Linear and Parametric Microphone Array Processing

Part II: Linear Spatial Processing

Emanuël A.P. Habets[1] and Sharon Gannot[2]

[1]International Audio Laboratories Erlangen, Germany

[2]Faculty of Engineering, Bar-Ilan University, Israel

ICASSP 2013, Vancouver, Canada

Outline

Linear Spatial Noise Reduction Techniques I

Families of Methods

1. Fixed beamforming Combine the microphone signals using a time-invariant filter-and-sum operation (data-independent)

   [Jan and Flanagan, 1996]; [Doclo and Moonen, 2003].

2. Blind Source Separation (BSS) Considers the received signals at the microphones as a mixture of all sound sources filtered by the RIRs. Utilizes Independent Component Analysis (ICA) techniques

   [Makino et al., 2007]; TRINICON, [Buchner et al., 2004].

3. Adaptive Beamforming Combine the spatial focusing of fixed beamformers with adaptive suppression of (spectrally and spatially time-varying) background noise

   General reading: [Cox et al., 1987]; [Van Veen and Buckley, 1988]; [Van Trees, 2002].

Linear Spatial Noise Reduction Techniques II

Some Criteria

1. Adaptive optimization [Sondhi and Elko, 1986]; [Kaneda and Ohga, 1986]; [Brandstein and Ward, 2001].
2. Minimum variance distortionless response (MVDR) and GSC [Van Compernolle, 1990]; [Affes and Grenier, 1997]; [Nordholm et al., 1993]; [Hoshuyama et al., 1999]; [Gannot et al., 2001]; [Herbordt, 2005]; [Gannot and Cohen, 2008].
3. Minimum mean square error (MMSE) - GSVD based spatial Wiener filter [Doclo and Moonen, 2002a].
4. Speech distortion weighted multichannel Wiener filter (SDW-MWF) [Doclo and Moonen, 2002b]; [Spriet et al., 2004]; [Doclo et al., 2005].
5. Maximum signal to noise ratio (SNR) [Warsitz and Haeb-Umbach, 2007].
6. Linearly constrained minimum variance (LCMV) [Markovich et al., 2009].

Linear Spatial Noise Reduction Techniques III

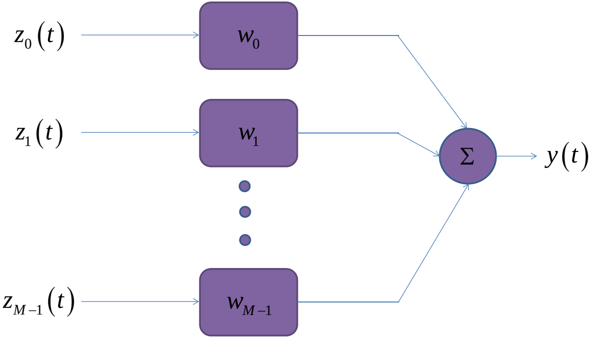
Some Books

1. Acoustic signal processing for telecommunication [Gay and Benesty, 2000].
2. Microphone Arrays: Signal Processing Techniques and Applications [Brandstein and Ward, 2001].
3. Speech Enhancement [Benesty et al., 2005].
4. Blind speech separation [Makino et al., 2007].
5. Microphone Array Signal Processing [Benesty et al., 2008a].
6. Springer handbook of speech processing [Benesty et al., 2008b].
7. Handbook on array processing and sensor networks [Haykin and Liu, 2010].
8. Speech processing in modern communication: Challenges and perspectives [Cohen et al., 2010].

Spatial Filters

Beamforming: Filter and Sum

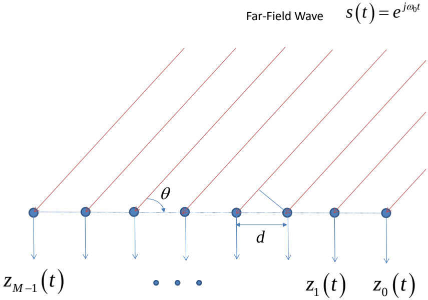
$$y(t) = \mathbf{w}^H(t)\mathbf{z}(t).$$



$\mathbf{w}$: $M \times 1$ beamforming vector of filters (or just gains).

Array Processing

Preliminaries



Narrow-band Signal

$$
\begin{aligned}
y(t) &= \sum_{m=0}^{M-1} w_m^* e^{j\omega_0(t-\tau_m)} \\
&= e^{j\omega_0 t} \sum_{m=0}^{M-1} w_m^* e^{-j\omega_0 \left(\frac{d\cos(\theta)}{c}\right)m} \\
&= e^{j\omega_0 t} \sum_{m=0}^{M-1} w_m^* e^{-j2\pi \frac{d}{\lambda_0}\cos(\theta)m}
\end{aligned}
$$

Beampattern is the DTFT of the weights

$$
y(t) = e^{j\omega_0 t} W\left(\frac{d}{\lambda_0}; \cos(\theta)\right)
$$

The Delay & Sum Beamformer

Uniform Linear Array (ULA)

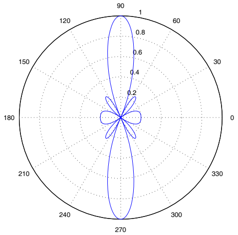
- $w_m = \frac{1}{M};\ m = 0, \ldots, M-1.$
- For simplicity, assume symmetric array.
- Steered to $\cos(\theta_0)$.
- Beampattern:

$$B(\theta) = \frac{1}{M} \cdot \frac{\sin\left(\frac{M}{2} 2\pi \frac{d}{\lambda_0}(\cos(\theta) - \cos(\theta_0))\right)}{\sin\left(\frac{1}{2} 2\pi \frac{d}{\lambda_0}(\cos(\theta) - \cos(\theta_0))\right)}$$

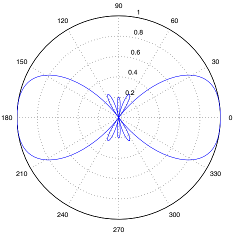
Beamformers

- Discriminate between angles.
- Can be steered by setting $\mathbf{w}$.
- Depends on the ratio $\frac{d}{\lambda_0}$.

