Carl von Ossietzky
Universität
Oldenburg

# Submissions E016, E019, E021

## Combining binaural LCMP beamforming and deep multi-frame filtering for joint dereverberation and interferer reduction in the Clarity-2021 Challenge

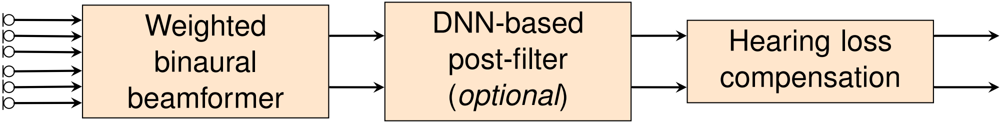
September 17, 2021 –

Marvin Tammen, Henri Gode, Hendrik Kayser,
Eike J. Nustede, Nils L. Westhausen,
Jörn Anemüller, Simon Doclo

Carl von Ossietzky Universität Oldenburg

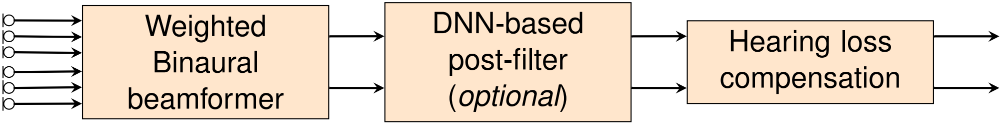
Carl von Ossietzky Universität Oldenburg

# Algorithm Description (Overview)




- reduce late reverberation and interferer
- preserve target speaker

- reduce interferer residuals
- preserve target speaker


Weighted
Binaural
beamformer
DNN-based
post-filter
(*optional*)
Hearing loss
compensation

Carl von Ossietzky Universität Oldenburg

# Motivation and Goal of wBLCMP Beamformer

## MPDR

- Minimizing output power
- Preserve target (distortionless constraint)
  - Target RTF required



Submissions E016, E019, E021
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Motivation and Goal of wBLCMP Beamformer

## wMPDR[1]

- Minimizing output power
- Preserve target (distortionless constraint)
  - Target RTF required
- Dereverberation
  - Multi-frame approach
  - Iterative reweighted optimization (weights $\beta$)





[1] Nakatani and Kinoshita 2019.

# Motivation and Goal of wBLCMP Beamformer

## wLCMP[2]

- Minimizing output power
- Preserve target (distortionless constraint)
  - Target RTF required
- Dereverberation
  - Multi-frame approach
  - Iterative reweighted optimization (weights $\beta$)
- Suppressing interferer with factor $\delta$ ($2^{nd}$ constraint)
  - Interferer RTF required

[2] Aroudi et al. 2020.

# Motivation and Goal of wBLCMP Beamformer

## wLCMP + ADD-ONs[3]

- Minimizing output power
- Preserve target (distortionless constraint)
  - Target RTF required
- Dereverberation
  - Multi-frame approach
  - Iterative reweighted optimization (weights $\beta$)
- Suppressing interferer with factor $\delta$ (2$^{nd}$ constraint)
  - Interferer RTF required
- Controlling the sparsity using shape parameter $p$
- Recursive algorithm with smoothing constant $\gamma$
- Binaural output

[3] Gode, Tammen, and Doclo 2021; Jukić et al. 2015.

# Weighted Binaural LCMP Beamformer

## Optimization Goal

- Minimizing a sparse version of the output power ($\ell_p$-norm)
  - Shape parameter $p$ controls the sparsity

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t}\right) = \sum_{n=1}^{t} \gamma^{t-n} |z_{m,n}|^p = \sum_{n=1}^{t} \gamma^{t-n} \left|\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t\right|^p$$

smoothing factor $\gamma$, shape parameter $p$

Carl von Ossietzky Universität Oldenburg

# Weighted Binaural LCMP Beamformer

## Optimization Goal

- Minimizing a sparse version of the output power ($\ell_p$-norm)
  - Shape parameter $p$ controls the sparsity

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t}\right) = \sum_{n=1}^{t} \gamma^{t-n} |z_{m,n}|^p = \sum_{n=1}^{t} \gamma^{t-n} \left|\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t\right|^p$$

- Linear constraints using RTFs
  - Combined constraint formulation $\quad \bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \tilde{\mathbf{C}} \overset{!}{=} \begin{bmatrix} 1 \\ \delta \end{bmatrix}$

smoothing factor $\gamma$, shape parameter $p$, target RTF $\tilde{\mathbf{a}}_m$, interferer RTF $\tilde{\mathbf{b}}_m$, RTF matrix $\tilde{\mathbf{C}}_m = \left[\tilde{\mathbf{a}}_m \quad \tilde{\mathbf{b}}_m\right]$



