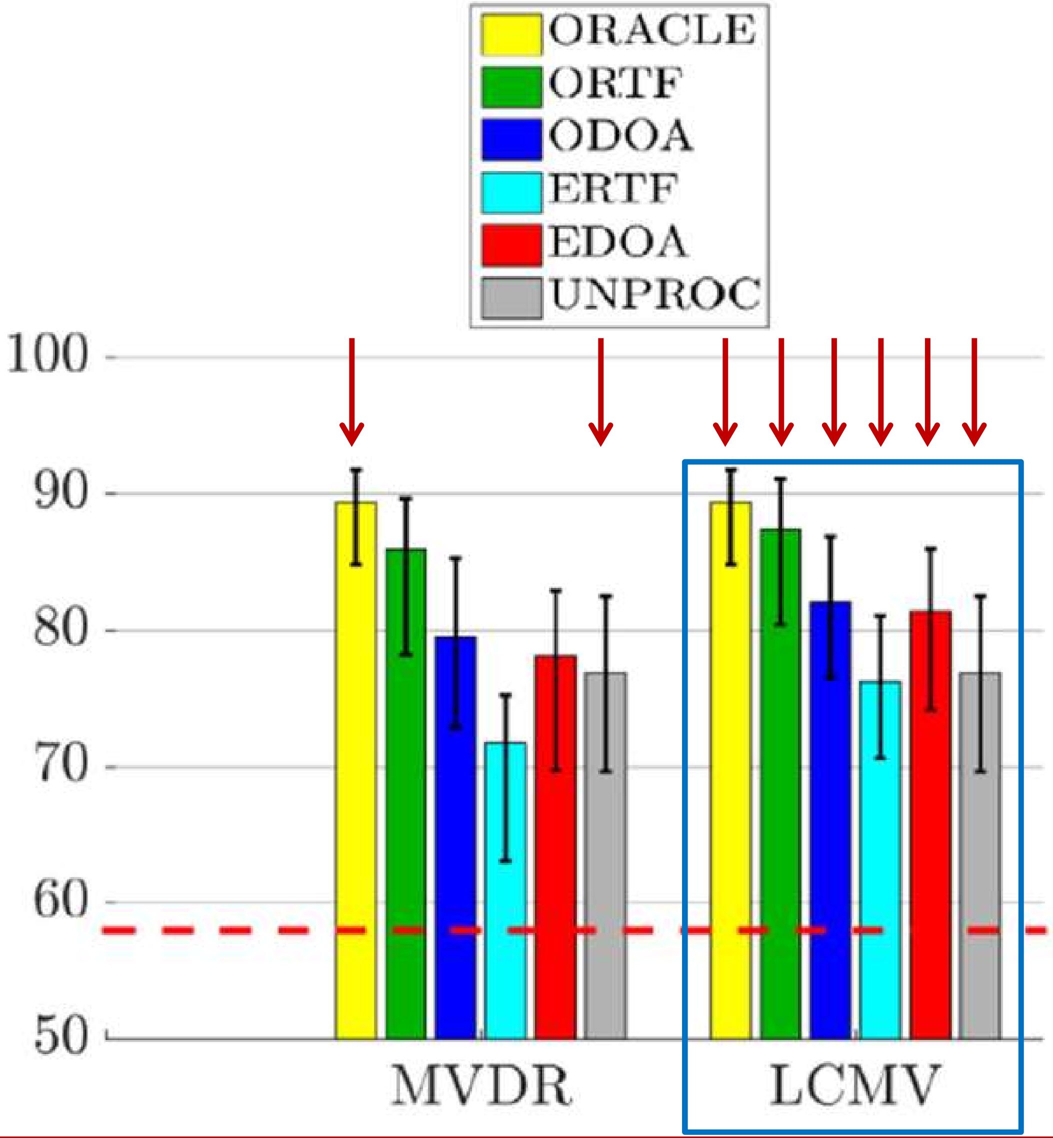
- **DOA estimation of speakers**
  - oracle DOA (ODOA)
  - estimated DOA (EDOA) from binaural microphone signals with SVM-based multi-source localization method using GCC-PHAT features