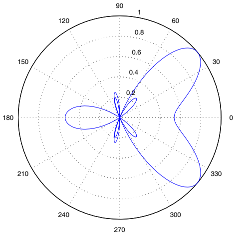
Beampattern



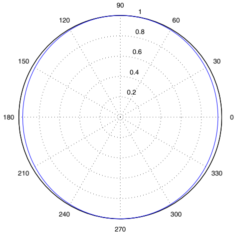
(a) $\theta_0 = 90^o$; $\frac{d}{\lambda_0} = \frac{1}{2}$



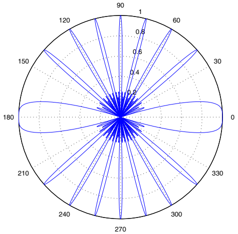
(b) $\theta_0 = 0^o$; $\frac{d}{\lambda_0} = \frac{1}{2}$



(c) $\theta_0 = 40^o$; $\frac{d}{\lambda_0} = \frac{1}{2}$

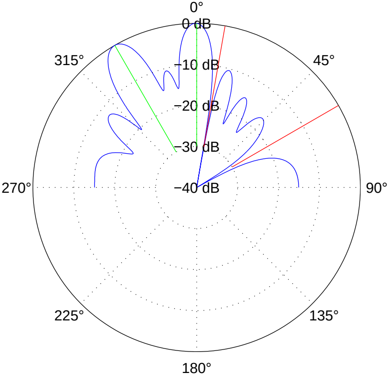


(d) $\theta_0 = 90^o$; $\frac{d}{\lambda_0} = \frac{1}{32}$



(e) $\theta_0 = 90^o$; $\frac{d}{\lambda_0} = \frac{4}{1}$

Additional Control on the Beampattern



- 10 microphone uniform linear array.
- 2 Desired sources in green and 2 interfering sources in red.
- Can be obtained by applying the LCMV criterion.

Directivity and White Noise Gain (WNG) [Van Trees, 2002] I

Definitions

- Propagation vector: $\mathbf{u} = \left[\, \sin(\theta)\cos(\phi)\ \sin(\theta)\sin(\phi)\ \cos(\theta) \,\right]^T$.
- Beampattern: $B(\phi, \theta)$.
- Beampower: $P(\phi, \theta) = |B(\phi, \theta)|^2$.

Directivity

- Assume that desired response is normalized: $P(\phi_0, \theta_0) = 1$.
- $D = \left(\frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} \sin(\theta) P(\phi, \theta) d\phi d\theta \right)^{-1}$.
- Directivity Index: $DI = 10 \log_{10}(D)[\text{dB}]$.
- Maximum Directivity for ULA with $d = \frac{\lambda}{2}$ is M. It is achieved by the delay & sum beamformer.

Directivity and White Noise Gain (WNG) [Van Trees, 2002] II

White Noise Gain

- SNR improvement for spatially white input: $A_{\mathbf{w}} = \frac{\mathrm{SNR_{out}}}{\mathrm{SNR_{in}}} = \|\mathbf{w}\|^{-2}$.
- Sensitivity to array weight imperfections and sensor misalignment is $T_{se} = \frac{1}{A_{\mathbf{w}}} = \|\mathbf{w}\|^2$ (hence, large WNG is better).

Maximum Directivity [Parsons, 1987]

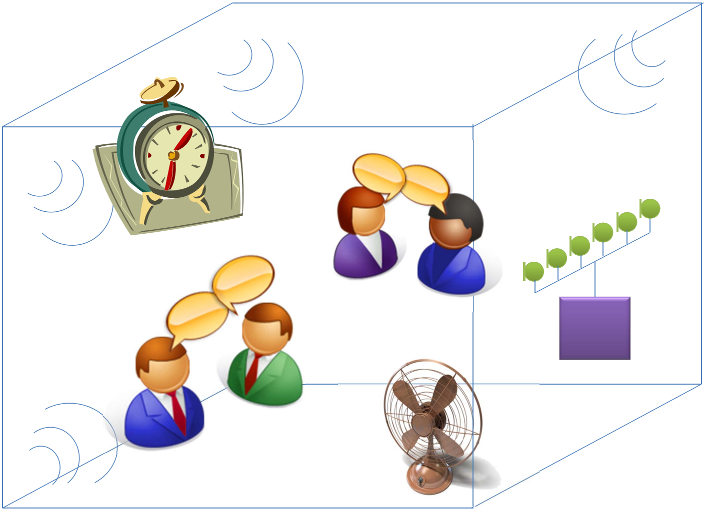
- MVDR criterion for diffuse noise field: super-directive beamformer.
- Obtained for linear endfire array with vanishingly small inter-sensor distance ($d \to 0$)!
- Maximum achievable directivity is M^2.
- In that case $T_{se} \to \infty$ [Gilbert and Morgan, 1955] (see extention AASP-L4, Levin, Gannot and Habets).
- Robust design limiting the sensitivity exists [Cox et al., 1986].
- Forms the basis of differential microphone arrays [Elko, 1996].

From Geometry to Linear Algebra

Array Design for Speech Propagating in Acoustic Environments

- Beampatterns:
  Array response as a function of the angle of arrival (AoA).
- In reverberant environments (especially for low DRR), sound propagation is more involved than merely the AoA.
- The steering vector (comprised of the AoA) generalizes to acoustic transfer function (ATF).
- The ATF summarizes all arrivals of the speech signals.
- The vector of received signals is treated as a vector in an abstract linear space.
- Linear Algebra methods are utilized to construct beamformers.
- AoA becomes less prominent.

A Noisy Example



Multiple Wideband Signals (e.g. Speech)

Short-Time Fourier Transform (STFT) - Multiplicative Transfer Function (MTF) Approximation

$t \overset{\text{STFT}}{\Longrightarrow} \{\ell, k\}$; Convolution $\overset{\text{STFT}}{\Longrightarrow}$ Multiplication (for long enough frames).

Microphone Signals ($m = 0, \ldots, M-1$):

$$z_m(\ell, k) = \sum_{j=1}^{P_d} s_j^d h_{jm}^d + \sum_{j=1}^{P_i} s_j^i h_{jm}^i + \sum_{j=1}^{P_n} s_j^n h_{jm}^n + n_m$$

Vector Formulation

$$\mathbf{z}(\ell, k) = \mathbf{H}^d \mathbf{s}^d + \mathbf{H}^i \mathbf{s}^i + \mathbf{H}^n \mathbf{s}^n + \mathbf{n} \triangleq \mathbf{H}\mathbf{s} + \mathbf{n}.$$

$$P = P_d + P_i + P_n \ \leq M$$

Beamforming in the STFT Domain

Apply filter & sum beamforming independently for each frequency bin.

Power Spectral Density (PSD)

Microphone Signals

$$\mathbf{z}(\ell, k) = \mathbf{H}^d \mathbf{s}^d + \mathbf{H}^i \mathbf{s}^i + \mathbf{H}^n \mathbf{s}^n + \mathbf{n} \triangleq \mathbf{H}\mathbf{s} + \mathbf{n}$$

The PSD of the Various Components:

- Stationary Sources: $\mathbf{\Phi}_{zz}^{\mathrm{stat}} = \mathbf{H}^n \mathbf{\Phi}_{s^n s^n} \left(\mathbf{H}^n\right)^H + \mathbf{\Phi}_{nn}$.
- Constraints Sources:

$$\mathbf{H}\mathbf{\Phi}_{ss}\mathbf{H}^H \triangleq \mathbf{H}^d \mathbf{\Phi}_{s^d s^d} \left(\mathbf{H}^d\right)^H + \mathbf{H}^i \mathbf{\Phi}_{s^i s^i} \left(\mathbf{H}^i\right)^H + \mathbf{H}^n \mathbf{\Phi}_{s^n s^n} \left(\mathbf{H}^n\right)^H.$$

