MULTICHANNEL JOINT DEREVERBRATION AND NOISE REDUCTION IN THE STFT DOMAIN

Eliav Benisty
Department of Electrical Engineering,
Technion Israel Institute of Technology

Supervised by: Prof. Israel Cohen
Prof. Jacob Benesty

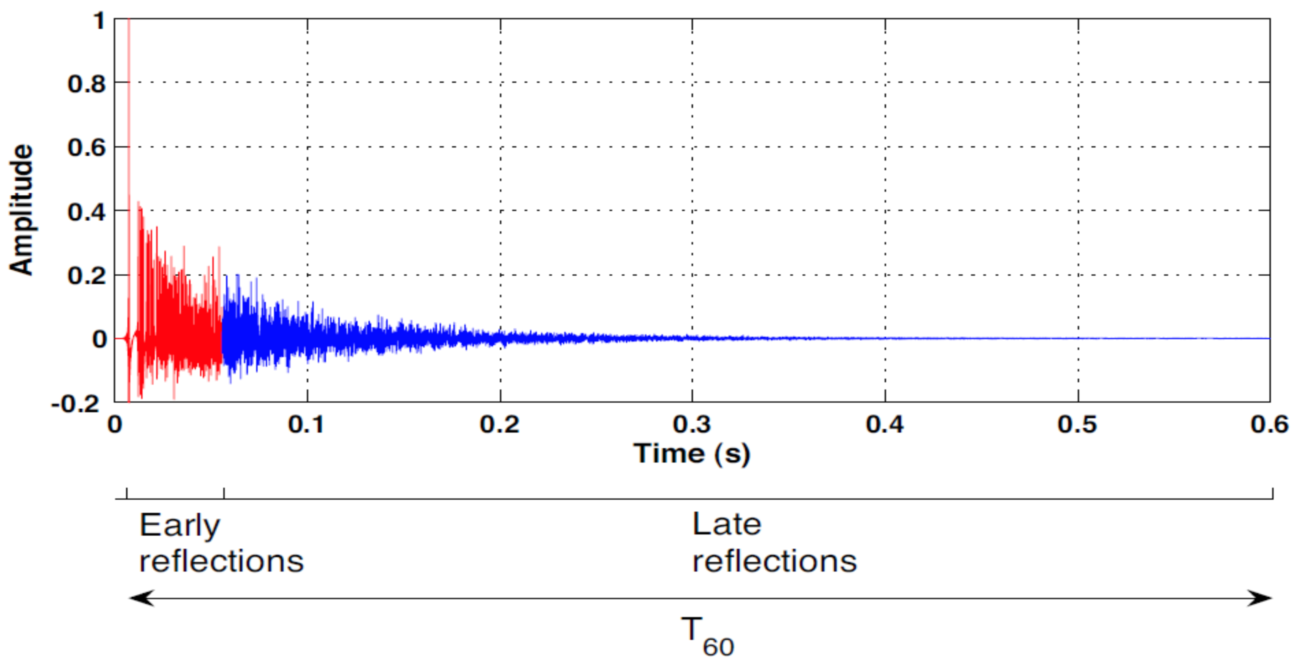
November 2014

Outline

1. Problem definition, previous work and main contributions
2. Optimal Beamformers for the JDNR problem
3. Reverberation Blockage beamformers
4. Summary and conclusions

Problem Definition

- In a typical environment the received speech signals are degraded by room reverberation and background noise.
- Noise is mostly additive and uncorrelated with the speech signal.
- Reverberations are the correlated delayed reflections of the source signal.
- The room impulse response (RIR) describes the linear relation between the speaker and the microphone.
- The RIR can be divided into late and early reflections.

