Generalized Coherence-based Signal Enhancement

Heinrich Löllmann, Andreas Brendel and Walter Kellermann

Chair of Multimedia Communications and Signal Processing (LMS)

Friedrich-Alexander University Erlangen-Nürnberg (FAU)

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

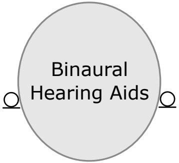
ICASSP2020
Barcelona

LMS

Introduction

Signal enhancement based on the coherent-to-diffuse power ratio (CDR)

- Well-established concept for two-channel speech dereverberation and noise reduction, e.g., [Habets 2007, Jeub 2012, Schwarz 2015]
- Beneficial for signal enhancement in hearing aids or mobile phones
- Algorithms for such applications have to cope with
  - heterogeneous signal quality at each microphone
  - erroneous or missing TDOA information about the desired speaker

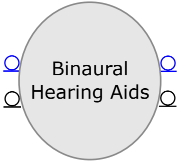


→ **New multichannel CDR-based signal enhancement approach for such a scenario will be presented**

Introduction

Signal enhancement based on the coherent-to-diffuse power ratio (CDR)

- Well-established concept for two-channel speech dereverberation and noise reduction, e.g., [Habets 2007, Jeub 2012, Schwarz 2015]
- Beneficial for signal enhancement in hearing aids or mobile phones
- Algorithms for such applications have to cope with
  - heterogeneous signal quality at each microphone
  - erroneous or missing TDOA information about the desired speaker



➜ **New multichannel CDR-based signal enhancement approach for such a scenario will be presented**

Overview

- Introduction
- CDR-based Speech Enhancement
- New N-channel CDR-based Speech Enhancement
- Evaluation
- Conclusions

Signal Model

Noisy reverberant speech at microphone i

$$x_i(k) = \sum_{\kappa=0}^{L-1} s(\kappa)\, h_i(k-\kappa) + n_i^{(\text{diff})}(k)$$

$$= \underbrace{\sum_{\kappa=0}^{K_{\text{early}}-1} s(\kappa)\, h_i(k-\kappa)}_{\text{desired speech } d_i(k)} + \underbrace{\sum_{\kappa=K_{\text{early}}}^{L-1} s(\kappa)\, h_i(k-\kappa) + n_i^{(\text{diff})}(k)}_{\text{reverberant speech } + \text{ diffuse noise } n_i^{(\text{diff})}(k)}$$

$s(k)$: speech signal
$h_i(k)$: room impulse response (RIR)
$n_i^{(\text{diff})}(k)$: diffuse noise
$i = 1, \ldots, N$: microphone index
K_{early} : onset of late reverberation

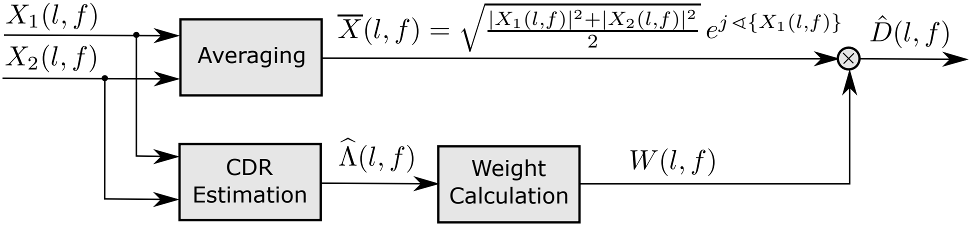
Signal Model

Coherent-to-diffuse power ratio (CDR)

$$\Lambda_i(l, f) = \frac{\Phi_{d_i,d_i}(l, f)}{\Phi_{n_i,n_i}(l, f)}$$

$\Phi_{d_i,d_i}(l, f)$: short time power spectral density (PSD) of desired speech
$\Phi_{n_i,n_i}(l, f)$: PSD of reverberant speech and diffuse noise
l : frame index
f : frequency