- Microphone Signals: $\mathbf{\Phi}_{zz}(\ell, k) = \mathbf{H}\mathbf{\Phi}_{ss}\mathbf{H}^H + \mathbf{\Phi}_{nn}$.
- Noise+Interference Sources:

$$\mathbf{\Phi}_{vv}(\ell, k) \triangleq \mathbf{H}^i \mathbf{\Phi}_{s^i s^i} \left(\mathbf{H}^i\right)^H + \mathbf{H}^n \mathbf{\Phi}_{s^n s^n} \left(\mathbf{H}^n\right)^H + \mathbf{\Phi}_{nn}$$

Linearly Constrained Minimum Variance Beamformer

[Er and Cantoni, 1983]; [Van Veen and Buckley, 1988]

LCMV Criterion

- $y(\ell, k) = \mathbf{w}^H(\ell, k)\mathbf{z}(\ell, k)$.
- Let $\mathbf{\Phi}_{nn} = E\{\mathbf{n}\mathbf{n}^H\}$ be the $M \times M$ correlation matrix of the unconstraint sources.
- Minimize noise power $\mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$
  Such that a linear constraint set is satisfied: $\mathbf{C}^H\mathbf{w} = \mathbf{g}$.
- $\mathbf{C}$: $M \times P$ constraints matrix.
- $\mathbf{g}$: $P \times 1$ response vector.

Closed-form Solution

$$\mathbf{w}(\ell, k) = \mathbf{\Phi}_{nn}^{-1}\mathbf{C}\left(\mathbf{C}^H\mathbf{\Phi}_{nn}^{-1}\mathbf{C}\right)^{-1}\mathbf{g}$$

Linearly Constrained Minimum Power (LCMP) Beamformer

[Van Trees, 2002]

LCMV vs. LCMP

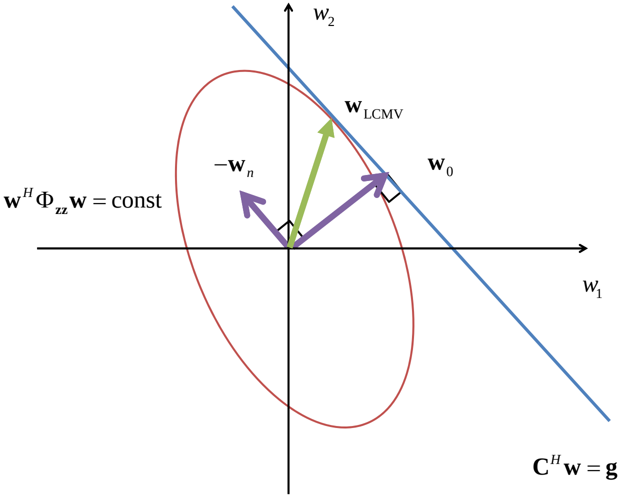
- Assume $\mathbf{C} = \mathbf{H}$ (all directional signals constrained).

$$\begin{aligned}
\mathbf{w}_{\mathrm{LCMP}} &= \underset{\mathbf{w}}{\operatorname{argmin}}\{\mathbf{w}^H \boldsymbol{\Phi}_{zz}\mathbf{w} \text{ s.t. } \mathbf{H}^H\mathbf{w} = \mathbf{g}\} \\
&= \underset{\mathbf{w}}{\operatorname{argmin}}\{\mathbf{w}^H (\mathbf{H}\boldsymbol{\Phi}_{ss}\mathbf{H}^H + \boldsymbol{\Phi}_{nn})\mathbf{w} \text{ s.t. } \mathbf{H}^H\mathbf{w} = \mathbf{g}\} \\
&= \underset{\mathbf{w}}{\operatorname{argmin}}\{\mathbf{g}^H \boldsymbol{\Phi}_{ss}\mathbf{g} + \mathbf{w}^H \boldsymbol{\Phi}_{nn}\mathbf{w} \text{ s.t. } \mathbf{H}^H\mathbf{w} = \mathbf{g}\} \\
&= \underset{\mathbf{w}}{\operatorname{argmin}}\{\mathbf{w}^H \boldsymbol{\Phi}_{nn}\mathbf{w} \text{ s.t. } \mathbf{H}^H\mathbf{w} = \mathbf{g}\} = \mathbf{w}_{\mathrm{LCMV}}
\end{aligned}$$

- If $\mathbf{H}$ is not accurately estimated, the LCMP beamformer exhibits self-cancellation and hence severe speech distortion.
- It is quite common in the literature to use only the term LCMV for both beamformers.

LCMV Minimization

Graphical Interpretation [Frost III, 1972]



The Minimum Variance Distortionless Beamformer

[Affes and Grenier, 1997]; [Hoshuyama et al., 1999]; [Gannot et al., 2001]

Beamformer Design:

- One desired signal $\Rightarrow$ Single constraint ($P = 1$).
- "Steer a beam" to desired source and minimize other directions.
- $\mathbf{C} = \mathbf{h}^d; \quad \mathbf{g} = 1.$

Closed-form Solution (MPDR eq. MVDR):

$$\mathbf{w}(\ell, k) = \frac{\mathbf{\Phi}_{zz}^{-1}\mathbf{h}^d}{(\mathbf{h}^d)^H \mathbf{\Phi}_{zz}^{-1}\mathbf{h}^d} = \frac{\mathbf{\Phi}_{nn}^{-1}\mathbf{h}^d}{(\mathbf{h}^d)^H \mathbf{\Phi}_{n}^{-1}\mathbf{h}^d}$$

Output signal:

$$y = s^d + \text{residual noise and interference signals}$$

Multiple Speech Distortion Weighted Multichannel Wiener Filter (MSDW-MWF)[Markovich-Golan et al., 2012b]

Notation (Reminder)

- Received signals: $\mathbf{z}(\ell, k) = \mathbf{H}\mathbf{s} + \mathbf{n}$.
- $P < M$ constrained sources: $\mathbf{s}(\ell, k) \triangleq \begin{bmatrix} s_1 & \cdots & s_P \end{bmatrix}^T$ and respective ATFs: $\mathbf{H}(\ell, k) \triangleq \begin{bmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_P \end{bmatrix}$.
- Sources covariance matrix: $\mathbf{\Phi}_{ss} = \text{diag}\left\{\phi_{s_1 s_1}, \ldots, \phi_{s_P s_P}\right\}$.
- Microphones covariance matrix: $\mathbf{\Phi}_{zz} \triangleq \mathbf{H}\mathbf{\Phi}_{ss}\mathbf{H}^\dagger + \mathbf{\Phi}_{nn}$.