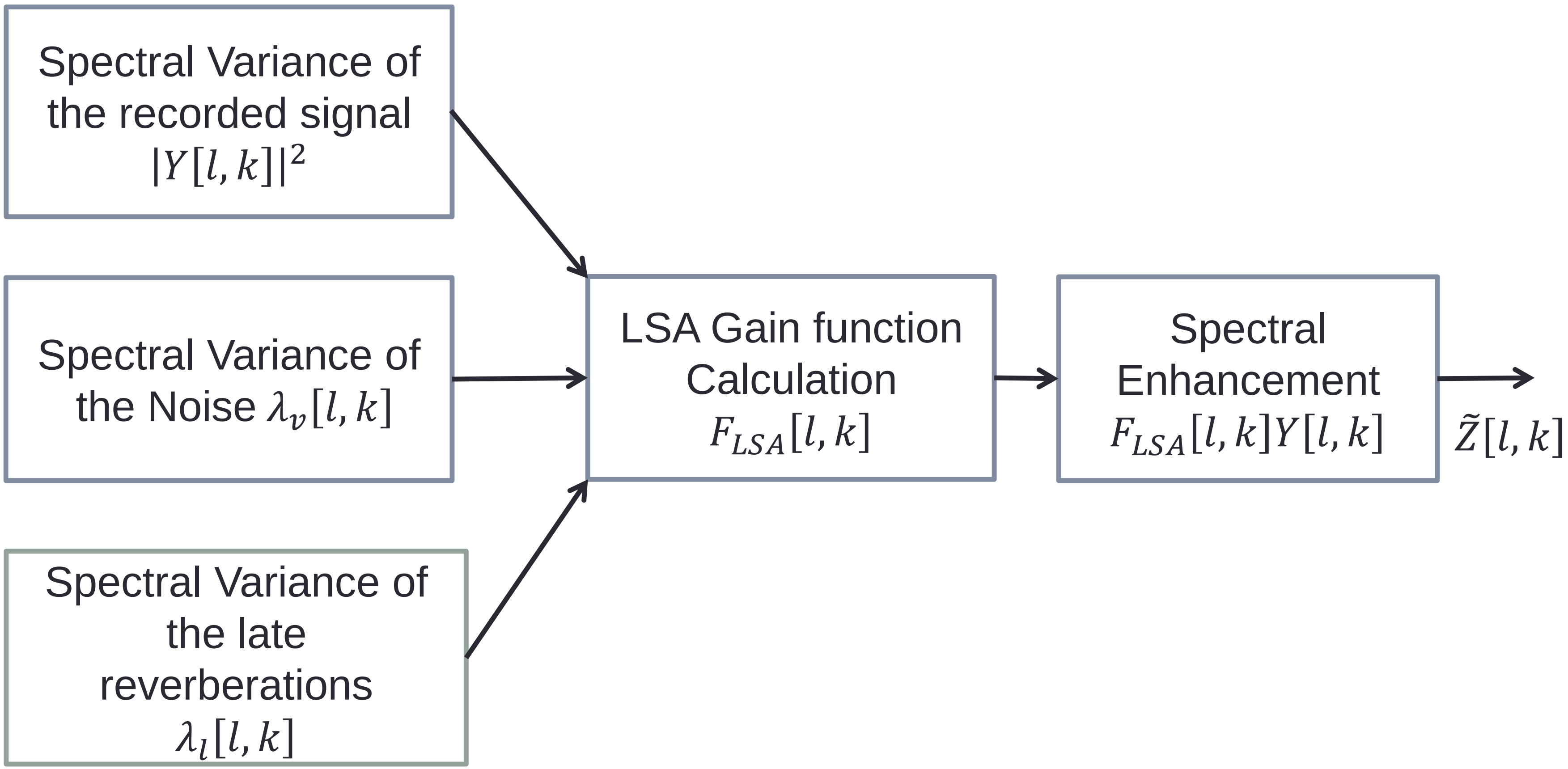


clean

Noise + Rev

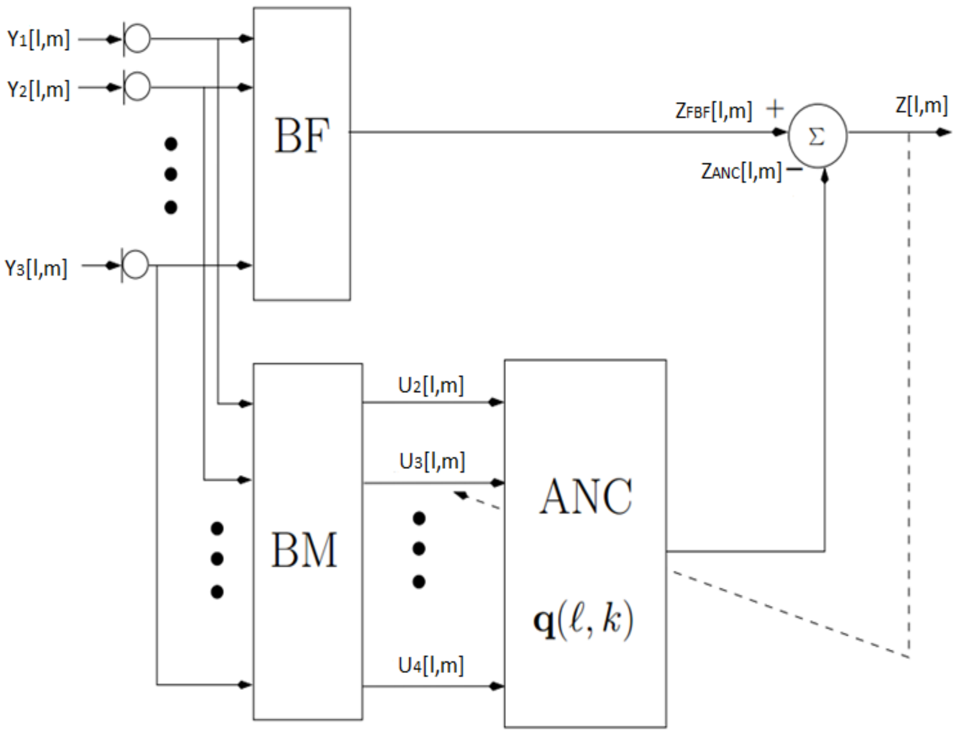
Dereverberation by Spectral Subtraction

E.A.P. Habets, S. Gannot and I. Cohen, “Late reverberant spectral variance estimation based on a statistical model,” IEEE Signal Processing Letters, vol. 16, pp. 770-773, Sept.2009



TF-GSC

S. Gannot, D. Burshtein and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech”, IEEE Trans Aug. (2001)



- The first block is a fixed beamformer, designed to satisfy the constraint.
- The second block is a blocking matrix, which blocks the desired signal and produces noise-only reference signals.
- The third block is an unconstrained adaptive algorithm that aims at canceling the residual noise at the fixed beamformer output

Drawbacks of previous dereverberation algorithms

- In most of the beamforming algorithms the focus was set on noise reduction.
- GSC based beamformers are sensitive to RTF/RIR estimation error which are common in reverberant environments → usually causes leakage.
- Most beamformers were designed to extract the clean but reverberant signal at the reference channel
- Reverberation at the look direction of the beamformer, are disregarded.
- The reverberation reduction problem is not addressed directly in the optimization scheme.

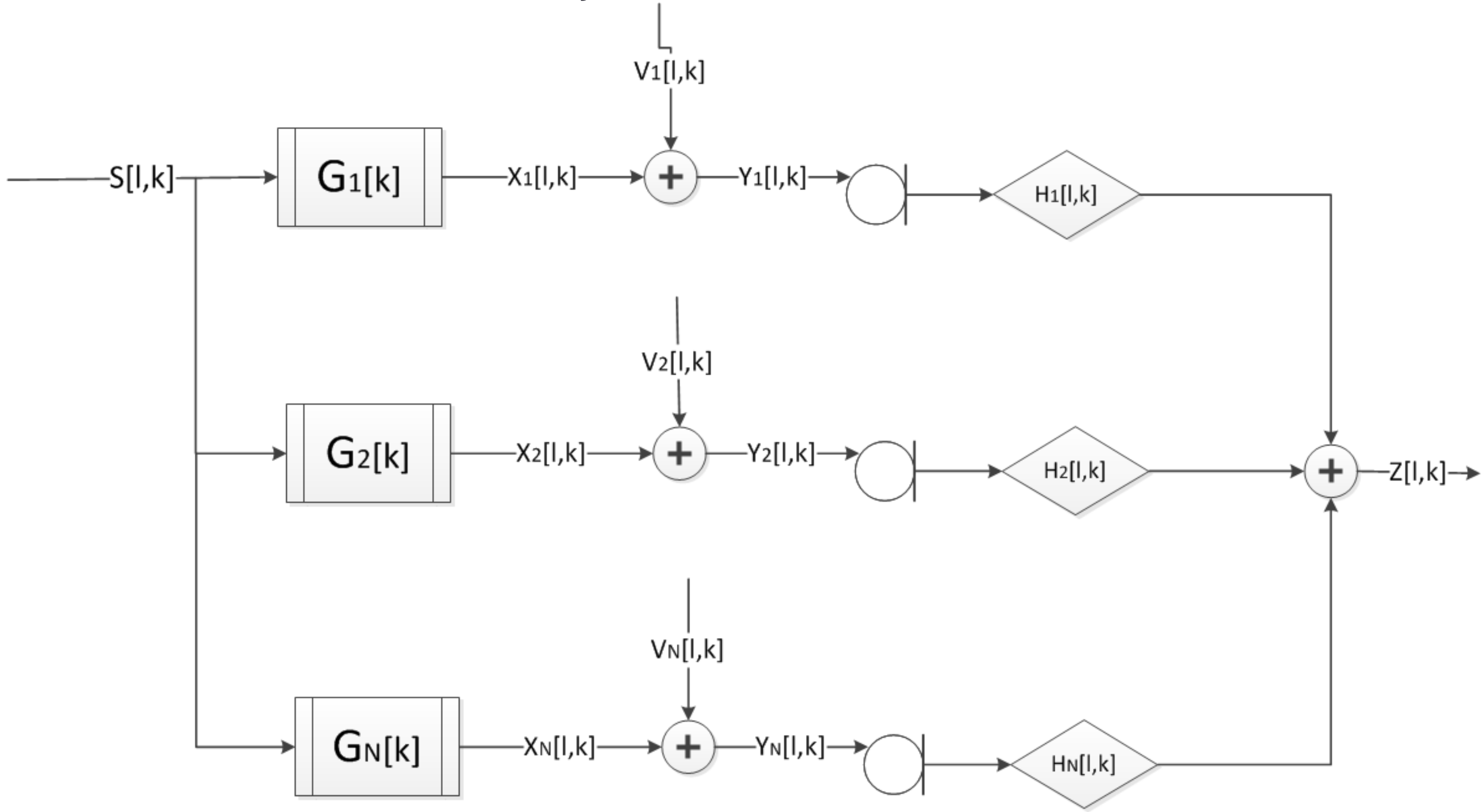
Main Contributions

- Propose an algorithm for the estimation of the full to early relative transfer function $\frac{\bar{G}\,[k]}{G_{1,e}[k]}$
- Use the full to early RTF estimator to adjust known beamformers to the problem of joint dereverberation and noise reduction.
- We neglect the GSC scheme and use close form solutions
- Propose an new signal decomposition model for the array observation vector.
- Use the model to design beamformers that address dereverberation directly in the optimization scheme