CDR-based Speech Enhancement



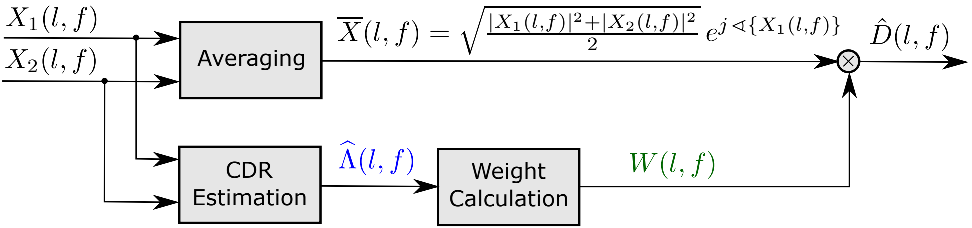
- Speech enhancement by common noise suppression schemes, e.g.,

$$W(l,f) = \max\left\{W_{\text{min}}, \left(1-\left(\frac{\mu}{\widehat{\Lambda}(l,f)+1}\right)^{\beta}\right)^{\alpha}\right\}$$

W_{min}: minimum weight, μ oversubtraction factor, α, β: determines weighting rule

- Implicit channel selection by choice of phase term $\sphericalangle\{X_1(l,f)\}$ for averaging operation

CDR-based Speech Enhancement



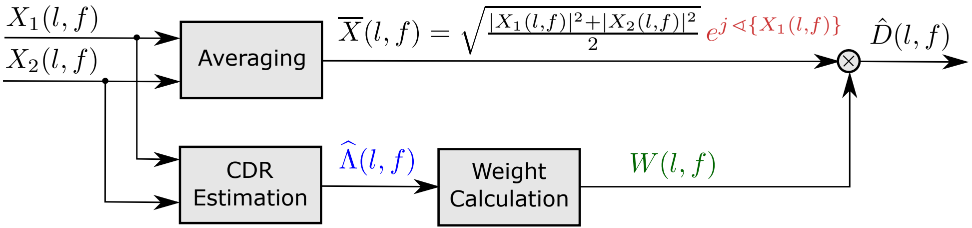
- Speech enhancement by common noise suppression schemes, e.g.,

$$W(l,f) = \max\left\{W_{\text{min}},\left(1-\left(\frac{\mu}{\widehat{\Lambda}(l,f)+1}\right)^{\beta}\right)^{\alpha}\right\}$$

W_{min}: minimum weight, μ oversubtraction factor, α, β: determines weighting rule

- Implicit channel selection by choice of phase term $\sphericalangle\{X_1(l,f)\}$ for averaging operation

CDR-based Speech Enhancement



- Speech enhancement by common noise suppression schemes, e.g.,

$$W(l,f) = \max\left\{W_{\text{min}}, \left(1 - \left(\frac{\mu}{\widehat{\Lambda}(l,f)+1}\right)^{\beta}\right)^{\alpha}\right\}$$

W_{min}: minimum weight, μ oversubtraction factor, α, β: determines weighting rule

- Implicit channel selection by choice of phase term $\sphericalangle\{X_1(l,f)\}$ for averaging operation

CDR Estimation

- Short time CDR for two microphone signals

$$\Lambda_{1,2}(l,f) = \frac{\Gamma_{n_1,n_2}(l,f) - \Gamma_{x_1,x_2}(l,f)}{\Gamma_{x_1,x_2}(l,f) - \Gamma_{d_1,d_2}(l,f)}$$

➜ Might become complex-valued due to model and/or estimation errors

➜ Solved by heuristic modifications, e.g., [Thiergart 2012]

$$\Lambda_{1,2}^{(\mathrm{Th})}(l,f) = \mathrm{Re}\left\{ \frac{\widehat{\Gamma}_{n_1,n_2}(l,f) - \widehat{\Gamma}_{x_1,x_2}(l,f)}{\widehat{\Gamma}_{x_1,x_2}(l,f) - e^{j\sphericalangle\left\{\widehat{\Gamma}_{x_1,x_2}(l,f)\right\}}} \right\}$$

How to exploit $N \geq 2$ microphone signals for CDR estimation?