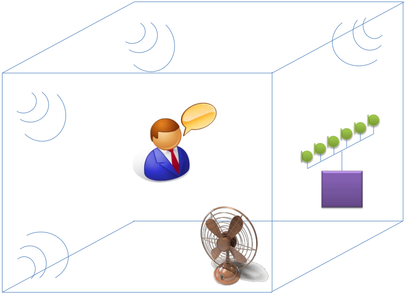
MSDW-MWF

- Control the distortion of each individual source.
- Minimize the weighted mean square error (MSE).
- Desired response for all constrained signals: $d(\ell, k) \triangleq \mathbf{g}^H \mathbf{s}(\ell, k)$.
- The beamformer output: $y(\ell, k) = \mathbf{w}^H \mathbf{z}(\ell, k)$.
- MSE: $\mathrm{E}\left\{|d(\ell) - y(\ell)|^2\right\}$.

Speech enhancement with a Single Source I

Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF)

[Doclo and Moonen, 2002b]; [Spriet et al., 2004]; [Doclo et al., 2005]



Speech enhancement with a Single Source II

Speech Distortion Weighted Multichannel Wiener Filter (SDW-MWF)

[Doclo and Moonen, 2002b]; [Spriet et al., 2004]; [Doclo et al., 2005]

The Multichannel Wiener Filter (MWF) Criterion

$$J_w \triangleq \mathrm{E}\left\{ |d(\ell) - y(\ell)|^2 \right\} = \left| g - (\mathbf{h}^d)^H \mathbf{w} \right|^2 \phi_{s^d s^d} + \mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$$

The Speech Distortion Weighted (SDW)-MWF Criterion

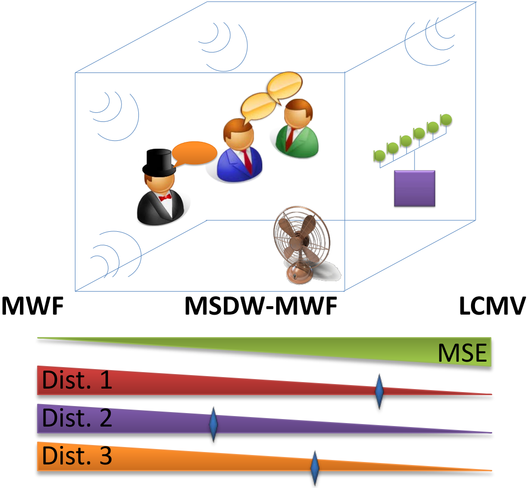
$$J_{\text{SDW-MWF}} = \left| g - (\mathbf{h}^d)^H \mathbf{w} \right|^2 \phi_{s^d s^d} + \mu \mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$$

The Speech Distortion Weighted (SDW)-MWF Solution

$$\mathbf{w} = \frac{\phi_{s^d s^d} \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d}{\mu + \phi_{s^d s^d} (\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d} g$$

Speech Enhancement with Multiple Sources I

[Markovich-Golan et al., 2012b]



Speech Enhancement with Multiple Sources II

[Markovich-Golan et al., 2012b]

The MSDW-MWF Criterion

$$J_{\text{MSDW-MWF}} \triangleq \left(\mathbf{g} - \mathbf{H}^H \mathbf{w}\right)^H \mathbf{\Lambda} \mathbf{\Phi}_{ss} \left(\mathbf{g} - \mathbf{H}^H \mathbf{w}\right) + \mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$$

- Diagonal weights matrix: $\mathbf{\Lambda} \triangleq \text{diag}\left\{\lambda_1, .., \lambda_P\right\}$.

MSDW-MWF Beamformer

$$\mathbf{w} \triangleq \left(\mathbf{H} \mathbf{\Lambda} \mathbf{\Phi}_{ss} \mathbf{H}^H + \mathbf{\Phi}_{nn}\right)^{-1} \mathbf{H} \mathbf{\Lambda} \mathbf{\Phi}_{ss} \mathbf{g}$$

Special Cases of $\boldsymbol{\Lambda}$

MWF

- $\boldsymbol{\Lambda} = \mathbf{I}$.
- $\mathbf{w} = \boldsymbol{\Phi}_{zz}^{-1}\mathbf{H}\boldsymbol{\Phi}_{ss}\mathbf{g}$.

SDW-MWF (Reminder: Single Source of Interest)

- $\boldsymbol{\Lambda} = \mu^{-1}$.
- $\mathbf{w} = \left(\mathbf{h}^d \phi_{s^d s^d} (\mathbf{h}^d)^H + \mu \boldsymbol{\Phi}_{nn}\right)^{-1} \mathbf{h}^d \phi_{s^d s^d} g$.
- $\lim_{\mu\to 0} \mathbf{w} = \frac{\boldsymbol{\Phi}_{nn}^{-1}\mathbf{h}^d}{(\mathbf{h}^d)^H \boldsymbol{\Phi}_{nn}^{-1}\mathbf{h}^d} g$ (MVDR eq. MPDR).

LCMV

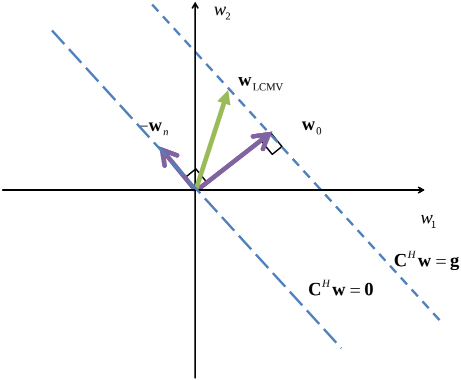
- $\boldsymbol{\Lambda} = \mu^{-1}\boldsymbol{\Phi}_{ss}^{-1}$.
- $\lim_{\mu\to 0} \mathbf{w} = \boldsymbol{\Phi}_{nn}^{-1}\mathbf{H}\left(\mathbf{H}^H\boldsymbol{\Phi}_{nn}^{-1}\mathbf{H}\right)^{-1}\mathbf{g}$ (LCMV eq. LCMP).

The Generalized Sidelobe Canceller Implementation

For Constrained Minimization [Griffiths and Jim, 1982]

Split the Beamformer

- $\mathbf{w} = \mathbf{w}_0 - \mathbf{w}_n$.
- Constraints Subspace: $\mathbf{w}_0 \in \mathrm{Span}\{\mathbf{C}\}$.
- Null Subspace: $\mathbf{w}_n \in \mathcal{N}\{\mathbf{C}\}$.
- $\mathbf{w}_n \triangleq \mathbf{Bq}$.
- $\mathbf{B}$: $M \times (M - P)$ matrix. Spans the Null Subspace.
- $\mathbf{q}$: vector of $M - P$ filters.
- $\Rightarrow \mathbf{w} = \mathbf{w}_0 - \mathbf{Bq}$.