OPTIMAL BEAMFORMERS FOR THE JDNR PROBLEM

STFT Beamforming

- Applying a complex weight to the output of each sensor, and summing across the aperture
  - $Y_j[l,k] = G_j[k]S[l,k] + V_j[l,k], \quad j = 1 \dots N$
  - $Z[l,k] = \sum_{j=1}^{N} Y_j[l,k]H_j[l,k]$



Signal Decomposition

- The vector $\bar{Y}[l,k]=\bar{G}[k]S[l,k]+\bar{V}[l,k]$ can be written as a function of the desired signal.
- $\bar{Y}[l,k]=\overline{d}_{\bar{x},X_{1,c}}[k]X_{1,c}[l,k]+\bar{V}[l,k]$
- $\bar{d}_{\bar{x},X_{1,c}}[k]$ is the array's Steering Vector, constructed out of the full to early RTF vector
- $\bar{d}_{\bar{x},X_{1,c}}[k]=\left[\frac{G_1[k]}{G_{1,c}[k]},\frac{G_2[k]}{G_{1,c}[k]},\dots,\frac{G_N[k]}{G_{1,c}[k]}\right]$ $=\frac{G_1[k]}{G_{1,c}[k]}x\underbrace{\left[1,\frac{G_2[k]}{G_1[k]},\dots,\frac{G_N[k]}{G_1[k]}\right]}_{Estimators\ exists}$
- We suggest an algorithm to estimate $\frac{G_1[k]}{G_{1,c}[k]}$

MSE criteria

- The Subband MSE criterion is defined as:
- $J\left[\bar{h}[l,k]\right] = E\left\{\left|\bar{h}^H[l,k]\bar{Y}[l,k] - X_{1,c}[l,k]\right|^2\right\}$
- Can be divided into two uncorrelated parts
- $J_d\left[\bar{h}[l,k]\right] = \phi_{X_{1,c}}[l,k]\left|\bar{h}^H[l,k]\bar{d}_{\bar{x},X_{1,c}}[k] - 1\right|^2$
- $J_r\left[\bar{h}[l,k]\right] = \bar{h}^H[l,k]\Phi_{vv}[l,k]\bar{h}[l,k]$
- Goal → extract $X_{1,c}[l,k]$ from $\bar{Y}[l,k]$ using:
- $Z[l,k] = \bar{h}^H[l,k]\bar{Y}[l,k]$
- $\bar{h}[l,k]$ is chosen to either minimize $J\left[\bar{h}[l,k]\right]$ or minimize $J_d\left[\bar{h}[l,k]\right]$ or $J_r\left[\bar{h}[l,k]\right]$ subject to some constraint.

Performance Measures

- We Analyze the Beamformer performance by 3 criteria:

1. Noise Reduction.

$$OSNR = \frac{1}{N_T}\sum_{l=0}^{N_T}\frac{\sum_{k=0}^{M}\phi_{X1,c}[l,k]\left|\bar{h}^H[l,k]\bar{d}_{\bar{x},X_{1,c}}[k]\right|^2}{\sum_{k=0}^{M}\bar{h}^H[l,k]\Phi_{vv}[l,k]\bar{h}[l,k]}$$

2. Speech Distortion.

$$\nu_{sd} = \frac{1}{N_T}\sum_{l=0}^{N_T}\sum_{k=0}^{M}\phi_{X1,c}[l,k]\left|\bar{h}^H[l,k]\bar{d}_{\bar{x},X_{1,c}}[k]-1\right|^2$$

3. Reverberation suppression

$$LSD[l] = \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}\left|\mathcal{L}\{Z[l,k]\}-\mathcal{L}\{X_{1,c}[l,k]\}\right|^2 dB}$$

Wiener MVDR and Trade-off Beamformers

- The Wiener beamformer minimizes $J\left[\bar{h}[l,k]\right]$.
- The MVDR beamformer minimizes $J_r\left[\bar{h}[l,k]\right]$, under no distortion constraint $J_d\left[\bar{h}[l,k]\right] = 0$.
- The Trade-off beamfomer minimizes $J_d\left[\bar{h}[l,k]\right]$,with the constraint $J_r\left[\bar{h}[l,k]\right] = \beta\phi_{V_1}[l,k]$ 1>β>0

Trade-off Beamformer

- The Trade-off Beamformer solves the constraint minimization problem
- $\bar{h}_{\mathrm{T},\mu}[l,k] = \underbrace{min}_{\bar{h}} J_d\left[\bar{h}[l,k]\right],\ \mathrm{S.T}\ J_r\left[\bar{h}[l,k]\right] = \beta\phi_{V_1}[l,k]$

  β>0, for a Lagrange multiplier $\mu > 0$
- $\bar{h}_{\mathrm{T},\mu}[l,k] = \frac{\phi_{X_{1,c}}[l,k]\Phi_{\bar{v}\bar{v}}^{-1}[l,k]\bar{d}_{\bar{x},X_{1,c}}[k]}{\mu+\phi_{X_{1,c}}[l,k]\bar{d}_{\bar{x},X_{1,c}}^{H}\Phi_{\bar{v}\bar{v}}^{-1}[l,k]\bar{d}_{\bar{x},X_{1,c}}[k]}$
- When $\mu = 0 \rightarrow$ MVDR $\mu = 1 \rightarrow$ Wiener
- When $0 < \mu < 1 \rightarrow$ trade off between the MVDR and the Wiener beamformers.
- When $\mu > 1 \rightarrow$ Strong noise reduction, high speech distortion

Relative Transfer function Estimation

- Estimation of the classical RTFs have been suggested in:
- I. Cohen, Relative Transfer Function Identification Using Speech Signals Special Issue of the IEEE Trans – using a recursive normalized WLMS scheme
- A. Krueger, E. Warsitz, and R. Haeb-Umbach, Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation- using GEVD on the noise correlation matrix
- E. Habets and J. Benesty, A perspective on frequency-domain beamformers in room acoustics- using $\frac{\{\Phi_{xx}[l,k]\}_{1:N}}{\{\Phi_{xx}[l,k]\}_{1,1}}$

Relative Transfer function Estimation

- Similar to the work in - I. Cohen “Relative Transfer Function Identification Using Speech Signals Special”, Issue of the IEEE Trans.
- From the signal model
- $\Phi_{Y_1,Y_1}[l,k] = \phi_{X_{1,c},X_1}[l,k] d_{X_{1,c},X_1}[l,k] + \phi_{V_1,V_1}[l,k]$
- For every frequency bin k we define L equations
- $$\begin{bmatrix} \hat{\phi}_{Y_1,Y_1}[1,k] - \hat{\phi}_{V_1,V_1}[1,k] \\ \vdots \\ \hat{\phi}_{Y_1,Y_1}[L,k] - \hat{\phi}_{V_1,V_1}[L,k] \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{X_{1,c},X_1}[1,k] \\ \vdots \\ \hat{\phi}_{X_{1,c},X_1}[L,k] \end{bmatrix} \hat{d}_{X_{1,c},X_1}[k] + \begin{bmatrix} \bar{\epsilon}_1[1,k] \\ \vdots \\ \bar{\epsilon}_1[L,k] \end{bmatrix}$$