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

weights $\beta_{n,i}$, smoothing factor $\gamma$, shape parameter $p$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $$\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$$

weights $\beta_{n,i}$, smoothing factor $\gamma$, shape parameter $p$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$
   - Using weighted covariance matrix $\bar{\mathbf{R}}_{y,t,i} = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^{\mathrm{H}}$

weights $\beta_{n,i}$, smoothing factor $\gamma$, shape parameter $p$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$
   - Using weighted covariance matrix $\bar{\mathbf{R}}_{y,t,i} = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^{\mathrm{H}}$
2. Weight update: $\beta_{n,i+1} = \frac{1}{\left| z_{m,n,i} \right|^{2-p}} = \frac{1}{\left| \bar{\mathbf{w}}_{m,n,i}^{\mathrm{H}} \bar{\mathbf{y}}_n \right|^{2-p}}$

weights $\beta_{n,i}$, smoothing factor $\gamma$, shape parameter $p$

# Practical Implementation of Weighted Binaural LCMP Beamformer

- Estimating RTFs and weighted covariance matrix
  - Interferer RTF estimated within first 2 s
  - Target RTF estimated and updated from 2 s on using covariance whitening

# Practical Implementation of Weighted Binaural LCMP Beamformer

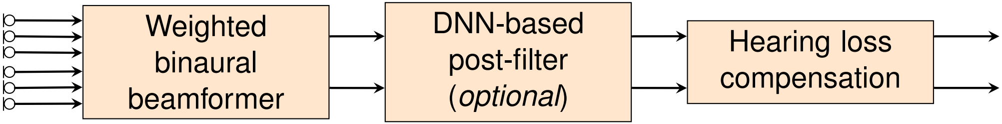
- Estimating RTFs and weighted covariance matrix
  - Interferer RTF estimated within first 2 s
  - Target RTF estimated and updated from 2 s on using covariance whitening

- Algorithm and framework parameters
  - Frame length: 5 ms
  - Frame shift: 2.5 ms

# Practical Implementation of Weighted Binaural LCMP Beamformer

- Estimating RTFs and weighted covariance matrix
  - Interferer RTF estimated within first 2 s
  - Target RTF estimated and updated from 2 s on using covariance whitening

- Algorithm and framework parameters
  - Frame length: 5 ms
  - Frame shift: 2.5 ms
  - Interferer suppression: $\delta = 0.1$
  - Smoothing constant: $\gamma = 1$
  - Shape parameter: $p = 0.5$


Weighted binaural beamformer
DNN-based post-filter (*optional*)
Hearing loss compensation

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |
| utilized correlations | spatial | temporal |

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |
| utilized correlations | spatial | temporal |
| speech component | $\mathbf{x}_{f,t} = S_{f,t}\mathbf{v}_{f,t}$ | $\mathbf{x}_{f,t} = S_{f,t}\gamma_{f,t}$ |

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |
| utilized correlations | spatial | temporal |
| speech component | $\mathbf{x}_{f,t} = S_{f,t}\mathbf{v}_{f,t}$ | $\mathbf{x}_{f,t} = S_{f,t}\boldsymbol{\gamma}_{f,t}$ |
| | $\mathbf{v}_{f,t}$: signal-independent, easier to estimate | $\boldsymbol{\gamma}_{f,t}$: signal-dependent, more difficult to estimate |



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |
| utilized correlations | spatial | temporal |
| speech component | $\mathbf{x}_{f,t} = S_{f,t}\mathbf{v}_{f,t}$<br>$\mathbf{v}_{f,t}$: signal-independent, easier to estimate | $\mathbf{x}_{f,t} = S_{f,t}\gamma_{f,t}$<br>$\gamma_{f,t}$: signal-dependent, more difficult to estimate |
| optimization goal | minimize undesired power while preserving *spatial* speech correlations $\mathbf{v}_{f,t}$ | minimize undesired power while preserving *temporal* speech correlations $\gamma_{f,t}$ |

# Deep Binaural MFMVDR Filter: Motivation

| | MVDR | multi-frame MVDR |
|---|---|---|
| | multi-microphone, single-frame | single-microphone, multi-frame |
| utilized correlations | spatial | temporal |
| speech component | $\mathbf{x}_{f,t} = S_{f,t}\mathbf{v}_{f,t}$<br>$\mathbf{v}_{f,t}$: signal-independent, easier to estimate | $\mathbf{x}_{f,t} = S_{f,t}\gamma_{f,t}$<br>$\gamma_{f,t}$: signal-dependent, more difficult to estimate |
| optimization goal | minimize undesired power while preserving *spatial* speech correlations $\mathbf{v}_{f,t}$ | minimize undesired power while preserving *temporal* speech correlations $\gamma_{f,t}$ |
| binaural extension | preserve speech correlations at left and right ear | |

# Deep Binaural MFMVDR Filter

## Optimization Goal

- **minimize undesired power spectral density** while

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \underset{\mathbf{w}^{\{l,r\}}}{\operatorname{argmin}}\, \mathbf{w}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{n}}^{\{l,r\}} \mathbf{w}^{\{l,r\}}$$

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter

## Optimization Goal

- **minimize undesired power spectral density** while **preserving** left and right **correlated speech components**:

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \underset{\mathbf{w}^{\{l,r\}}}{\operatorname{argmin}}\, \mathbf{w}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{n}}^{\{l,r\}} \mathbf{w}^{\{l,r\}}, \quad \mathrm{s.t.} \quad \mathbf{w}^{\{l,r\}^H} \gamma_{\mathbf{x}}^{\{l,r\}} = 1$$



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter

## Optimization Goal

- **minimize undesired power spectral density** while **preserving** left and right **correlated speech components**:

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \underset{\mathbf{w}^{\{l,r\}}}{\operatorname{argmin}}\, \mathbf{w}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{n}}^{\{l,r\}} \mathbf{w}^{\{l,r\}}, \quad \text{s.t.} \quad \mathbf{w}^{\{l,r\}^H} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}} = 1$$

$$= \frac{\mathbf{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}{\boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}$$

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter

## Optimization Goal

- **minimize undesired power spectral density** while **preserving** left and right **correlated speech components**:

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \underset{\mathbf{w}^{\{l,r\}}}{\operatorname{argmin}}\, \mathbf{w}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{n}}^{\{l,r\}} \mathbf{w}^{\{l,r\}}, \quad \text{s.t.} \quad \mathbf{w}^{\{l,r\}^H} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}} = 1$$

$$= \frac{\mathbf{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}{\boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}^H} \mathbf{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}$$

- $\mathbf{\Phi}_{\mathbf{u}}^{\{l,r\}}$, $\boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}$ estimated using DNNs

# Deep Binaural MFMVDR Filter

## Optimization Goal

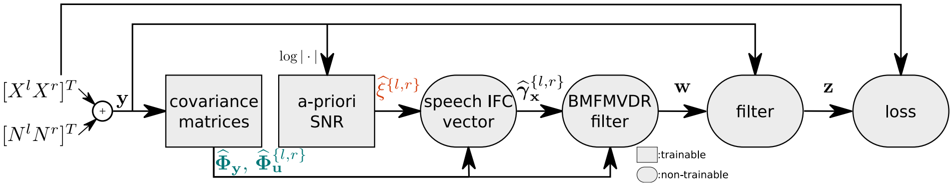
- **minimize undesired power spectral density** while **preserving** left and right **correlated speech components**:

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \underset{\mathbf{w}^{\{l,r\}}}{\operatorname{argmin}}\, \mathbf{w}^{\{l,r\}^H} \boldsymbol{\Phi}_{\mathbf{n}}^{\{l,r\}} \mathbf{w}^{\{l,r\}}, \quad \mathrm{s.t.} \quad \mathbf{w}^{\{l,r\}^H} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}} = 1$$

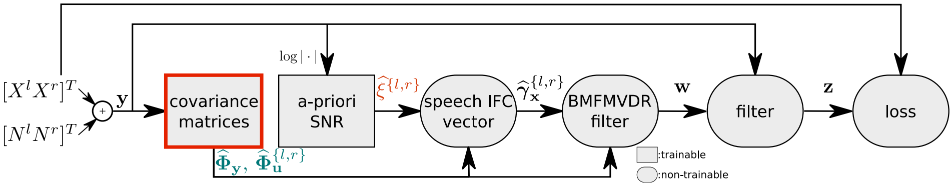
$$= \frac{\boldsymbol{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}{\boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}^H} \boldsymbol{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}}$$

- $\boldsymbol{\Phi}_{\mathbf{u}}^{\{l,r\}}$, $\boldsymbol{\gamma}_{\mathbf{x}}^{\{l,r\}}$ estimated using DNNs
  - binaural extension of Tammen and Doclo 2021

# Deep Binaural MFMVDR Filter: Block Diagram



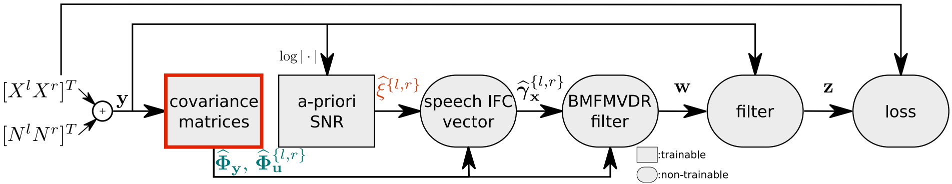
# Deep Binaural MFMVDR Filter: Block Diagram




## covariance matrices $\widehat{\mathbf{\Phi}}_{\mathbf{y}}$ and $\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\{l,r\}}$

- DNN inputs: concatenated (binaural) STFT log magnitude & cos of phase

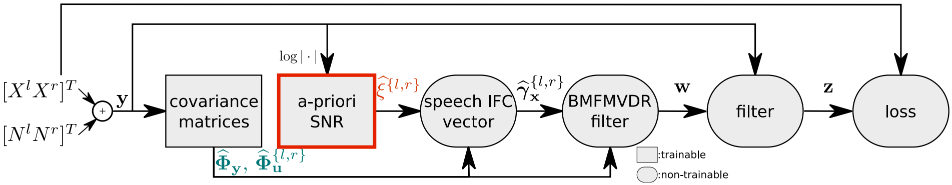
# Deep Binaural MFMVDR Filter: Block Diagram




## covariance matrices $\widehat{\mathbf{\Phi}}_{\mathbf{y}}$ and $\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\{l,r\}}$

- DNN inputs: concatenated (binaural) STFT log magnitude & cos of phase
- DNN is **trained using speech enhancement loss**