- **MVDR/LCMV beamformer**
  - *Noise covariance matrix:* diffuse noise assumption
  - *Relative transfer functions:*
    - oracle reverberant RTFs (ORTF)
    - estimated reverberant RTFs (ERTF)
    - anechoic RTFs using oracle DOAs (ODOA)
    - anechoic RTFs using estimated DOAs (EDOA)

- **Auditory attention decoding**
  - trial length: **30 seconds**
  - oracle AAD (OAAD) or estimated AAD (AAD)

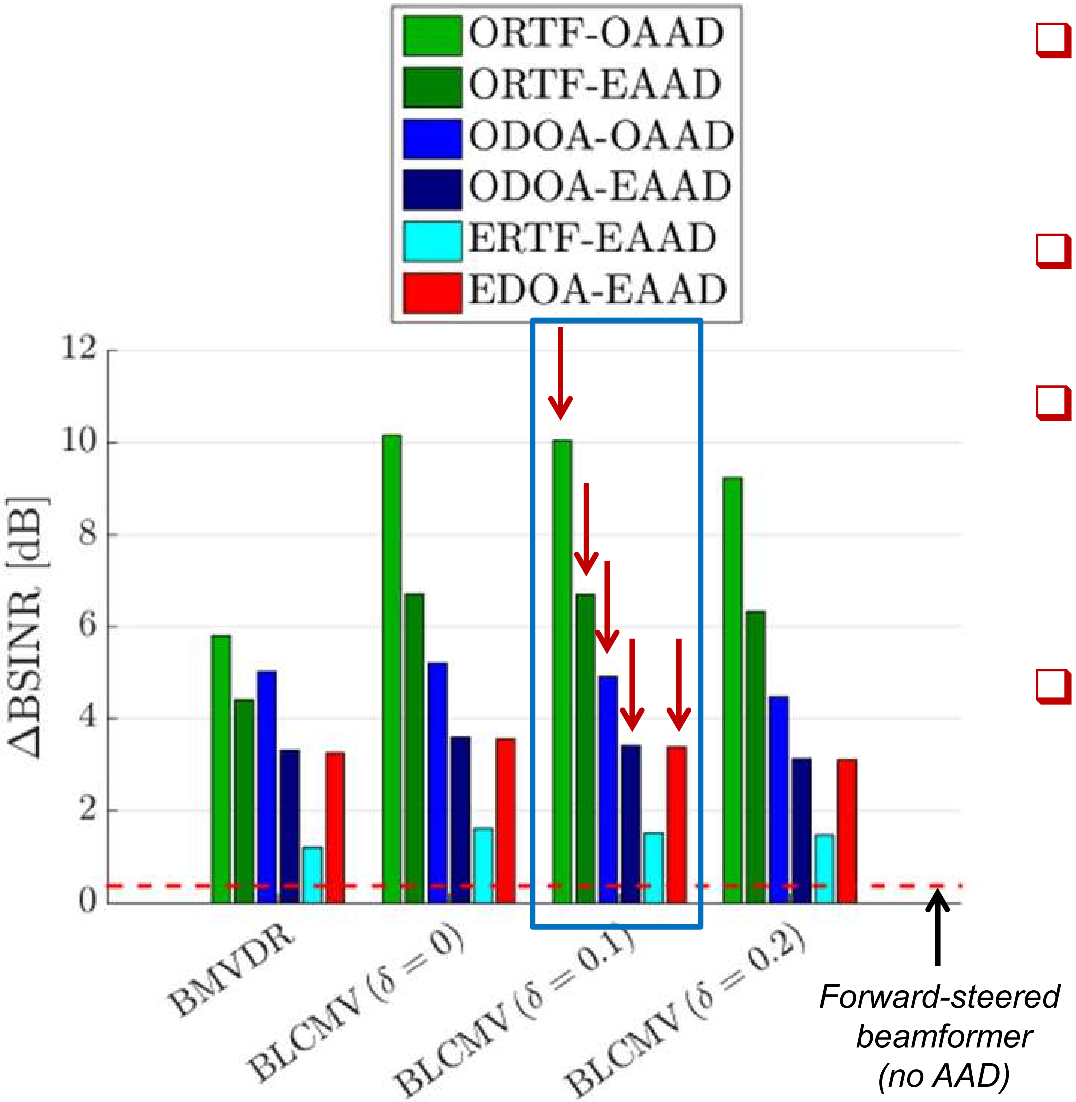


Experimental results: AAD performance



- **Oracle signals** (no noise, no interfering speaker): decoding performance ≈ 89%
- **Unprocessed microphone signals**: decoding performing ≈ 77%
- **Decoding performance larger for LCMV beamformer than for MVDR beamformer,** (larger interference suppression)
  - Best performance when using oracle reverberant RTFs (≈ 87%)
  - Worst performance when using estimated reverberant RTFs
  - **Anechoic RTFs** decrease AAD performance (≈ 82%) compared to reverberant RTFs, but can be **used in practice**

Experimental results: speech enhancement performance



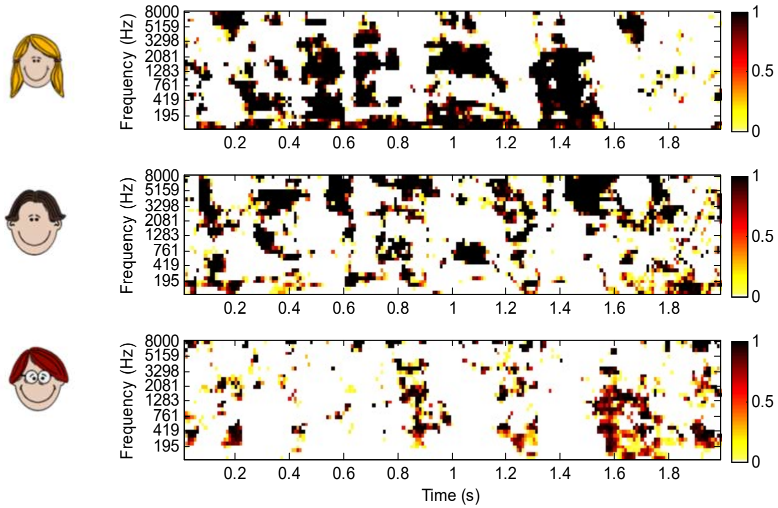
- **Large binaural SINR improvement of cognitive-driven beamformers** compared to forward-steered beamformer
- **Better performance by binaural LCMV beamformer than MVDR beamformer**
- **Oracle AAD:**
  - Best performance when using oracle reverberant RTFs (≈ 10.1 dB)
