Joint Dereverberation and Noise Reduction Using Beamforming and a Single-Channel Speech Enhancement Scheme

B. Cauchi, I. Kodrasi, R. Rehr,
S. Gerlach, T. Gerkmann,
S. Doclo, S. Goetze

Fraunhofer IDMT,
Project Group Hearing, Speech and Audio Technology

Oldenburg University,
Signal Processing Group

Florence, May 10th 2014

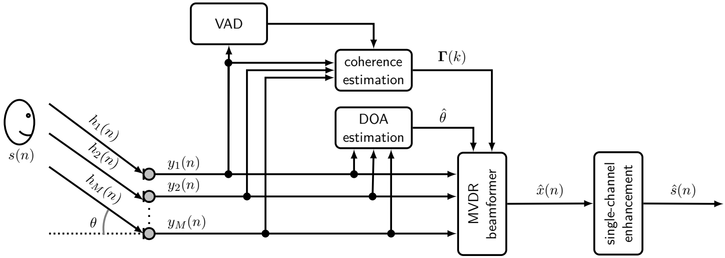
benjamin.cauchi@idmt.fraunhofer.de
phone 0441 2172-450

Introduction

- Overview of the proposed system
- Design of the MVDR beamformer
  - DOA estimated using MUSIC
  - Estimated noise covariance
- Single-channel enhancement scheme
  - Combination and optimization of published estimators
- Results
  - Objective measures
  - MUSHRA scores
  - WER using a baseline recognizer

1. Proposed System

Overview



- Beamformer: towards estimated direction of arrival (DOA)
- Single-channel enhancement: based on statistical estimators
  - Late reverberant spectral variance (LRSV)
  - Noise spectral variance (NSV)
  - Speech spectral variance (SSV)

2. MVDR Beamformer

With $Y_m(k,\ell)$ the STFT of the input signal in the m-th microphone we define

$$\mathbf{Y}(k,\ell) = [Y_1(k,\ell)\ Y_2(k,\ell)\ \ldots\ Y_M(k,\ell)]^T$$

The output $\hat{X}(k,\ell)$ of the beamformer is obtained as

$$\hat{X}(k,\ell) = \mathbf{W}_\theta^H(k)\mathbf{Y}(k,\ell)$$

where

$$\mathbf{W}_\theta(k) = \frac{\mathbf{\Gamma}^{-1}(k)\mathbf{d}_\theta(k)}{\mathbf{d}_\theta^H(k)\mathbf{\Gamma}^{-1}(k)\mathbf{d}_\theta(k)}$$

- Noise coherence matrix: $\mathbf{\Gamma}(k) \rightarrow$ estimated using a VAD.
- Steering vector: $\mathbf{d}_\theta(k) \rightarrow$ from $\hat{\theta}$ using a far-field assumption.

Estimation of noisefield coherence

- Noise periods identified with a VAD
  - Comparison between the long-term spectral envelope and the average noise spectrum
- $\mathbf{\Gamma}(k)$ is estimated using detected noise-only frames
- Alternatively, a theoretically diffuse noise field is used:

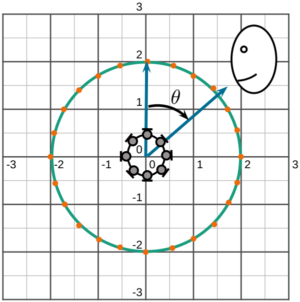
$$\mathbf{W}_\theta(k) = \frac{(\mathbf{\Gamma}_{\mathrm{diff}}(k)+\varrho(k)\mathbf{I}_M)^{-1}\mathbf{d}_\theta(k)}{\mathbf{d}_\theta^H(k)(\mathbf{\Gamma}_{\mathrm{diff}}(k)+\varrho(k)\mathbf{I}_M)^{-1}\mathbf{d}_\theta(k)}$$

with $\varrho(k)$ a constraint such that

$$\mathbf{W}_\theta^H(k)\mathbf{W}_\theta(k) \leq \mathrm{WNG}_{\mathrm{max}} = 10 \text{ dB}$$

Ramirez, J., Segura, J.C., Benitez, C., de la Torre, A., and Rubio, A., *Efficient voice activity detection algorithms using long-term speech information*, 2003.