Relative Transfer function Estimation

- Use recursive normalized LMS solution :
- $\hat{d}_{X_1,X_{1,c}}[l,k] = \hat{d}_{X_1,X_{1,c}}[l-1,k]$
  $+ \mu_{RTF} \frac{\bar{\bar{I}}[k]}{\widehat{\phi}_{X_{1,c},X_1}[l,k]} \bar{\epsilon}_1[l,k]$
- $I[l,k] = \begin{cases} 1 \ if \ p[l,k] \geq p_0 \\ 0 \qquad otherwise \end{cases}$, p_0 is a predefine threshold
- μ_{RTF} is the LMS Update step
- $\bar{\epsilon}_1[l,k] = \widehat{\phi}_{Y_1,Y_1}[l,k] - \widehat{\phi}_{V_1,V_1}[l,k] - \widehat{\phi}_{X_{1,c},X_1}[l,k]\hat{d}_{X_1,X_{1,c}}[l-1,k]$
- This estimator is known as the minimum variance estimator.

PSD estimation

- The cross PSD Φ_{yy} is estimated by applying a first order recursive smoothing approach:
- $\Phi_{yy}[l,k] = \alpha_y \Phi_{yy}[l-1,k] + (1$

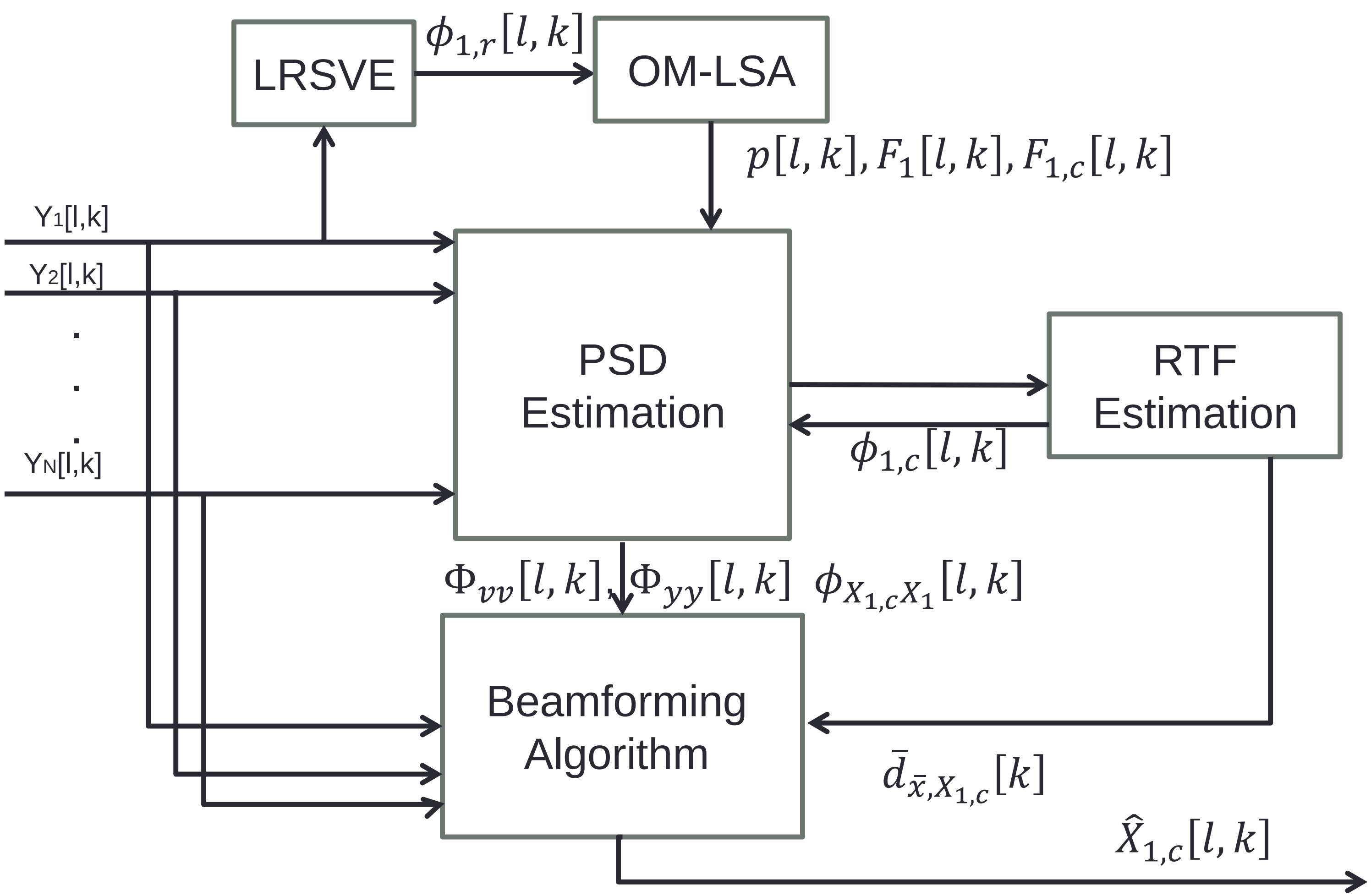
Desired and reverberant cross signal PSD Estimation

- $\Phi_{x1,x1_c}$ is calculated using two log-spectral amplitude gain functions
- $\Phi_{x1,x1_c}[l,k] = \alpha_y \Phi_{x1,x1_c}[l-1,k] + (1$

Desired signal PSD Estimation

- Using the estimate for $\Phi_{x1,x1_c}$ we use it to find an estimate for full to early RTF $d_{X_{1,c},X_1}[l,k]$.
- We incorporate $d_{X_{1,c},X_1}[l,k]$ into an estimator for the desired signal PSD
- $\Phi_{x1_c}[l,k] = \alpha_y \Phi_{x1_c}[l-1,k] + (1$

System flow

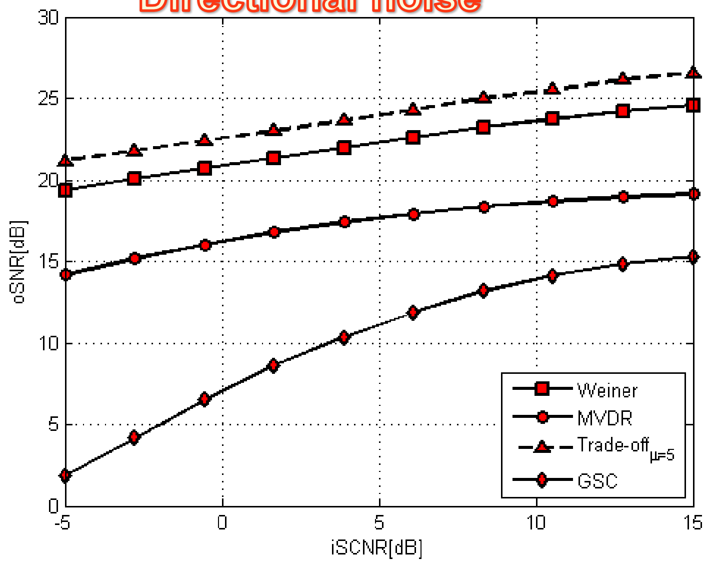

LRSVE
$\phi_{1,r}[l,k]$
OM-LSA
$p[l,k], F_1[l,k], F_{1,c}[l,k]$
Y1[l,k]
Y2[l,k]
YN[l,k]
PSD Estimation
RTF Estimation
$\phi_{1,c}[l,k]$
$\Phi_{vv}[l,k], \Phi_{yy}[l,k]$ $\phi_{X_{1,c}X_1}[l,k]$
Beamforming Algorithm
$\bar{d}_{\bar{x},X_{1,c}}[k]$
$\hat{X}_{1,c}[l,k]$