New CDR Estimator

Based on the generalized magnitude-squared coherence (GMSC) [Ramirez 2008]

Generalized Magnitude Coherence (GMC)

$$\gamma_x(l,f) = \frac{1}{N-1}\left(\lambda_x^{(\mathsf{max})}(l,f) - 1\right)$$

$\lambda_x^{(\mathsf{max})}(l,f)$: largest eigenvalue of the $N \times N$ coherence matrix

$$\boldsymbol{C}_x(l,f) = \begin{bmatrix} 1 & \Gamma_{x_1,x_2}(l,f) & \dots & \Gamma_{x_1,x_N}(l,f) \\ \Gamma_{x_2,x_1}(l,f) & 1 & \dots & \Gamma_{x_2,x_N}(l,f) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{x_N,x_1}(l,f) & \Gamma_{x_N,x_2}(l,f) & \dots & 1 \end{bmatrix}$$

New CDR Estimator

N-channel CDR estimator

$$\Lambda_{\text{gen}}(l,f) = \frac{\gamma_n(l,f) - \gamma_x(l,f)}{\gamma_x(l,f) - 1}$$

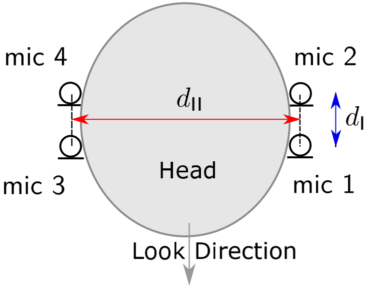
$\gamma_x(l,f)$: GMC of coherence matrix $\boldsymbol{C}_x(l,f)$ estimated from input signals

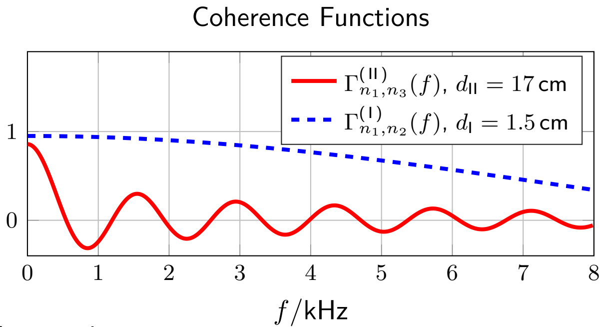
$\gamma_n(l,f)$: GMC of coherence matrix $\boldsymbol{C}_n(l,f)$ obtained by noise coherence models

Properties

- Real-valued and positive
- No TDOA information of desired source $d(k)$ needed
- $\Lambda_{\text{gen}}(l,f) = \frac{|\Gamma_n(l,f)| - |\Gamma_x(l,f)|}{|\Gamma_x(l,f)| - 1}$ for special case $N = 2$

GMC for Binaural Hearing Aids (HAs)





- Diffuse noise coherence model for one device:

$$\Gamma^{(\mathsf{I})}_{n_1,n_2}(f) = a_{\mathsf{s}}\,\mathrm{sinc}\left(2\pi f \frac{d_{\mathsf{I}}}{c}\right) \text{ with } 0 \ll a_{\mathsf{s}} \leq 1$$

- Diffuse noise coherence model between both devices:

$\Gamma^{(\mathsf{II})}_{n_1,n_3}(f)$ can be determined, e.g., by semi-analytical expressions [Jeub 2012]