The Generalized Sidelobe Canceller Implementation

GSC Output

$$y = \mathbf{w}_0^H \mathbf{z} - \mathbf{q}^H \underbrace{\mathbf{B}^H \mathbf{z}}_{\mathbf{u}(\ell,k)}$$

Constraints Subspace ($\mathbf{w}_0 \in \mathrm{Span}\{C\}$):

$$\mathbf{w}_0(\ell, k) \triangleq \mathbf{C}\left(\mathbf{C}^H \mathbf{C}\right)^{-1} \mathbf{g}$$

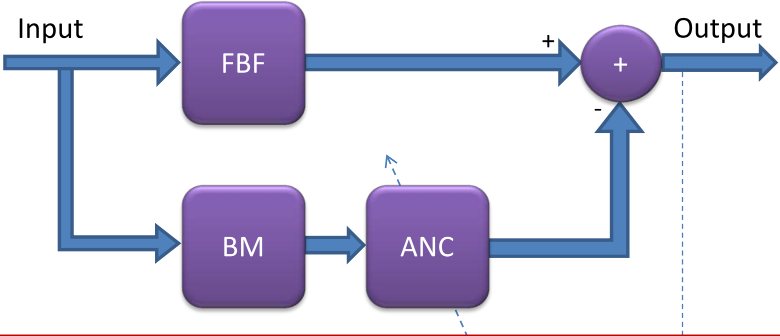
Null Subspace (columns of **B** span $\mathcal{N}\{C\}$):

$$\mathbf{B}(\ell, k) \triangleq \mathbf{I}_{\mathrm{M}\times\mathrm{M}} - \mathbf{C}\left(\mathbf{C}^H \mathbf{C}\right)^{-1} \mathbf{C}^H; \ (\text{verify } \mathbf{B}^H \mathbf{C} = \mathbf{0}).$$

Noise Cancelling Filters (orthogonality principle):

$$E\left\{\mathbf{u}\left(\mathbf{z}^H \mathbf{w}_0 - \mathbf{u}^H \mathbf{q}\right)\right\} \Rightarrow \mathbf{q}(\ell, k) = \left(\mathbf{B}^H \mathbf{\Phi}_{zz} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{\Phi}_{zz} \mathbf{w}_0$$

The GSC Structure [Griffiths and Jim, 1982]



GSC Blocks

- Fixed beamformer (FBF) - satisfies the constraints ($\mathbf{w}_0$).
- Blocking matrix (BM) - generates $M - P$ unconstrained signals ($\mathbf{B}$).
- Noise canceller (ANC) - adaptively (LMS) suppresses the residual noise utilizing $M - P$ degrees of freedom (DoF) ($\mathbf{q}$) [Widrow et al., 1975]; [Shynk, 1992].

GSC Implementation of the MVDR Beamformer

Blocks [Griffiths and Jim, 1982]:

$$\mathbf{w}_0(\ell, k) = \frac{\mathbf{h}^d}{\|\mathbf{h}^d\|^2}$$

$$\mathbf{B}(\ell, k) \triangleq \mathbf{I}_{\mathrm{M}\times\mathrm{M}} - \frac{\mathbf{h}^d \left(\mathbf{h}^d\right)^H}{\|\mathbf{h}^d\|^2}$$

$$\mathbf{q}(\ell, k) = \left(\mathbf{B}^H \mathbf{\Phi}_{zz} \mathbf{B}\right)^{-1} \mathbf{B}^H \mathbf{\Phi}_{zz} \mathbf{w}_0$$

$\mathbf{q}(\ell, k)$ can be recursively updated using the LMS algorithm [Shynk, 1992].

The Relative Transfer Function GSC (TF-GSC)

Relax Dereverberation Requirement [Gannot et al., 2001]

Modified Constraint Set:

$$\mathbf{C}(\ell, k) = \mathbf{h}^d(\ell, k); \quad \tilde{\mathbf{g}}(\ell, k) = (h_0^d(\ell, k))^*$$

$$\Rightarrow (\mathbf{h}^d(\ell, k))^H \mathbf{w} = (h_0^d(\ell, k))^*$$

Equivalent to:

$$\tilde{\mathbf{C}}(\ell, k) = \tilde{\mathbf{h}}^d(\ell, k) \triangleq \frac{\mathbf{h}^d}{h_0^d} = \left[1 \ \frac{h_1^d}{h_0^d} \ \ldots \ \frac{h_{M-1}^d}{h_0^d} \right]^T$$

$$\mathbf{g}(\ell, k) = 1.$$

The Relative Transfer Function

$\tilde{\mathbf{h}}^d(\ell, k)$ - The ratio of all ATFs to the reference ATF (#0 in this case).

The Transfer Function GSC utilizing RTF I

[Gannot et al., 2001]

FBF:

$$\mathbf{w}_0(\ell, k) = \tilde{\mathbf{h}}^d / \|\tilde{\mathbf{h}}^d\|^2$$

Blocking matrix

- Noise reference signals: $\mathbf{u} = \mathbf{B}^H \mathbf{z}$.
- Efficient implementation of the BM with $M - 1$ filters exists.

$$\mathbf{B}(\ell, k) = \begin{bmatrix} -(\tilde{h}_1^d)^* & -(\tilde{h}_2^d)^* & \dots & -(\tilde{h}_{M-1}^d)^* \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \dots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Compactly, $u_0 = 0$; $u_m = z_m - \tilde{h}_m^d z_0$, $m \neq 0$.

The Transfer Function GSC utilizing RTF II

[Gannot et al., 2001]

Output signal:

$$y(\ell, k) = \underbrace{h_0^d s^d}_{\tilde{s}_0^d(\ell,k)} + \text{residual noise and interference signals}$$

Tradeoff:

Noise reduction is sacrificed if dereverberation is required [Habets et al., 2010].

Multi-Constraint Beamformer

Based on LCMV Beamforming [Markovich et al., 2009]

Applications:

- Conference call scenario with multiple participants.
- Hands-free cellular phone conversation in a car environment with several passengers.
- Cocktail Party scenario, in which desired conversation blend with many simultaneous conversations.

Problem Formulation (Reminder):

$$\mathbf{z} = \mathbf{H}^d \mathbf{s}^d + \mathbf{H}^i \mathbf{s}^i + \mathbf{H}^n \mathbf{s}^n + \mathbf{n}$$

GSC Formulation

GSC Implementation of the LCMV (exists [Breed and Strauss, 2002])

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{B}\mathbf{q}$$

Fixed Beamformer (in Constraints Subspace)

$$\mathbf{w}_0 = \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{g}$$

Blocking Matrix (in Constraints Null Subspace)

$$\mathbf{B} = \mathbf{I}_{M\times M} - \mathbf{C}\left(\mathbf{C}^H\mathbf{C}\right)^{-1}\mathbf{C}^H$$

Can be efficiently implemented: $(M-P)\times P$ filters [Markovich-Golan et al., 2012a].