Simulation Description

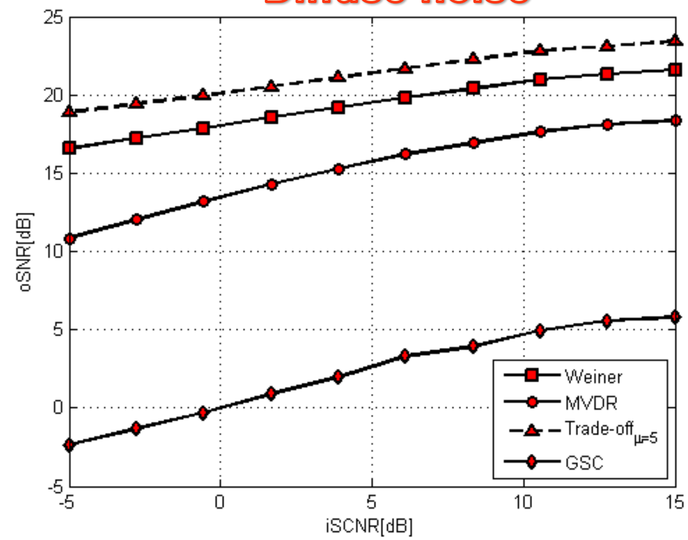
- The system is constructed out of 7 narrow band microphones array.
- The distance between each to microphones was set to 5[cm].
- The room dimensions were set to $[6\text{x}6\text{x}4][m]^{3}$
- The source was placed in $[X, Y, Z] = [3{,}3{,}2][m]$
- The reverberation time T_{60} is adaptable
- A noise source was placed at $[X, Y, Z] = [4{,}4{,}2][m]$ from the speaker, with an adaptable - iSCNR.
- To simulate the microphones thermal noise, white noise with 20[dB] SNR was added to each
- TF-GSC without post filter, BM was constructed out of the estimated RTF.

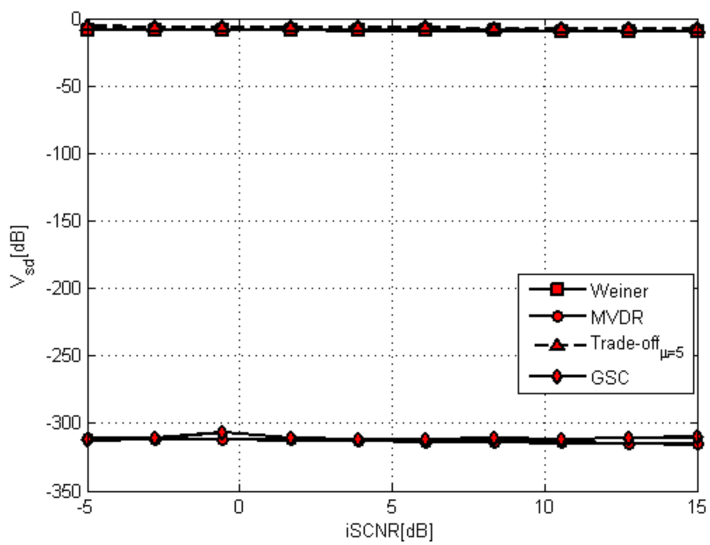
Influence of iSCNR T60=0.8[sec]

