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)
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
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
Noisy reverberant speech at microphone $i$

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

\[
= \sum_{\kappa=0}^{K_{\text{early}}-1} s(\kappa) h_i(k - \kappa) + \sum_{\kappa=K_{\text{early}}}^{L-1} s(\kappa) h_i(k - \kappa) + n_i^{(\text{diff})}(k)
\]

desired speech $d_i(k)$

reverberant speech + 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

$K_{\text{early}}$: onset of late reverberation
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

\[
X_1(l, f) \quad X_2(l, f) \quad \text{Averaging} \quad \bar{X}(l, f) = \sqrt{\frac{|X_1(l, f)|^2 + |X_2(l, f)|^2}{2}} e^{j\angle\{X_1(l, f)\}} \quad \hat{D}(l, f)
\]

\[
\hat{\Lambda}(l, f) \quad \text{Weight Calculation} \quad W(l, f)
\]

- Speech enhancement by common noise suppression schemes, e.g.,
  \[
  W(l, f) = \max \left\{ W_{\min}, \left(1 - \left(\frac{\mu}{\hat{\Lambda}(l, f) + 1}\right)^\beta\right)^\alpha \right\}
  \]
  \(W_{\min}\): minimum weight, \(\mu\) oversubtraction factor, \(\alpha, \beta\): determines weighting rule

- Implicit channel selection by choice of phase term \(\angle\{X_1(l, f)\}\) for averaging operation
Speech enhancement by common noise suppression schemes, e.g.,

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

- \(W_{\text{min}}\): minimum weight, \(\mu\) oversubtraction factor, \(\alpha, \beta\): determines weighting rule

Implicit channel selection by choice of phase term \(\angle \{X_1(l, f)\}\) for averaging operation.
CDR-based Speech Enhancement

\[ \bar{X}(l, f) = \sqrt{\frac{|X_1(l, f)|^2 + |X_2(l, f)|^2}{2}} e^{j\angle\{X_1(l, f)\}} \]

- Speech enhancement by common noise suppression schemes, e.g.,
  \[ W(l, f) = \max \left\{ W_{\text{min}}, \left( 1 - \frac{\mu}{\tilde{\Lambda}(l, f) + 1} \right)^\alpha \right\} \]
  \( W_{\text{min}} \): minimum weight, \( \mu \): oversubtraction factor, \( \alpha, \beta \): determines weighting rule

- Implicit channel selection by choice of phase term \( \angle\{X_1(l, f)\} \) 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]

\[ \Lambda_{1,2}^{(Th)}(l, f) = \text{Re} \left\{ \frac{\hat{\Gamma}_{n_1,n_2}(l, f) - \hat{\Gamma}_{x_1,x_2}(l, f)}{\hat{\Gamma}_{x_1,x_2}(l, f) - e^{j \angle \hat{\Gamma}_{x_1,x_2}(l,f)}} \right\} \]

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)**

\[
\gamma_x(l, f) = \frac{1}{N - 1} \left( \lambda_{x}^{(\text{max})}(l, f) - 1 \right)
\]

\(\lambda_{x}^{(\text{max})}(l, f)\): largest eigenvalue of the \(N \times N\) coherence matrix

\[
\mathbf{C}_x(l, f) = \begin{bmatrix}
1 & \Gamma_{x_1, x_2}(l, f) & \cdots & \Gamma_{x_1, x_N}(l, f) \\
\Gamma_{x_2, x_1}(l, f) & 1 & \cdots & \Gamma_{x_2, x_N}(l, f) \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{x_N, x_1}(l, f) & \Gamma_{x_N, x_2}(l, f) & \cdots & 1
\end{bmatrix}
\]
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 \( C_x(l, f) \) estimated from input signals

\( \gamma_n(l, f) \) : GMC of coherence matrix \( C_n(l, f) \) 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 \( N = 2 \)
Diffuse noise coherence model for one device:
\[ \Gamma^{(1)}_{n_1,n_2}(f) = a_s \text{sinc} \left( 2\pi f \frac{d_1}{c} \right) \text{ with } 0 \ll a_s \leq 1 \]

Diffuse noise coherence model between both devices:
\[ \Gamma^{(II)}_{n_1,n_3}(f) \text{ can be determined, e.g., by semi-analytical expressions [Jeub 2012]} \]
GMC for Binaural Hearing Aids (HAs)

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

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

- Assumed to be time-invariant
  - $\gamma_n(f)$ can be calculated in advance
New $N$-channel CDR-based Speech Enhancement

\[ X_1(l, f), X_N(l, f) \]

Signal Selection

\[ X_{i_{\text{opt}}}(l, f) \]

CDR Estimation

\[ i_{\text{opt}}(l), \Lambda_{\text{gen}}(l, f) \]

Weight Calculation

\[ \hat{D}(l, f), W(l, f) \]
New $N$-channel CDR-based Speech Enhancement

$$X_1(l, f) \rightarrow \text{Signal Selection} \rightarrow X_{i_{\text{opt}}}(l, f) \rightarrow \hat{D}(l, f)$$

$$X_N(l, f) \vdots \vdots$$

$$X_1(l, f) \rightarrow \text{CDR Estimation} \rightarrow i_{\text{opt}}(l) \rightarrow \Lambda_{\text{gen}}(l, f) \rightarrow W(l, f)$$
Largest coefficient $v_{x}^{(\text{max})}(l, f_m, i)$ of principal eigenvector of $C_x(l, f)$ provides an indicator for signal contribution of microphone $i$, cf. [Ramirez 2008]

Exploited for signal selection

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

$$\bar{i}_{\text{opt}}(l) = \frac{1}{M} \sum_{m=0}^{M-1} \text{arg} \max_{i} \left\{ |v_{x}^{(\text{max})}(l, f_m, i)| \right\}, \quad M: \text{no. of frequency bands}$$
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 \( T_{60} = 1.25 \text{s}, 12 \text{ HRIR sets} \)
  - courtyard \( T_{60} = 0.9 \text{s}, 12 \text{ HRIR sets} \)
  - first office room \( T_{60} = 0.4 \text{s}, 37 \text{ HRIR sets} \)
  - second office room \( T_{60} = 0.3 \text{s}, 8 \text{ HRIR sets} \)
- Babble noise added to achieve **CDR of 5 dB** for microphone \# 1 \( f_s = 16 \text{ kHz} \)
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

Wideband PESQ

- 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

Löllmann et al.: Generalized Coherence-based Signal Enhancement
ICASSP 2020
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,

[Schwarz 2015] A. Schwarz and W. Kellermann,
“Coherent-to-Diffuse Power Ratio Estimation,”

[Ramirez 2008] D. Ramirez, J. Via, and I. Santamaría,
“A Generalization of the Magnitude Squared Coherence Spectrum for More than Two Signals: Definition, Properties and Estimation,”

[Jeub 2012] M. Jeub,
Joint Dereverberation and Noise Reduction for Binaural Hearing Aids and Mobile Phones,

[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,”

“Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses,”

“A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech,”