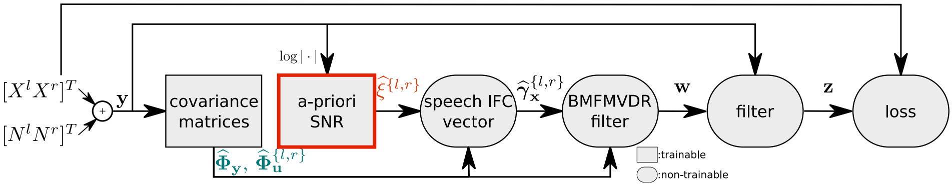
# Deep Binaural MFMVDR Filter: Block Diagram




## a-priori SNR $\widehat{\xi}^{\{l,r\}}$

- DNN inputs: logarithm of noisy STFT magnitudes

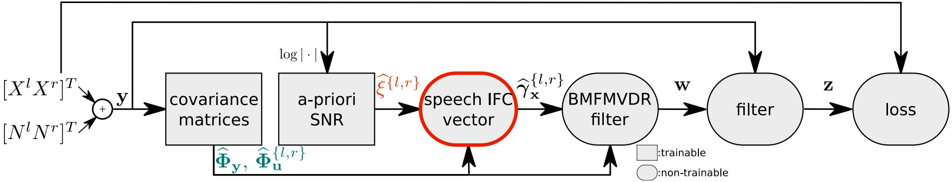
# Deep Binaural MFMVDR Filter: Block Diagram




## a-priori SNR $\widehat{\xi}^{\{l,r\}}$

- DNN inputs: logarithm of noisy STFT magnitudes
- DNN is **trained using speech enhancement loss**

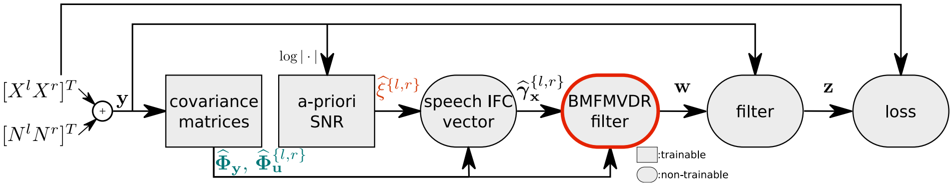
# Deep Binaural MFMVDR Filter: Block Diagram




## speech inter-frame correlation (IFC) vector $\widehat{\gamma}_{\mathbf{x}}^{\{l,r\}}$

$$\widehat{\gamma}_{\mathbf{x}}^{\{l,r\}} = \frac{1+\widehat{\xi}^{\{l,r\}}}{\widehat{\xi}^{\{l,r\}}} \frac{\widehat{\mathbf{\Phi}}_{\mathbf{y}} \mathbf{e}^{\{l,r\}}}{\mathbf{e}^{\{l,r\}\top} \widehat{\mathbf{\Phi}}_{\mathbf{y}} \mathbf{e}^{\{l,r\}}} - \frac{1}{\widehat{\xi}^{\{l,r\}}} \frac{\widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\{l,r\}} \mathbf{e}^{\{l,r\}}}{\mathbf{e}^{\{l,r\}\top} \widehat{\mathbf{\Phi}}_{\mathbf{u}}^{\{l,r\}} \mathbf{e}^{\{l,r\}}}$$

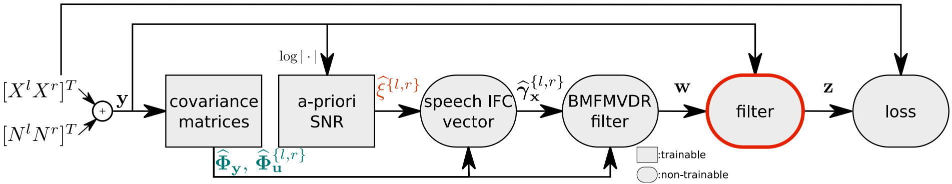
# Deep Binaural MFMVDR Filter: Block Diagram




## deep binaural MFMVDR filter

$$\mathbf{w}^{\mathrm{BMFMVDR}\{l,r\}} = \frac{\widehat{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \widehat{\gamma}_{\mathbf{x}}^{\{l,r\}}}{\widehat{\gamma}_{\mathbf{x}}^{\{l,r\}^H} \widehat{\Phi}_{\mathbf{u}}^{\{l,r\}^{-1}} \widehat{\gamma}_{\mathbf{x}}^{\{l,r\}}}$$

# Deep Binaural MFMVDR Filter: Block Diagram




## filter

$$\mathbf{z} = \begin{bmatrix} Z^l \\ Z^r \end{bmatrix} = \begin{bmatrix} \mathbf{w}^{\mathrm{BMFMVDR}\{l\}^H}\mathbf{y} \\ \mathbf{w}^{\mathrm{BMFMVDR}\{r\}^H}\mathbf{y} \end{bmatrix}$$




**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP





[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context





[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context
- DNN architecture:

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\mathrm{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]





[4] Bai, Kolter, and Koltun 2018.