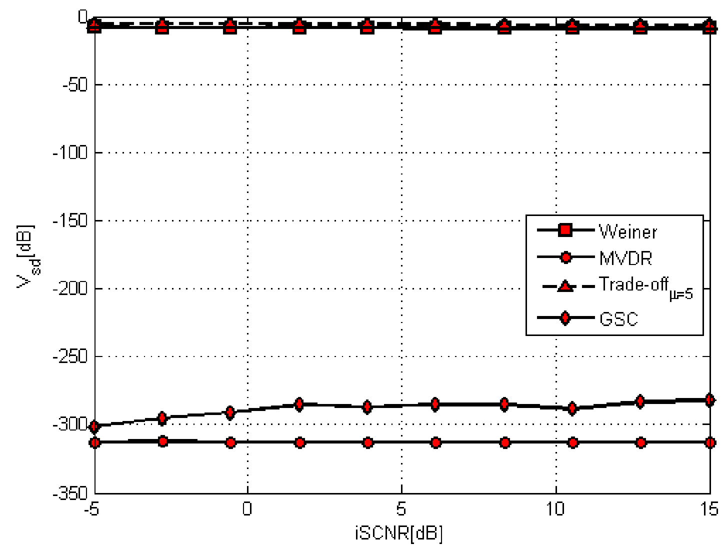
Directional noise



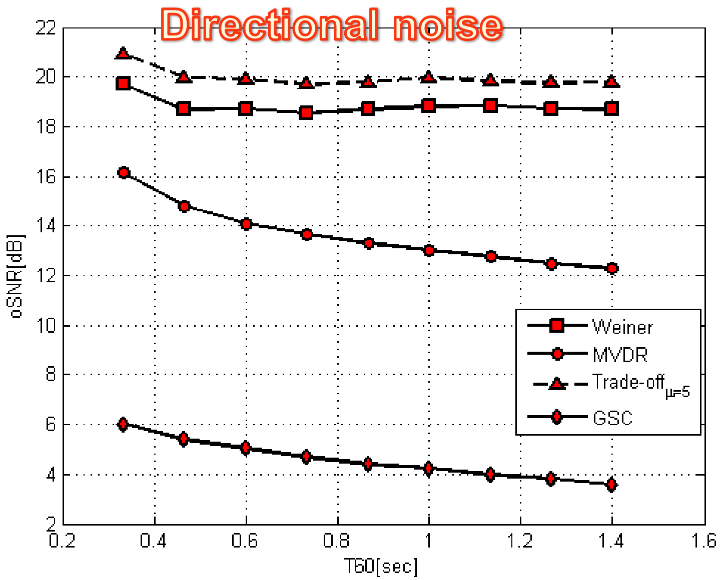
Diffuse noise

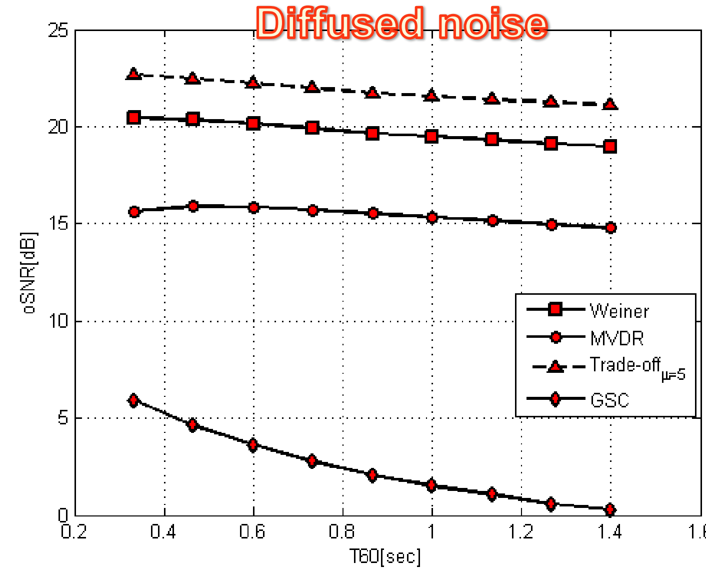


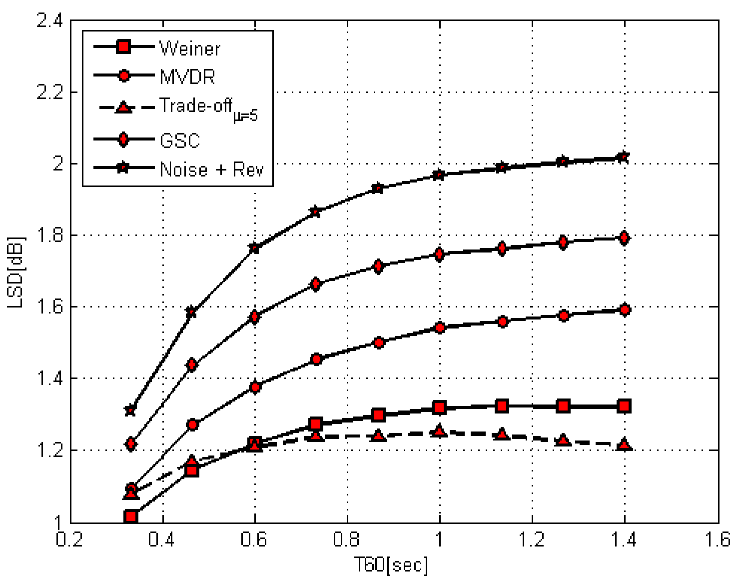


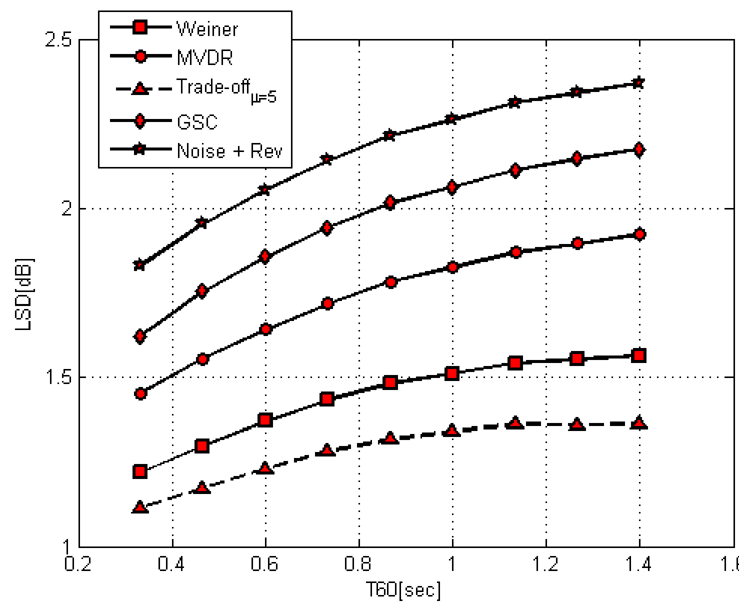


Influence of T60 iSCNR = 5[db]









REVERBERATION BLOCKAGE BEAMFORMERS

New Signal decomposition

- We preform a simple manipulation on the signal model
- $\bar{Y}[l,k] = \bar{r}[k]X_{1,c}[l,k] + i_N X_{1,c}[l,k] + \bar{V}[l,k]$
- $\bar{r}[k]$ is the a vector holding the ratio between the unwanted parts of the RIR to the desired part
- $\bar{r}[k] = \bar{d}_{X_1,X_{1,c}}[k] - i_N = \left[\frac{G_1[k]}{G_{1,c}[k]} - 1, \frac{G_2[k]}{G_{1,c}[k]} - 1, \dots, \frac{G_N[k]}{G_{1,c}[k]}\right.$

MSE criteria

- The Sub band MSE criterion is defined as:
- $J\left[\bar{h}[l,k]\right] = \phi_{X_{1,c}}[l,k]\left[\left|\bar{h}^H[l,k]i_N - 1\right|^2 + 2Re\left\{\left(\bar{h}^H[l,k]i_N\right.\right.\right.$

Performance Measures

- We Analyze the Beamformer performance using the following broadband measures

1. Noise Reduction + rev suppression

$$oSNR_{BB}[l] = \frac{\sum_{k=0}^{M} \phi_{X1,c}[l,k] \left| \bar{h}^H[l,k] i_N \right|^2}{\sum_{k=0}^{M} \bar{h}^H[l,k] \Phi_{tt}[l,k] \bar{h}[l,k]}$$

$$iSNR_{BB}[l] = \frac{\sum_{k=0}^{M} \phi_{X1,c}[l,k]}{\sum_{k=0}^{M} \phi_{X1,c}[l,k] \left[\left| \bar{r}_1 \quad [k] \right|^2 + 2RE\{ \bar{r}_1 \quad [k] \} \right] + \phi_{V1}[l,k]}$$

$$A_{BB} = \frac{1}{N_T} \sum_{l=0}^{N_T} \frac{oSNR_{BB}}{iSNR_{BB}}$$

Performance Measures

- We Analyze the Beamformer performance using the following broadband measures

1. Speech Distortion. ν_{sd}

$$= \frac{1}{N_T} \sum_{l=0}^{N_T} \frac{\sum_{k=0}^{M} \phi_{X1,c}[l,k] \left| \bar{h}^H[l,k] \bar{d}[k] \ -1 \right|^2}{\sum_{k=0}^{M} \phi_{X1,c}[l,k]}$$

2. Reverberation suppression index

$$\nu_{sr} = \frac{1}{N_T} \sum_{l=0}^{N_T} \frac{\sum_{k=0}^{M} \phi_{X1,c}[l,k] \left| \bar{h}^H[l,k] \bar{r}[k] \right|^2}{\sum_{k=0}^{M} \phi_{X1,c}[l,k]}$$

The Reverberation Block Wiener Beamformer

- The RBW Beamformer solves the constraint minimization problem
- $\bar{h}_{RBW}[l,k] = \underbrace{min}_{\bar{h}}\left\{J_d\left[\bar{h}[l,k]\right] + J_{rn}\left[\bar{h}[l,k]\right]\right\}$

$$\mathrm{S.T}\ \bar{h}^H[l,k]\bar{r} = 0$$

- $\bar{h}_{RBW}[l,k] = \phi_{X_{1,c}}[l,k]\Phi_{pp}^{-1}[l,k]\left[I_{NxN} - \frac{\bar{r}[k]\bar{r}^H[k]\Phi_{pp}^{-1}[l,k]}{\bar{r}^H[k]\Phi_{pp}^{-1}[l,k]\bar{r}[k]}\right] i_N$
- $\Phi_{pp}^{-1}[l,k] = \Phi_{vv}^{-1}[l,k] - \frac{\phi_{X_{1,c}}[l,k]\Phi_{vv}^{-1}[l,k]\ i_N i_N^T \Phi_{vv}^{-1}[l,k]}{1+\phi_{X_{1,c}}[l,k] i_N^T \Phi_{vv}^{-1}[l,k] i_N}$
- The RBW Beamformer takes the reverberation suppression index to 0