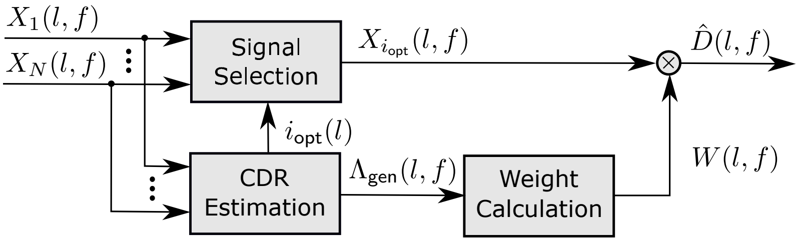
GMC for Binaural Hearing Aids (HAs)

- GMC for diffuse noise $\gamma_n(f)$ determined by noise coherence matrix

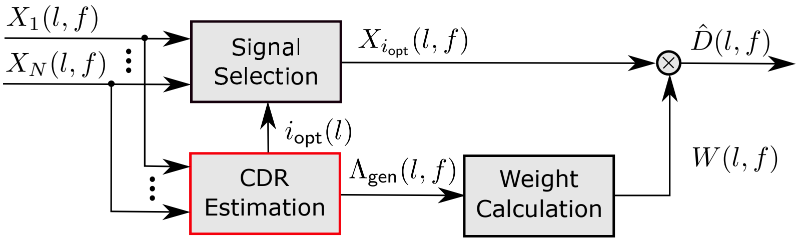
$$\boldsymbol{C}_n(f) = \begin{bmatrix} 1 & \Gamma^{(\mathsf{I})}_{n_1,n_2}(f) & \Gamma^{(\mathsf{II})}_{n_1,n_3}(f) & \Gamma^{(\mathsf{II})}_{n_1,n_4}(f) \\ \Gamma^{(\mathsf{I})}_{n_2,n_1}(f) & 1 & \Gamma^{(\mathsf{II})}_{n_2,n_3}(f) & \Gamma^{(\mathsf{II})}_{n_2,n_4}(f) \\ \Gamma^{(\mathsf{II})}_{n_3,n_1}(f) & \Gamma^{(\mathsf{II})}_{n_3,n_2}(f) & 1 & \Gamma^{(\mathsf{I})}_{n_3,n_4}(f) \\ \Gamma^{(\mathsf{II})}_{n_4,n_1}(f) & \Gamma^{(\mathsf{II})}_{n_4,n_2}(f) & \Gamma^{(\mathsf{I})}_{n_4,n_3}(f) & 1 \end{bmatrix}$$

- Assumed to be time-invariant
  ➜ $\gamma_n(f)$ can be calculated in advance

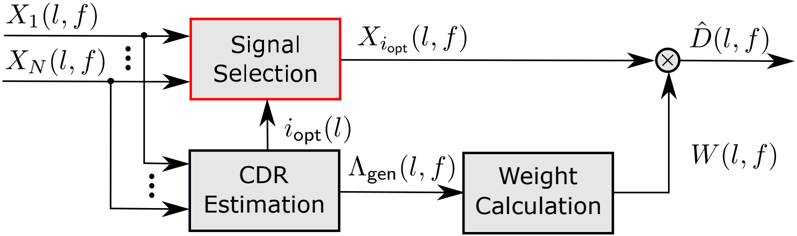
New N-channel CDR-based Speech Enhancement



New N-channel CDR-based Speech Enhancement



New N-channel CDR-based Speech Enhancement



- Largest coefficient $v_x^{(\text{max})}(l, f_m, i)$ of principal eigenvector of $\boldsymbol{C}_x(l,f)$ provides an indicator for signal contribution of microphone i, cf. [Ramirez 2008]
- Exploited for signal selection

$$i_\text{opt}(l) = \text{round}\left\{\alpha\, i_\text{opt}(l-1) + (1-\alpha)\, \bar{i}_\text{opt}(l)\right\}, \quad 0 < \alpha < 1$$

$$\bar{i}_\text{opt}(l) = \frac{1}{M}\sum_{m=0}^{M-1} \arg\max_i \left\{\left|v_x^{(\text{max})}(l, f_m, i)\right|\right\}, \quad M\text{: no. of frequency bands}$$