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\text{s} > 2\,\text{s}$

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\mathrm{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\mathrm{s} > 2\,\mathrm{s}$
  - 2 stacks of 8 layers, kernel size $3 \rightarrow 3.02\,\mathrm{M}$ parameters

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\text{s} > 2\,\text{s}$
  - 2 stacks of 8 layers, kernel size $3 \rightarrow 3.02\,\text{M}$ parameters
- DNN training:

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\text{s} > 2\,\text{s}$
  - 2 stacks of 8 layers, kernel size $3 \rightarrow 3.02\,\text{M}$ parameters
- DNN training:
  - official Clarity-2021 training data

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\mathrm{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\mathrm{s} > 2\,\mathrm{s}$
  - 2 stacks of 8 layers, kernel size $3 \rightarrow 3.02\,\mathrm{M}$ parameters
- DNN training:
  - official Clarity-2021 training data
  - scale-invariant signal-to-distortion-ratio loss function

[4] Bai, Kolter, and Koltun 2018.

# Deep Binaural MFMVDR Filter: Settings

- same STFT parameters as wBLCMP
- filter length $N = 4 \rightarrow 12.5\,\text{ms}$ context
- DNN architecture:
  - causal temporal convolutional network[4]
  - receptive field size of $2.56\,\text{s} > 2\,\text{s}$
  - 2 stacks of 8 layers, kernel size $3 \rightarrow 3.02\,\text{M}$ parameters
- DNN training:
  - official Clarity-2021 training data
  - scale-invariant signal-to-distortion-ratio loss function
  - AdamW optimizer, learning rate $10^{-3}$, 67 epochs, batch size 4

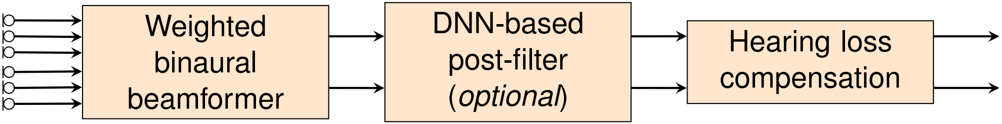
[4] Bai, Kolter, and Koltun 2018.

1 Algorithm Description

- Overview
- Weighted Binaural LCMP Beamformer
- Deep Binaural Multi-Frame MVDR Filter
- Hearing Loss Compensation

2 Experimental Results

3 Conclusions



Carl von Ossietzky Universität Oldenburg

# Hearing Loss Compensation

1. submission E016:
   "half-gain rule (HGR)": simple hearing loss (HL)-dependent **broadband gain**, i.e., $\mathrm{HGR} = \frac{1}{6}(HL_{500\,\mathrm{Hz}} + HL_{1000\,\mathrm{Hz}} + HL_{2000\,\mathrm{Hz}})$





[5] Moore et al. 1999.

# Hearing Loss Compensation

1. submission E016:
   "half-gain rule (HGR)": simple hearing loss (HL)-dependent **broadband gain**, i.e., $\mathrm{HGR} = \frac{1}{6}(HL_{500\,\mathrm{Hz}} + HL_{1000\,\mathrm{Hz}} + HL_{2000\,\mathrm{Hz}})$
2. submissions E019, E021:
   baseline multi-band dynamic range compressor (DRC): **frequency- and HL-dependent gain**, compressive Camfit prescription rule[5]

[5] Moore et al. 1999.

# Algorithms Summary

| | E016 | E019 | E021 |
|---|---|---|---|
| beamformer | ✓ | ✓ | ✓ |
| deep post-filter | × | × | ✓ |
| HL compensation | broadband gain | baseline DRC | |

Carl von Ossietzky Universität Oldenburg

# Experimental Results – Objective Evaluation

- all submitted systems outperform baseline system
- insignificant differences between submitted algorithms

| | E016 | E019 | E021 |
|---|---|---|---|
| beamformer | ✓ | ✓ | ✓ |
| deep post-filter | × | × | ✓ |
| HL compensation | broadband gain | baseline DRC | |

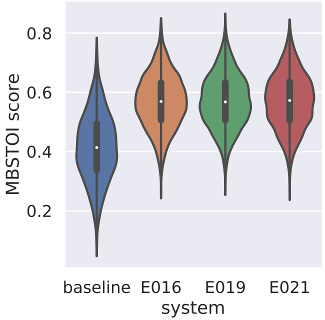



Figure: hearing loss-dependent MBSTOI results, development dataset



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Experimental Results – Objective Evaluation

- all submitted systems outperform baseline system
- insignificant differences between submitted algorithms

| | E016 | E019 | E021 |
|---|---|---|---|
| beamformer | ✓ | ✓ | ✓ |
| deep post-filter | × | × | ✓ |
| HL compensation | broadband gain | baseline DRC | |

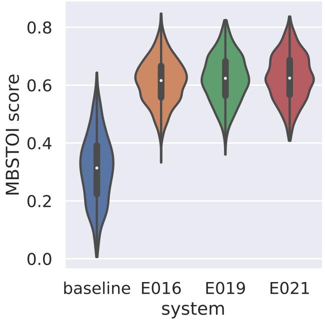



Figure: hearing loss-dependent MBSTOI results, evaluation dataset



Submissions E016, E019, E021

# Experimental Results – Subjective Evaluation

- all submitted systems outperform baseline system
- drastic drop in correctness for system E021 in some interfering speaker scenarios

| | **E019** | **E021** |
|---|---|---|
| beamformer | ✓ | ✓ |
| deep post-filter | × | ✓ |
| HL compensation | baseline DRC | |