DOA Estimation



$$\hat{\theta} = \underset{\theta}{\text{argmax}} \frac{1}{K} \sum_{k_{\text{low}}}^{k_{\text{high}}} U_\theta(k, \ell),$$

where $U_\theta(k, \ell)$ is the MUSIC pseudo-spectra:

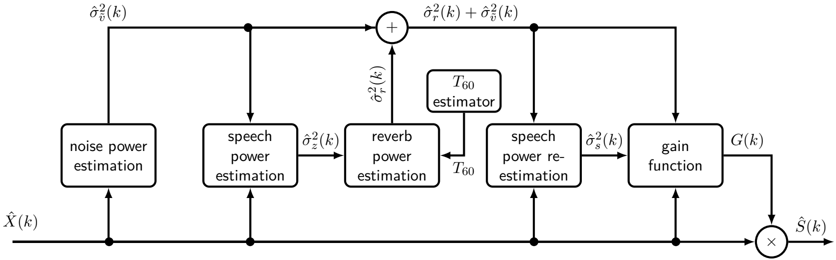
$$U_\theta(k, \ell) = \frac{1}{\mathbf{d}_\theta^H(k)\boldsymbol{E}(k,\ell)\boldsymbol{E}^{\mathrm{H}}(k,\ell)\mathbf{d}_\theta(k)}$$

$$\boldsymbol{E}(k, \ell) = [\boldsymbol{e}_{Q+1}(k, \ell) \ldots \boldsymbol{e}_M(k, \ell)]$$

with $\boldsymbol{e}_m$ denoting eigenvectors of the covariance matrix of $\mathbf{Y}(k, \ell)$.

Schmidt, R., *Multiple emitter location and signal parameter estimation*, 1986.

Overview

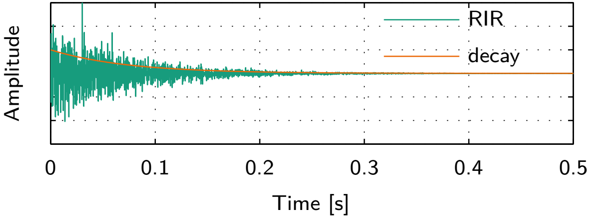


- $\sigma^2_{\tilde{v}}(k,\ell)$ estimated using Minimum Statistics
- $\sigma^2_s(k,\ell)$ estimated using Cepstral Smoothing
- $\sigma^2_r(k,\ell)$ estimated using Lebart's approach

Martin, R., *Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics*, 2001.
Breithaupt, C., Gerkmann T. and Martin, R., *A Novel A Priori SNR Estimation Approach Based on Selective Cepstro-Temporal Smoothing*, 2008.
Eaton, J., Gaubitch, N.D., Naylor, P.A., *Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost*, 2012.
Lebart, K., Boucher J.M. and Denbigh, P., *A new method based on spectral subtraction for speech dereverberation*, 2013.

LRSV estimation

- RIR modeled as Gaussian noise with decay $\Delta = \frac{3 \ln 10}{T_{60} f_s}$



- Representing the variance of the reverberant speech as:
  - $\sigma_z^2(k,\ell) = \sigma_r^2(k,\ell) + \sigma_s^2(k,\ell)$
- Leads to the estimator
  - $\hat{\sigma}_r^2\,(k,\ell) = \mathrm{e}^{-2\Delta T_d f_s} \sigma_z^2(k, \ell - T_d/T_s)$

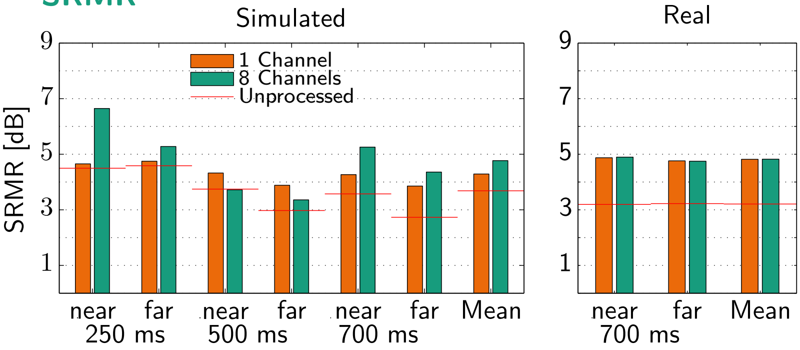
Lebart, K., Boucher J.M. and Denbigh, P., *A new method based on spectral subtraction for speech dereverberation*, 2001.

Gain function

- The output $\hat{X}(k,\ell)$ of the beamformer contains the anechoic speech, remaining noise and spatially filtered reverberation
  - $\hat{X}(k,\ell) = S(k,\ell) + \tilde{V}(k,\ell) + R(k,\ell)$
- We aim to compute a real gain such that:
  - $\hat{S}(k,\ell) = G(k,\ell)\hat{X}(k,\ell)$
- Computation of $G(k,\ell)$ using an MMSE estimation of the speech amplitude based on a super Gaussian speech model.