Normalized MVDR Beamformer

- The NMVDR Beamformer solves the constraint minimization problem
- $\bar{h}_{NMVDR}[l,k] = \underbrace{min}_{\bar{h}}\left\{J_{rn}\left[\bar{h}[l,k]\right] + J_{rc}\left[\bar{h}[l,k]\right]\right\}$

$$\mathrm{S.T}\ \bar{h}^{H}[l,k]i_N = 1$$

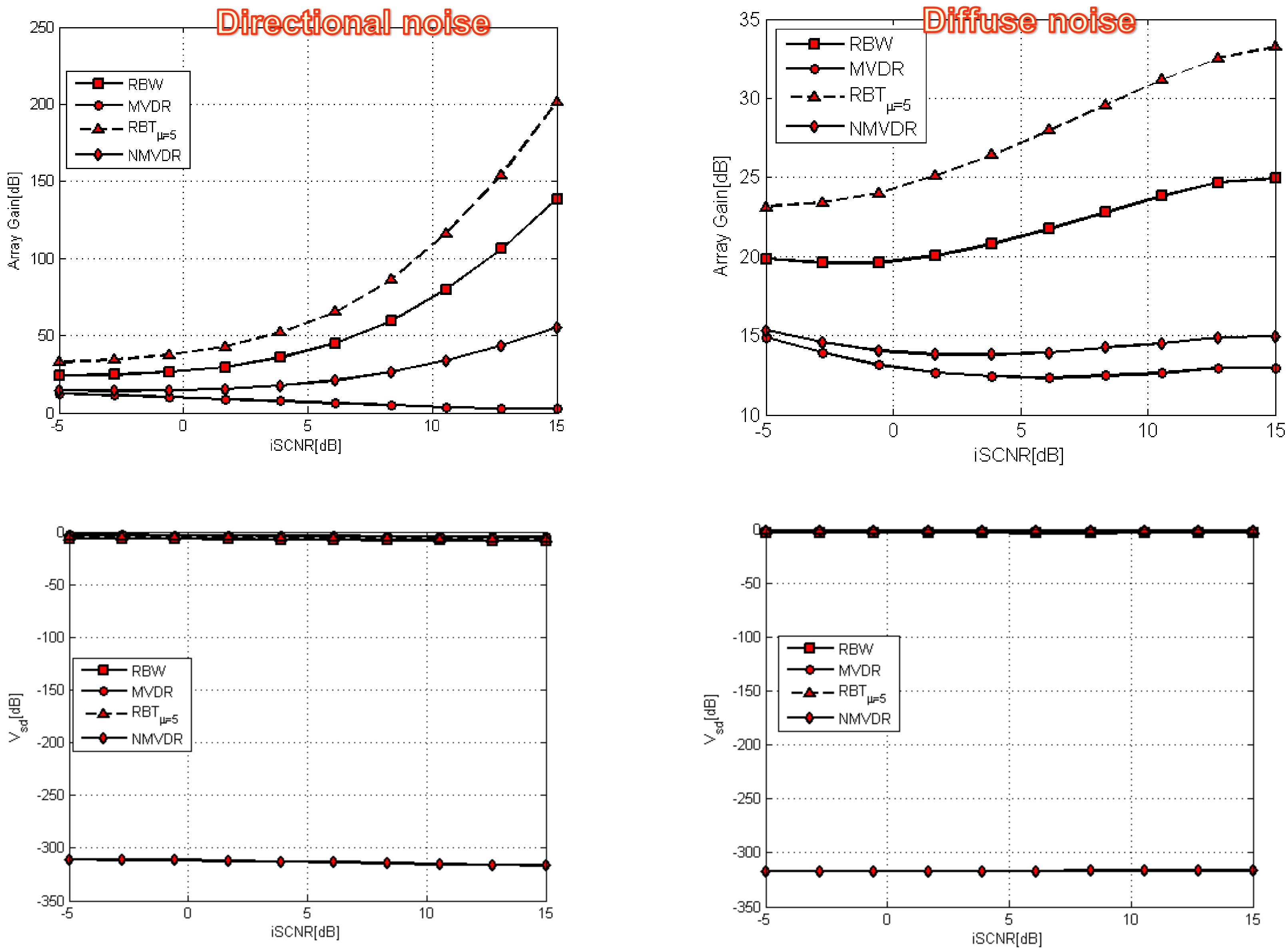
$$\bar{h}_{NMVDR}[k,l] = \frac{\Phi_{tt}^{-1}[l,k]i_N}{{i_N}^{T}\Phi_{tt}^{-1}[l,k]i_N}$$

- $\nu_{sd}\left[\bar{h}_{NMVDR}[l,k]\right] = 0$
- Special case of the NLCMV beamformer.
- The NLCMV enforces both reverberation blockage and no distortion.

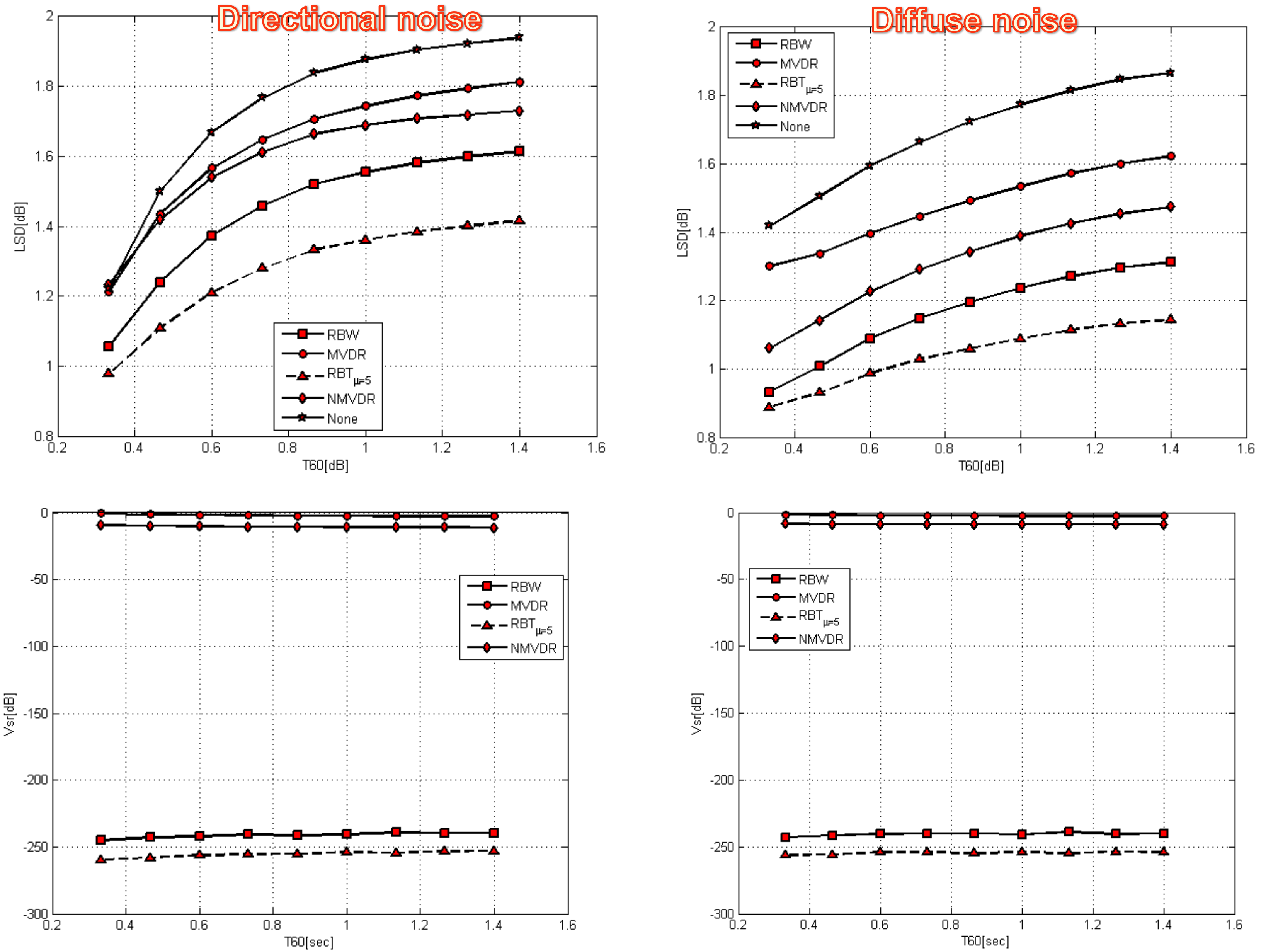
Reverberation Block Trade-off Beamformer