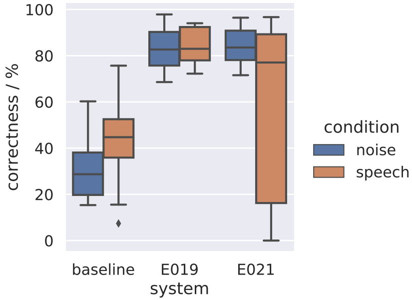



Figure: listening test results, evaluation dataset

Carl von Ossietzky Universität Oldenburg

# Experimental Results – Subjective Evaluation

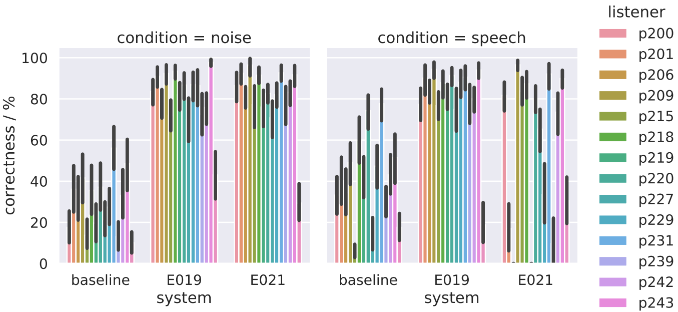



Figure: listening test results per listener, evaluation dataset

# Experimental Results – Audio Demos

| system | | # | interfering speech | interfering noise |
|---|---|---|---|---|
| input signals | | 1 | ▷ | ▷ |
| baseline | | 2 | ▷ | ▷ |
| E016 | beamformer + broadband gain | 3 | ▷ | ▷ |
| E019 | beamformer + baseline DRC | 4 | ▷ | ▷ |
| E021 | beamformer + deep post-filter + baseline DRC | 5 | ▷ | ▷ |
| | | | S08501, L104 | S08502, L106 |

# Conclusions

- proposed combinations of beamformer, DNN-based post-filter, and hearing loss compensation modules

# Conclusions

- proposed combinations of beamformer, DNN-based post-filter, and hearing loss compensation modules
- all submitted systems achieved considerable objective and subjective intelligibility improvements, outperforming baseline system

# Conclusions

- proposed combinations of beamformer, DNN-based post-filter, and hearing loss compensation modules
- all submitted systems achieved considerable objective and subjective intelligibility improvements, outperforming baseline system
- MBSTOI scores:
  similar performance of all systems

Carl von Ossietzky Universität Oldenburg

# Conclusions

- proposed combinations of beamformer, DNN-based post-filter, and hearing loss compensation modules
- all submitted systems achieved considerable objective and subjective intelligibility improvements, outperforming baseline system
- MBSTOI scores:
  similar performance of all systems
- listening test results:
  DNN-based post-filter may degrade performance in some scenarios

# Conclusions

- proposed combinations of beamformer, DNN-based post-filter, and hearing loss compensation modules
- all submitted systems achieved considerable objective and subjective intelligibility improvements, outperforming baseline system
- MBSTOI scores:
  similar performance of all systems
- listening test results:
  DNN-based post-filter may degrade performance in some scenarios

Thank you for your attention!

# 4 Backup Slides

# Weighted Binaural LCMP Beamformer

- Convolutional signal model per STFT-bin (frequency index omitted)

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{n}_t = \underbrace{\sum_{l=0}^{L-1} \mathbf{a}_l s_{x,t-l}}_{\text{target}} + \underbrace{\sum_{l=0}^{L-1} \mathbf{b}_l s_{n,t-l}}_{\text{interferer}} = \underbrace{\sum_{l=0}^{\tau-1} \mathbf{C}_l \begin{bmatrix} s_{x,t-l} \\ s_{n,t-l} \end{bmatrix}}_{\text{direct/early}} + \underbrace{\sum_{l=\tau}^{L-1} \mathbf{C}_l \begin{bmatrix} s_{x,t-l} \\ s_{n,t-l} \end{bmatrix}}_{\text{reverb}}$$

Filter length $L$, prediction delay $\tau$, target clean speech $s_{x,t}$, interferer clean signal $s_{n,t}$, target ATF $\mathbf{a}_l$, interferer ATF $\mathbf{b}_l$,

ATF matrix $\mathbf{C}_l = [\mathbf{a}_l \quad \mathbf{b}_l]$

Carl von Ossietzky Universität Oldenburg

# Weighted Binaural LCMP Beamformer

- Convolutional signal model per STFT-bin (frequency index omitted)

$$\boxed{\mathbf{y}_t = \mathbf{x}_t + \mathbf{n}_t = \underbrace{\sum_{l=0}^{L-1} \mathbf{a}_l s_{x,t-l}}_{\text{target}} + \underbrace{\sum_{l=0}^{L-1} \mathbf{b}_l s_{n,t-l}}_{\text{interferer}} = \underbrace{\sum_{l=0}^{\tau-1} \mathbf{C}_l \begin{bmatrix} s_{x,t-l} \\ s_{n,t-l} \end{bmatrix}}_{\text{direct/early}} + \underbrace{\sum_{l=\tau}^{L-1} \mathbf{C}_l \begin{bmatrix} s_{x,t-l} \\ s_{n,t-l} \end{bmatrix}}_{\text{reverb}}}$$