Noise Canceler

$$\mathbf{q} = \left(\mathbf{B}^H\mathbf{\Phi}_{zz}\mathbf{B}\right)^{-1}\mathbf{B}^H\mathbf{\Phi}_{zz}(\ell,k)\mathbf{w}_0$$

The Constraints Set

Original

$$\mathbf{C} \triangleq \mathbf{H} = \left[\mathbf{H}^d \ \mathbf{H}^i \ \mathbf{H}^n \right]$$

$$\mathbf{g} \triangleq \left[\underbrace{1 \ \ldots \ 1}_{P_d} \ \underbrace{0 \ \ldots \ 0}_{P-P_d} \right]^T$$

LCMV output

Since all directional signals are constrained, $\mathbf{q} = 0$ if $\mathbf{\Phi}_{nn}$ is spatially-white.

$$y = \sum_{j=1}^{P_d} s_j^d + \text{noise components}$$

An Equivalent Constraints Set

An orthonormal basis $\mathbf{Q}$:

- Noise+Interference Sources PSD (no desired sources):

$$\mathbf{\Phi}_{vv}(\ell,k) \triangleq \mathbf{H}^i \mathbf{\Phi}_{s^i s^i} \left(\mathbf{H}^i\right)^H + \mathbf{H}^n \mathbf{\Phi}_{s^n s^n} \left(\mathbf{H}^n\right)^H + \mathbf{\Phi}_{nn}$$

- Eigenvalue decomposition: $\mathbf{\Phi}_{vv}(\ell,k) = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$.
- Replace $\left[\, \mathbf{H}^i \;\; \mathbf{H}^n \,\right]$ with $\mathbf{Q}$, comprised of the eigenvectors that correspond to the significant eigenvaleues (# of significant eigenvaleues is, hopefully, $P_i + P_n$).

$$\dot{\mathbf{C}}^H \mathbf{w} = \mathbf{g}$$

$$\dot{\mathbf{C}} \triangleq \left[\, \mathbf{H}^d \;\; \mathbf{Q} \,\right]$$

A Modified Constraints Set

Relax the dereverberation requirements using RTFs:

$$\tilde{\mathbf{g}} \triangleq \Big[\underbrace{(h_{10}^d)^* \ \ldots \ (h_{P_d 0}^d)^*}_{P_d} \ \underbrace{0 \ \ldots \ 0}_{P-P_d} \Big]^T$$

$$\Rightarrow \tilde{\mathbf{h}}_j^d \triangleq \mathbf{h}_j^d / h_{j0}^d; \quad \mathbf{g} \triangleq \Big[\underbrace{1 \ \ldots \ 1}_{P_d} \ \underbrace{0 \ \ldots \ 0}_{P-P_d} \Big]^T$$

Hence, a modified constraints set: $\tilde{\mathbf{C}} \triangleq \left[\tilde{\mathbf{H}}^d \ \mathbf{Q} \right]$.

LCMV output:

$$y = \sum_{j=1}^{P_d} h_{j0}^d s_j^d + \text{noise components}$$

LCMV and MVDR Beamformers using ATFs & RTFs

Features & Drawbacks of the Proposed Beamformers

+ No need for sensor position calibration.
+ Beamformer components estimated from the received signals.
+ High amount of noise and interference reduction.
+ Low speech distortion.
- Number of filter coefficients to be estimated tends to be very large.
- Hence frame length tends to be large as well (can be mitigated at the expense of increased complexity. See CTF approximation).
- Limited performance in diffuse noise fields (can be mitigated by using postfiltering).

Performance Analysis

Theoretical and practical comparison of MVDR and LCMV beamformers can be found in [Markovich et al., 2008]; [Habets et al., 2009].

Objective Performance Measures

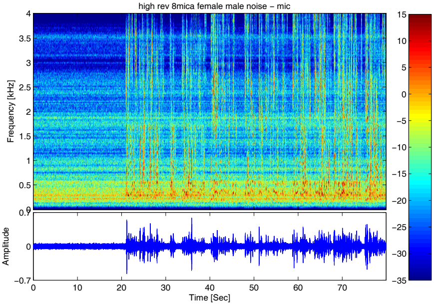
Desired > nonstationary by 6dB; Desired > stationary by 13dB

T_{60}	Source	FBF SIR			Total SIR			SSNR	LSD
		s_1^i	s_2^i	s_1^n	s_1^i	s_2^i	s_1^n		
150ms	s_1^d	18.8	22.4	19.1	18.5	21.7	24.0	9.6	1.1
	s_2^d	18.7	22.3	19.1	18.7	21.9	24.2	10.2	1.7
200ms	s_1^d	18.1	20.6	19.5	18.3	21.3	24.7	7.2	1.5
	s_2^d	18.1	20.7	19.6	18.9	21.9	25.2	8.4	2.0
250ms	s_1^d	18.5	19.8	19.9	18.4	20.9	24.5	7.0	1.8
	s_2^d	18.5	19.8	19.9	19.4	22.0	25.6	7.7	2.4
300ms	s_1^d	17.6	17.6	19.5	18.3	19.3	23.6	6.9	2.2
	s_2^d	17.4	17.5	19.3	18.6	19.7	24.0	7.7	1.8

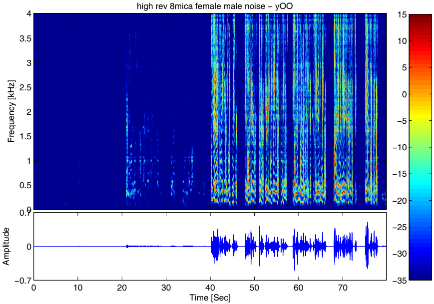
Table: 2 desired sources, 2 competing speakers, 1 stationary noise source. The desired signal at the input is larger than the competing signal by 6dB and larger than the stationary noise by 13dB. 10 microphones simulated environment. LSD & SSNR are the distortion measures between desired signal components at the output and at the input microphone #1.

Single Desired Speaker

Directional Noise Field



(a) Noisy at mic. #1

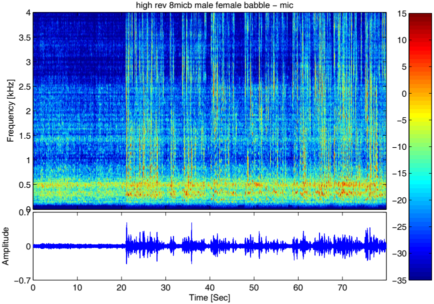


(b) Enhanced signal

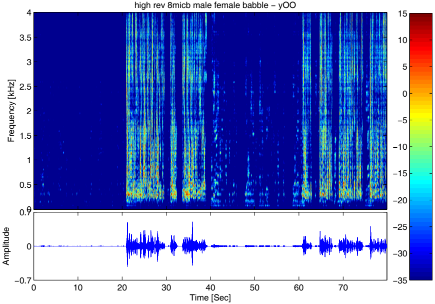
Figure: Female (desired) and male (interference) with Directional noise. 8 microphones recorded at BIU acoustic lab set to $T_{60} = 300$ms.