- The RBT Beamformer solves the constraint minimization problem
- $\bar{h}_{\mathrm{RBT},\mu}[l,k] = \underbrace{min}_{\bar{h}} J_d\left[\bar{h}[l,k]\right],$
- $\mathrm{S.T}\; J_{rn}\left[\bar{h}[l,k]\right] = \beta\phi_{V_1}[l,k]$ 1>β>0, $\bar{h}^H[l,k]\bar{r}[k] = 0$
- Using a Lagrange multiplier $\mu > 0$ we get
- $\bar{h}_{\mathrm{RBT},\mu}[l,k] = \phi_{X_{1,c}}[l,k]\Phi_{pp}^{-1}[l,k]\left[I_{NxN} - \frac{\bar{r}[k]\bar{r}^H[k]\Phi_{pp}^{-1}[l,k]}{\bar{r}^H[k]\Phi_{pp}^{-1}[l,k]\bar{r}[k]}\right] i_N$
- $\Phi_{pp}^{-1}[l,k] = \frac{1}{\mu}\left[\Phi_{vv}^{-1}[l,k] - \frac{\phi_{X_{1,c}}[l,k]\Phi_{vv}^{-1}[l,k]\, i_N i_N^T \Phi_{vv}^{-1}[l,k]}{\mu+\phi_{X_{1,c}}[l,k] i_N^T \Phi_{vv}^{-1}[l,k] i_N}\right]$
- When $\mu \rightarrow 0 \rightarrow$ NLCMV, When $\mu = 1 \rightarrow$ RBW
- When $0 < \mu < 1 \rightarrow$ lower speech distortion at the expense of residual noise reduction.
- When $\mu > 1 \rightarrow$ better noise reduction, at the expense of high speech distortion.

Influence of iSCNR T60=0.8[sec]

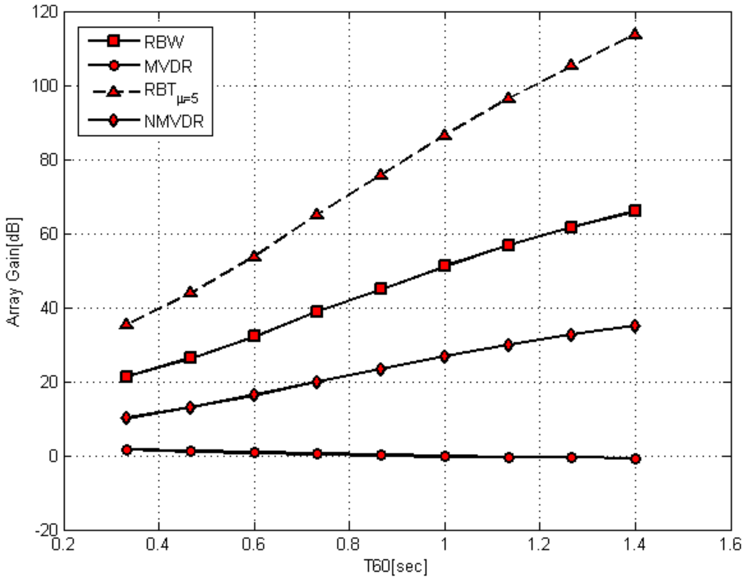


Influence of T60 iSCNR=5[dB]

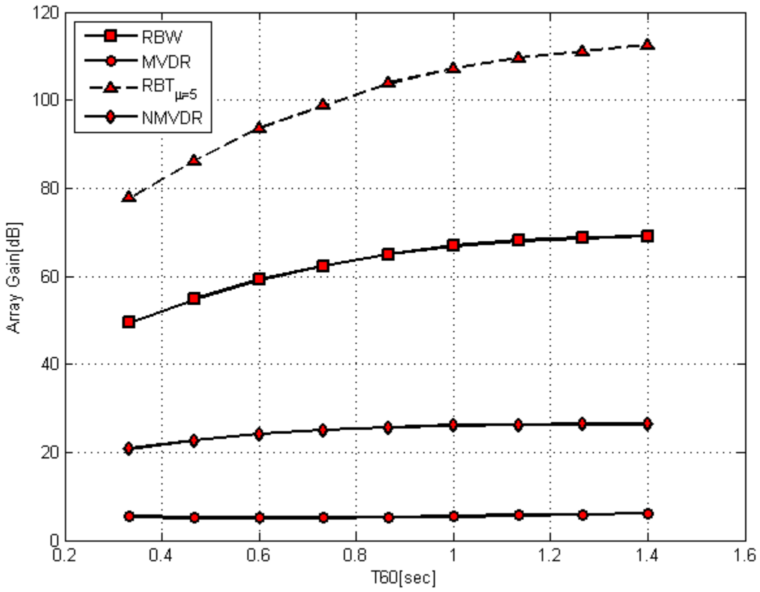


Influence of T60 iSCNR=5[dB]

Directional noise



Diffuse noise



Simulation Results
Different Beamformers

- First approach
- Directional noise field ISCNR 5[dB] T60-0.8[sec]
- Noisy Signal+ rev –
- Wiener GSC MVDR TF-mu=5

- Second approach
- Directional noise field ISCNR 5[dB] T60-0.8[sec]
- Noisy Signal+ rev –
- RBW NMVDR MVDR RBT-mu=5
-

Work Summary

- ***First approach***

➢ Manages to adapt the MVDR/Wiener/Trade-off Beamformers to reverberant environment

➢ Shows promising results in aspects of both denoising and dereverberation, compared to the TF-GSC.

- ***Second approach***

➢ Shows high reverberation blockage capabilities

➢ Outperforms the standard MVDR

➢ The reverberation suppression process is an acting part in the minimization problem

Thank you for listening

- Eliav.benisty@gmail.com