- Convolutional beamformer $\bar{\mathbf{w}}_{m,t}$

$$\boxed{z_{m,t} = \bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t} \quad \text{with} \quad \bar{\mathbf{y}}_t = \begin{bmatrix} \mathbf{y}_t^{\mathrm{T}} & \mathbf{y}_{t-\tau}^{\mathrm{T}} & \mathbf{y}_{t-\tau-1}^{\mathrm{T}} & \cdots & \mathbf{y}_{t-L+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$$

Filter length $L$, prediction delay $\tau$, target clean speech $s_{x,t}$, interferer clean signal $s_{n,t}$, target ATF $\mathbf{a}_l$, interferer ATF $\mathbf{b}_l$,

ATF matrix $\mathbf{C}_l = \begin{bmatrix} \mathbf{a}_l & \mathbf{b}_l \end{bmatrix}$

# Weighted Binaural LCMP Beamformer

## Optimization Goal

- Minimizing a sparse version of the output power ($\ell_p$-norm)
  - Shape parameter $p$ controls the sparsity

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t}\right) = \sum_{n=1}^{t} \gamma^{t-n} |z_{m,n}|^p = \sum_{n=1}^{t} \gamma^{t-n} \left|\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t\right|^p$$

smoothing factor $\gamma$, shape parameter $p$

Carl von Ossietzky Universität Oldenburg

# Weighted Binaural LCMP Beamformer

## Optimization Goal

- Minimizing a sparse version of the output power ($\ell_p$-norm)
  - Shape parameter $p$ controls the sparsity

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t}\right) = \sum_{n=1}^{t} \gamma^{t-n} |z_{m,n}|^p = \sum_{n=1}^{t} \gamma^{t-n} \left|\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t\right|^p$$

- Linear constraints using RTFs
  - Distortionless constraint for target $\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \tilde{\mathbf{a}} \overset{!}{=} 1$
  - Suppressing constraint for interferer $\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \tilde{\mathbf{b}} \overset{!}{=} \delta$

smoothing factor $\gamma$, shape parameter $p$, target RTF $\tilde{\mathbf{a}}_m$, interferer RTF $\tilde{\mathbf{b}}_m$

Carl von Ossietzky Universität Oldenburg

# Weighted Binaural LCMP Beamformer

## Optimization Goal

- Minimizing a sparse version of the output power ($\ell_p$-norm)
  - Shape parameter $p$ controls the sparsity

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t}\right) = \sum_{n=1}^{t} \gamma^{t-n} |z_{m,n}|^p = \sum_{n=1}^{t} \gamma^{t-n} \left|\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \bar{\mathbf{y}}_t\right|^p$$

- Linear constraints using RTFs
  - Combined constraint formulation $\bar{\mathbf{w}}_{m,t}^{\mathrm{H}} \tilde{\mathbf{C}} \stackrel{!}{=} \begin{bmatrix} 1 \\ \delta \end{bmatrix}$

smoothing factor $\gamma$, shape parameter $p$, target RTF $\tilde{\mathbf{a}}_m$, interferer RTF $\tilde{\mathbf{b}}_m$, RTF matrix $\tilde{\mathbf{C}}_m = [\tilde{\mathbf{a}}_m \quad \tilde{\mathbf{b}}_m]$

Carl von Ossietzky Universität Oldenburg

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

weights $\beta_{n,i}$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$

weights $\beta_{n,i}$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$
   - Using weighted covariance matrix $\bar{\mathbf{R}}_{y,t,i} = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^{\mathrm{H}}$

weights $\beta_{n,i}$

# Weighted Binaural LCMP Beamformer Using IRLS

## Iterative solution

- Reweighted cost function of $\ell_2$-norm subproblem

$$\mathcal{C}\left(\bar{\mathbf{w}}_{m,t,i}\right) = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \left| z_{m,n,i} \right|^2$$

1. Filter update: $\boxed{\bar{\mathbf{w}}_{m,t,i} = \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \left[ \tilde{\mathbf{C}}_m^{\mathrm{H}} \bar{\mathbf{R}}_{y,t,i}^{-1} \tilde{\mathbf{C}}_m \right]^{-1} \begin{bmatrix} 1 \\ \delta \end{bmatrix}}$
   - Using weighted covariance matrix $\bar{\mathbf{R}}_{y,t,i} = \sum_{n=1}^{t} \gamma^{t-n} \beta_{n,i} \bar{\mathbf{y}}_n \bar{\mathbf{y}}_n^{\mathrm{H}}$
2. Weight update: $\beta_{n,i+1} = \frac{1}{\left| z_{m,n,i} \right|^{2-p}} = \frac{1}{\left| \bar{\mathbf{w}}_{m,n,i}^{\mathrm{H}} \bar{\mathbf{y}}_n \right|^{2-p}}$

weights $\beta_{n,i}$



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\mathbf{\Phi}}_{\nu} \in \mathbb{C}^{2N\times 2N}$