  - Anechoic RTFs decrease performance (≈ 4.9 dB)
- **Estimated AAD:** AAD errors degrade binaural SINR improvement (attended speaker wrongly suppressed)
  - Best performance when using oracle reverberant RTFs (≈ 6.7 dB)
  - **Anechoic RTFs** decrease AAD performance (≈ 3.2 dB), but can be **used in practice**

Summary

- **Least-squares-based AAD method**
  - **clean speech signals** are not available as reference signals in practice
  - decoding performance significantly decreases when reference signals contain **interfering speaker**
  - **Improved decoding performance using LCMV output signals**
- **Cognitive-driven binaural beamformer system**
  - **Large binaural SINR improvement although AAD errors degrade performance**
  - Better performance by **binaural LCMV beamformer** than MVDR beamformer: larger SINR improvement, controlled suppression of interfering speaker

Next steps to reality…

- **Beamforming:** convolutional LCMV beamforming
- **Acoustic scenarios:** multiple and moving speakers
  → computational acoustic scene analysis (CASA)
- **Decoding:** faster and more reliable
- **Closed-loop system**
- **EEG hardware:** less electrodes (e.g. cEEGrid)



[Aroudi et al., *Proc. IEEE Workshop on Machine Learning for Signal Processing*, 2020.] [Bleichner & Debener, Front. Hum. Neurosci., 2017]