Single Desired Speaker

Pseudo-Babble Noise Field



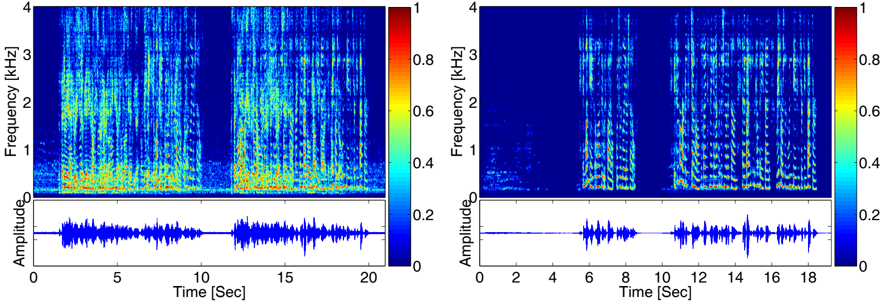
(a) Noisy at mic. #1



(b) Enhanced signal

Figure: Male (desired) and Female (interference) contaminated by pseudo-babble noise. 8 microphones recorded at BIU acoustic lab set to $T_{60} = 300$ms.

Multi-Speaker

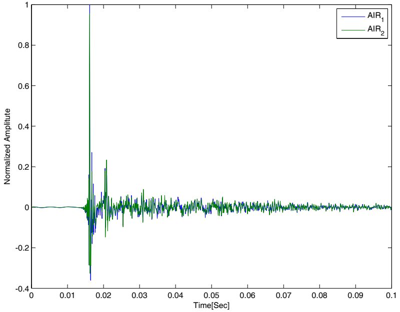


(a) Noisy at mic. #1 (b) Enhanced signal

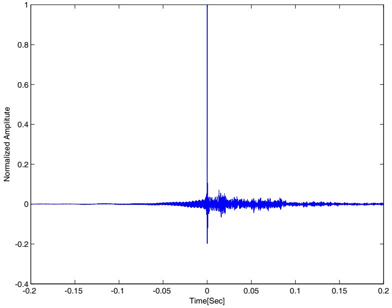
Figure: 1 desired source and 3 competing speakers. 8 microphones recorded at BIU acoustic lab set to $T_{60} = 300$ms.

Approximately 20dB SIR and SNR improvement.

The Importance of the RTF



(a) Room Impulse Responses



(b) Relative Impulse Response

Features

- Generalizes the time difference of arrival (TDOA) to ratio of ATFs.
- Usually exhibits "better behaviour" than the ATF.
- RTF is equivalent to Interaural Transfer Function (ITF).
- Drawback: Non-causal (in severe cases can cause "pre-echo").

Relative Transfer Function Estimation

Single Desired Source with Stationary Noise

System Perspective:

$$z_m(\ell, k) = \tilde{h}_m^d(\ell, k)(\ell, k) z_0(\ell, k) + u_m(\ell, k)$$

System Identification:

$$\hat{\Phi}_{z_m z_0}(\ell, k) = \tilde{h}_m^d(\ell, k) \hat{\Phi}_{z_0 z_0}(\ell, k) + \Phi_{u_m z_0}(\ell, k) + \varepsilon_m(\ell, k)$$

Estimation is Biased:

$u_m(\ell, k)$ and $z_0(\ell, k)$ are correlated $\Rightarrow$ Biased estimator for $\tilde{h}_m^d(\ell, k)$.

Relative Transfer Function Estimation I

Based on Speech Non-stationarity [Shalvi and Weinstein, 1996]; [Gannot et al., 2001]

Assumptions:

- System is Time-Invariant.
- Noise has only stationary components.
- Speech is non-stationary (use frames ℓ_i, $i = 1, \ldots, I$).

$$\begin{bmatrix} \hat{\Phi}_{z_m z_0}(\ell_1, k) \\ \hat{\Phi}_{z_m z_0}(\ell_2, k) \\ \vdots \\ \hat{\Phi}_{z_m z_0}(\ell_I, k) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}_{z_0 z_0}(\ell_1, k) & 1 \\ \hat{\Phi}_{z_0 z_0}(\ell_2, k) & 1 \\ \vdots & \\ \hat{\Phi}_{z_0 z_0}(\ell_I, k) & 1 \end{bmatrix} \begin{bmatrix} \tilde{h}_m^d(k) \\ \Phi_{u_m z_0}(k) \end{bmatrix} + \begin{bmatrix} \varepsilon_m(\ell_1, k) \\ \varepsilon_m(\ell_2, k) \\ \vdots \\ \varepsilon_m(\ell_I, k) \end{bmatrix}$$

Relative Transfer Function Estimation II

Based on Speech Non-stationarity [Shalvi and Weinstein, 1996]; [Gannot et al., 2001]

Solution

For $m = 1, \ldots, M-1$:

$$\hat{\tilde{h}}_m^d(k) = \frac{< \hat{\Phi}_{z_m z_0} \hat{\Phi}_{z_0 z_0} > (k) - < \hat{\Phi}_{z_m z_0} > (k) < \hat{\Phi}_{z_0 z_0} > (k)}{< \hat{\Phi}^2_{z_0 z_0} > (k) - < \hat{\Phi}_{z_0 z_0} >^2 (k)}$$

where, T_i the length of segment T_i and

$$< \Psi > (k) = \frac{\sum_{i=1}^{I} T_i \Psi(\ell_i, k)}{\sum_{i=1}^{I} T_i}.$$

An extension to two nonstationary sources in stationary noise exists [Reuven et al., 2008].

Alternative Estimation Procedures

- Assume direct-path model for the RIR and use TDOA estimation.
- Use speech presence probability and spectral subtraction [Cohen, 2004].
- ...

Subspace tracking [Affes and Grenier, 1997]

- Normalize by the, assumed to be known, norm:

$$\mathbf{z} = \mathbf{h}_d s^d + \mathbf{n} = \frac{\mathbf{h}_d}{\|\mathbf{h}_d\|}(\|\mathbf{h}_d\| s^d) + \mathbf{n} \triangleq \bar{\mathbf{h}}_d \bar{s}^d + \mathbf{n}$$

- Use PASTd [Yang, 1995] to recursively track the rank-1 eigenvector:

$$\hat{\bar{\mathbf{h}}}_d(\ell+1) = \hat{\bar{\mathbf{h}}}_d(\ell) + \mu(\ell)\mathbf{u}(\ell)\bar{y}^*_{\mathrm{FBF}}(\ell)$$

where, $\bar{y}_{\mathrm{FBF}} = \bar{\mathbf{h}}_d^H \mathbf{z}$. $\hat{\mathbf{h}}_d$ is obtained by using the ATF norm.

- Related to robust GSC [Hoshuyama et al., 1999].

Multi-Sources Case [Markovich et al., 2009]

Implementing the GSC Necessitates:

- Desired sources RTFs, $\tilde{\mathbf{H}}^d(\ell, k)$.
- Interferences subspace basis, $\mathbf{Q}(\ell, k)$.

Assumptions and Observations

- The ATFs are slowly-time varying.
- Segments with non-overlapping activity of desired and interference speakers are available.
- Double-talk within the group is allowed.
- Stationary sources are always active.