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} \quad \end{bmatrix}$$



**Submissions E016, E019, E021**

M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\mathbf{\Phi}}_{\nu} \in \mathbb{C}^{2N \times 2N}$
fill main diagonal ($2N$ real-valued coefficients);

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix}$$



Submissions E016, E019, E021
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

Carl von Ossietzky Universität Oldenburg

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\mathbf{\Phi}}_{\nu} \in \mathbb{C}^{2N \times 2N}$
fill main diagonal ($2N$ real-valued coefficients);
fill strictly upper triangle with real components ($\frac{2N(2N+1)}{2} = \frac{(2N)^2-2N}{2}$ real-valued coefficients);

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} a & d & f \\ & b & e \\ & & c \end{bmatrix}$$

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\boldsymbol{\Phi}}_{\nu} \in \mathbb{C}^{2N\times 2N}$
fill main diagonal ($2N$ real-valued coefficients);
fill strictly upper triangle with real components ($\frac{2N(2N+1)}{2} = \frac{(2N)^2-2N}{2}$ real-valued coefficients);
fill strictly upper triangle with imaginary components (also $\frac{(2N)^2-2N}{2}$ real-valued coefficients);

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} a & d+gj & f+ij \\ & b & e+hj \\ & & c \end{bmatrix}$$



**Submissions E016, E019, E021**
M Tammen, H Gode, H Kayser, EJ Nustede, NL Westhausen, J Anemüller, S Doclo – University of Oldenburg

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\boldsymbol{\Phi}}_{\nu} \in \mathbb{C}^{2N \times 2N}$
fill main diagonal ($2N$ real-valued coefficients);
fill strictly upper triangle with real components ($\frac{2N(2N+1)}{2} = \frac{(2N)^2 - 2N}{2}$ real-valued coefficients);
fill strictly upper triangle with imaginary components (also $\frac{(2N)^2 - 2N}{2}$ real-valued coefficients);
make Hermitian: $\mathbf{H}_{\nu} \leftarrow \mathbf{H}_{\nu} + \mathbf{H}_{\nu}^{H}$;

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$
$$\downarrow$$
$$\begin{bmatrix} 2a + 0j & d + gj & f + ij \\ d - gj & 2b + 0j & e + hj \\ f - ij & e - hj & 2c + 0j \end{bmatrix}$$

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\nu} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\mathbf{\Phi}}_{\nu} \in \mathbb{C}^{2N \times 2N}$
fill main diagonal ($2N$ real-valued coefficients);
fill strictly upper triangle with real components ($\frac{2N(2N+1)}{2} = \frac{(2N)^2-2N}{2}$ real-valued coefficients);
fill strictly upper triangle with imaginary components (also $\frac{(2N)^2-2N}{2}$ real-valued coefficients);
make Hermitian: $\mathbf{H}_{\nu} \leftarrow \mathbf{H}_{\nu} + \mathbf{H}_{\nu}^{H}$;
make PSD: $\widehat{\mathbf{\Phi}}_{\nu} = \mathbf{H}_{\nu}\mathbf{H}_{\nu}^{H}$;

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$
$$\downarrow$$
$$\begin{bmatrix} 2a+0j & d+gj & f+ij \\ d-gj & 2b+0j & e+hj \\ f-ij & e-hj & 2c+0j \end{bmatrix}$$

# Deep Binaural MFMVDR Filter: Covariance Matrices

**Input:** DNN output vector $\mathbf{h}_{\boldsymbol{\nu}} \in \mathbb{R}^{4N^2}$
**Output:** Hermitian-PSD matrix $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{\nu}} \in \mathbb{C}^{2N \times 2N}$
fill main diagonal ($2N$ real-valued coefficients);
fill strictly upper triangle with real components ($\frac{2N(2N+1)}{2} = \frac{(2N)^2-2N}{2}$ real-valued coefficients);
fill strictly upper triangle with imaginary components (also $\frac{(2N)^2-2N}{2}$ real-valued coefficients);
make Hermitian: $\mathbf{H}_{\boldsymbol{\nu}} \leftarrow \mathbf{H}_{\boldsymbol{\nu}} + \mathbf{H}_{\boldsymbol{\nu}}^H$;
make PSD: $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{\nu}} = \mathbf{H}_{\boldsymbol{\nu}}\mathbf{H}_{\boldsymbol{\nu}}^H$;
$\rightarrow$ in total, $2N + 2\frac{(2N)^2-2N}{2} = 4N^2$ real-valued coefficients required

$$\begin{bmatrix} a \\ \vdots \\ i \end{bmatrix}$$
$$\downarrow$$
$$\begin{bmatrix} 2a+0j & d+gj & f+ij \\ d-gj & 2b+0j & e+hj \\ f-ij & e-hj & 2c+0j \end{bmatrix}$$

# Hearing Loss Compensation Parameters

| | E016 | E019 | E021 |
|---|---|---|---|
| output gain (dB) | HGR | 10 | 10 |
| maximum output level (dB) | — | 120 | 120 |
| soft clipper attack time (s) | — | 0.002 | 0.001 |
| soft clipper decay time (s) | — | 0.01 | 0.01 |
| soft clipper threshold (dB) | 117 | 117 | 117 |