Evaluation

Setup

- Speech enhancement for binaural HAs with 4 microphones (2 per device)
- Speech signal convolved with head-related impulse responses (HRIRs) [Kayser 2009]:
  - cafeteria ($T_{60} = 1.25\,$s, 12 HRIR sets)
  - courtyard ($T_{60} = 0.9\,$s, 12 HRIR sets)
  - first office room ($T_{60} = 0.4\,$s, 37 HRIR sets)
  - second office room ($T_{60} = 0.3\,$s, 8 HRIR sets)
- Babble noise added to achieve **CDR of 5 dB** for microphone $\#$ 1 ($f_\mathsf{s} = 16\,$kHz)

Evaluation

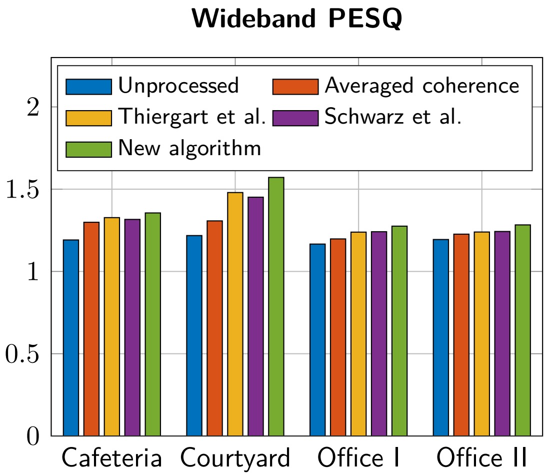
CDR-based speech enhancement using

- biased DOA-independent CDR estimator of [Thiergart 2012]
- unbiased DOA-independent CDR estimator of [Schwarz 2012]
- unbiased DOA-independent CDR estimator of [Schwarz 2012] with averaged coherence estimates (see paper for details)
- new CDR estimator with signal selection

Same weighting rule for all algorithms

- Spectral magnitude subtraction ($\beta = 1/2, \;\; \alpha = 1$)
- Overlapp-add method (512 frequency bands)

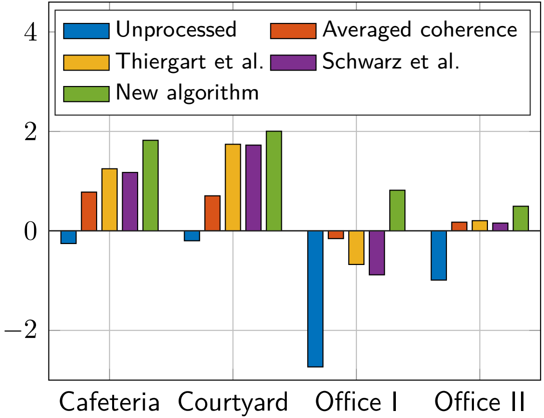
Evaluation



- Clean speech taken as reference signal
- Average results for each room
- Similar performance for all algorithms
- New algorithm shows best performance for all scenarios

Evaluation

Frequency-weighted segmental SNR in dB

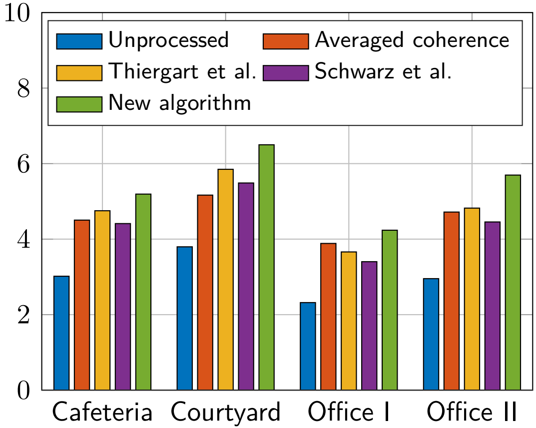


- Clean speech taken as reference signal
- New algorithm shows best performance

→ Outperforms averaging of coherence estimates

Evaluation

Speech-to-reverberation modulation energy ratio



- Non-intrusive measure
- Quality assessment for dereverberated speech [Falk 2010]
- New algorithm shows best performance