Breithaupt, C., Krawczyk, M., and Martin, R., *Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech*, 2008.

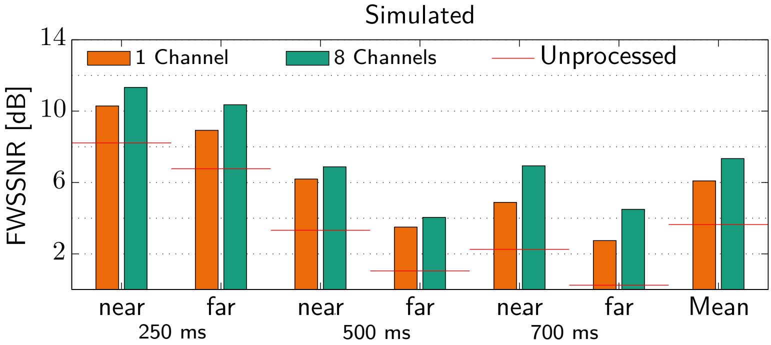
SRMR



- Illustrates dereverberation performance in all condition
- Better dereverberation achieved by multichannel, except for T_{60}=500 ms

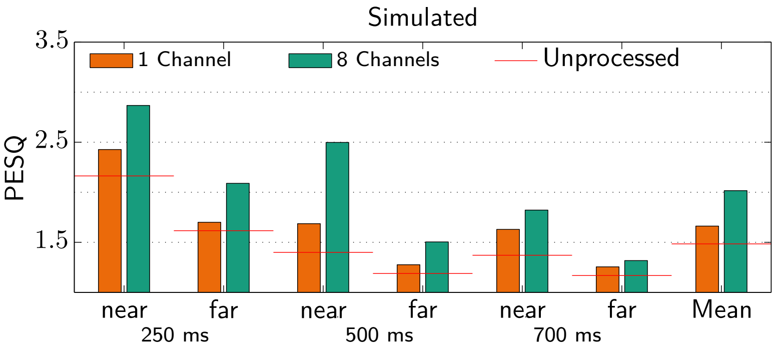
4. Objective Measures

FWSSNR



- Illustrates noise reduction in all condition
- Beamforming step advantageous for the noise reduction

PESQ



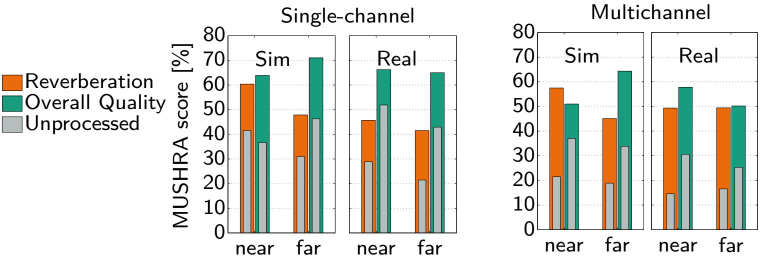
- Improvement of PESQ score in all condition illustrate the overall improvement in speech quality

5. Subjective Tests

MUSHRA test

Intermediate results of the subjective test ran by the organizers:

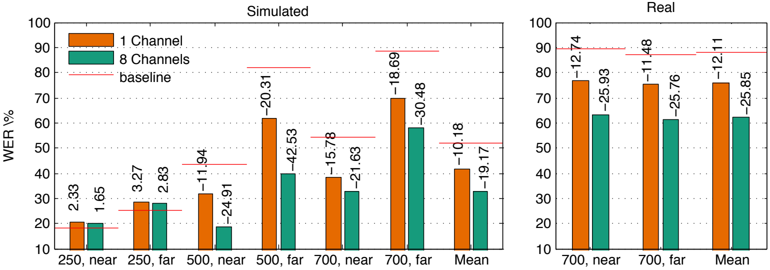
- Tests carried out separately for 1 and 8 channels



- Improvement for all tested condition
- Higher improvement of the overall quality

Word Error Rate

- Baseline recognizer provided by the organizers
- Using pre-trained models on clean data



7. Conclusion

- System based on combination of MVDR beamformer and spectral enhancement
- All parameters are blindly estimated
- Speech enhancement achieved in all conditions in terms of:
  - Objective measures
  - Subjective tests
  - Word error rate

Thank you very much for your attention

House of Hearing, Oldenburg

Questions ?

Fraunhofer IDMT
Project Group Hearing, Speech and Audio Technology

Oldenburg University
Signal Processing Group

benjamin.cauchi@idmt.fraunhofer.de