Conclusions

Generalized CDR-based signal enhancement algorithm presented

- Adopts concept of generalized magnitude coherence (GMC)
- Exploits inherently an arbitrary number of microphone signals
- Most appropriate microphone for signal enhancement determined implicitly without requiring TDOA information
- Achieves a consistently better signal quality than related approaches for binaural speech enhancement
- Also of interest for speech enhancement in mobile phones or sensor networks
  ➜ Subject for further investigations

Conclusions

Generalized CDR-based signal enhancement algorithm presented

- Adopts concept of generalized magnitude coherence (GMC)
- Exploits inherently an arbitrary number of microphone signals
- Most appropriate microphone for signal enhancement determined implicitly without requiring TDOA information
- Achieves a consistently better signal quality than related approaches for binaural speech enhancement
- Also of interest for speech enhancement in mobile phones or sensor networks
  ➜ Subject for further investigations

Conclusions

Generalized CDR-based signal enhancement algorithm presented

- Adopts concept of generalized magnitude coherence (GMC)
- Exploits inherently an arbitrary number of microphone signals
- Most appropriate microphone for signal enhancement determined implicitly without requiring TDOA information
- Achieves a consistently better signal quality than related approaches for binaural speech enhancement
- Also of interest for speech enhancement in mobile phones or sensor networks
  ➜ Subject for further investigations

Conclusions

Generalized CDR-based signal enhancement algorithm presented

- Adopts concept of generalized magnitude coherence (GMC)
- Exploits inherently an arbitrary number of microphone signals
- Most appropriate microphone for signal enhancement determined implicitly without requiring TDOA information
- Achieves a consistently better signal quality than related approaches for binaural speech enhancement
- Also of interest for speech enhancement in mobile phones or sensor networks
  ➜ Subject for further investigations

Conclusions

Generalized CDR-based signal enhancement algorithm presented

- Adopts concept of generalized magnitude coherence (GMC)
- Exploits inherently an arbitrary number of microphone signals
- Most appropriate microphone for signal enhancement determined implicitly without requiring TDOA information
- Achieves a consistently better signal quality than related approaches for binaural speech enhancement
- Also of interest for speech enhancement in mobile phones or sensor networks
  ➜ Subject for further investigations

Thank you for your attention!

References

[Habets 2007] E. A. P. Habets,
Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement,
Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands, 2007.

[Schwarz 2015] A. Schwarz and W. Kellermann,
"Coherent-to-Diffuse Power Ratio Estimation,"
IEEE/ACM Trans. on Audio, Speech, and Language Processing, vol. 23, no. 6, pp. 1006–1018, 6 2015.

[Ramirez 2008] D. Ramirez, J. Via, and I. Santamaria,
"A Generalization of the Magnitude Squared Coherence Spectrum for More than Two Signals: Definition, Properties and Estimation,"
in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas (Nevada), USA., Mar. 2008, pp. 3769–3772.

[Jeub 2012] M. Jeub,
Joint Dereverberation and Noise Reduction for Binaural Hearing Aids and Mobile Phones,
Ph.D. thesis, RWTH Aachen University, Aachen, Germany, 2012.

[Thiergart 2012] O. Thiergart, G. Del Galdo, and E. A. P. Habets,
"Signal-to-Reverberant Ratio Estimation Based on the Complex Spatial Coherence Between Omnidirectional Microphones,"
in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 309–312.

[Kayser 2009] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier,
"Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses,"
EURASIP Journal on Advances in Signal Processing, vol. 2009, pp. 1–10, 2009.

[Falk 2010] T. H. Falk, C. Zheng, and W. Chan,
"A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech,"
IEEE Trans. on Audio, Speech, and Language Processing, vol. 18, no. 7, pp. 1766–1774, Sept. 2010.