Interferences Subspace Estimation

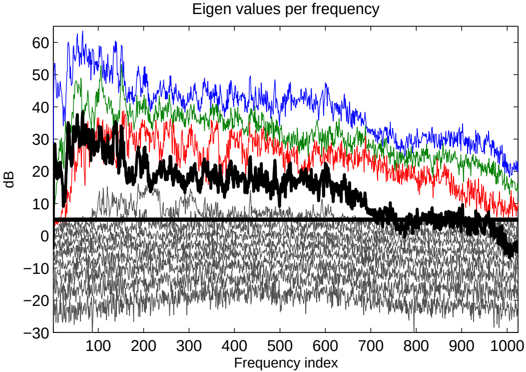
Step 1

EVD and Pruning

- Estimate the signals subspace at each time segment without any desired sources active

$$\hat{\mathbf{\Phi}}_{zz}(\ell_i, k) = \bar{\mathbf{E}}_i \mathbf{\Lambda}_i \bar{\mathbf{E}}_i^H$$

- All eigenvectors corresponding to "weak" eigenvalues are discarded



Interferences Subspace Estimation

Step 2

Union of Estimates

- Straightforward:

$$\mathbf{E}(k) \triangleq \bigcup_{i=1}^{N_{seg}} \bar{\mathbf{E}}_i(k)$$

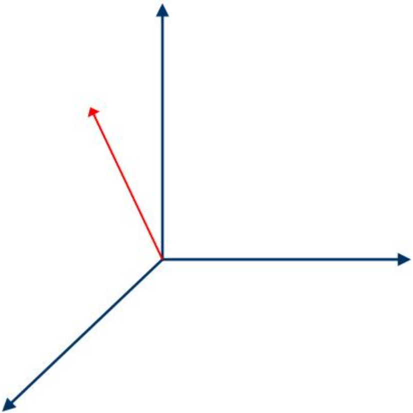
- Practical use QRD

$$\left[\bar{\mathbf{E}}_1(k)\bar{\mathbf{\Lambda}}_1^{\frac{1}{2}}(k) \ \ldots \ \bar{\mathbf{E}}_{N_{seg}}(k)\bar{\mathbf{\Lambda}}_{N_{seg}}^{\frac{1}{2}}(k) \right] \mathbf{P}(k) = \mathbf{Q}(k)\mathbf{R}(k)$$

- Discard vectors from the basis $\mathbf{Q}(k)$ that correspond to “weak” coefficients in $\mathbf{R}(k)$.

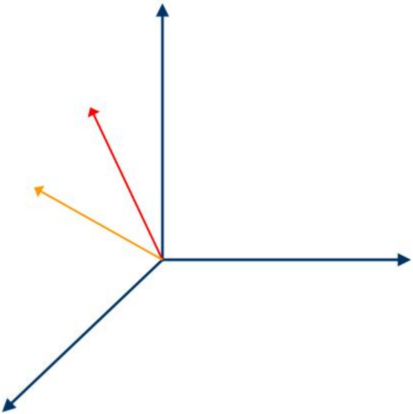
EVD per Frame - Graphical Interpretation

Frame 1, strong eigenvectors



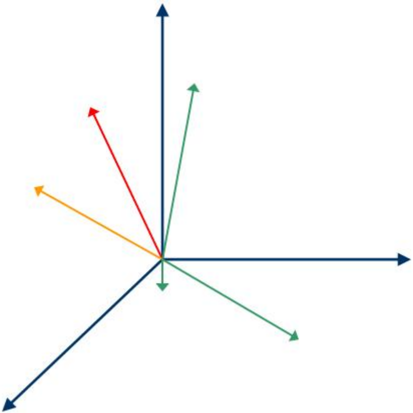
EVD per Frame - Graphical Interpretation

Frame 2, strong eigenvectors



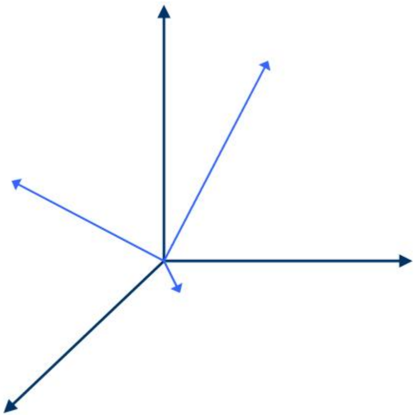
EVD per Frame - Graphical Interpretation

Frame 3, strong eigenvectors



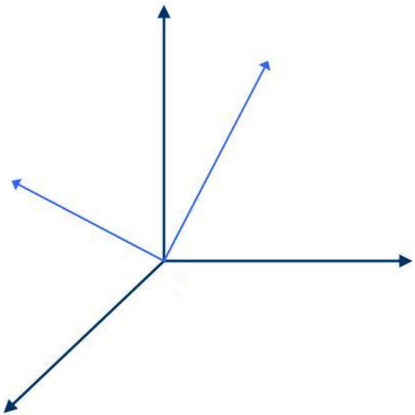
QRD Calculation

Graphical Interpretation



QRD Pruning

Graphical Interpretation



Desired Sources RTF Estimation

One Concurrent Desired Speaker

PSD Estimation

- Stationary noise PSD:

$$\mathbf{\Phi}_{zz}^{\mathrm{stat}} = \mathbf{H}^n \mathbf{\Phi}_{s^n s^n} \left(\mathbf{H}^n\right)^H + \mathbf{\Phi}_{nn}$$

- One desired source (i_0), no non-stationary source:

$$\hat{\mathbf{\Phi}}_{zz}^{d,i_0} \approx \phi_{i_0}^d \mathbf{h}_{i_0}^d \left(\mathbf{h}_{i_0}^d\right)^H + \mathbf{\Phi}_{zz}^{\mathrm{stat}}$$

Largest Generalized Eigenvector

$$\hat{\mathbf{\Phi}}_{zz}^{d,i_0} \mathbf{f}_{i_0} = \lambda_{i_0} \mathbf{\Phi}_{zz}^{\mathrm{stat}} \mathbf{f}_{i_0} \Rightarrow \hat{\tilde{\mathbf{h}}}_{i_0}^d \triangleq \frac{\mathbf{\Phi}_{zz}^{\mathrm{stat}} \mathbf{f}_{i_0}}{\left(\mathbf{\Phi}_{zz}^{\mathrm{stat}} \mathbf{f}_{i_0}\right)_0}$$

Multichannel Post-filtering (for single desired source)

Using matrix inversion lemma [Simmer et al., 2001]; [Doclo et al., 2010]

Why Postfiltering?

- In diffuse noise field multichannel processing is not enough!
- For nonstationary signals advanced single microphone spectral enhancement methods are beneficial [Cohen and Gannot, 2008].

MWF for estimating speech component at reference microphone (#0)

$$\mathbf{w}_{\text{SDW-MWF}} = \frac{\phi_{s^d s^d} \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d}{\mu + \phi_{s_d s_d} (\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d} (h_0^d)^* = \underbrace{\frac{\mathbf{\Phi}_{nn}^{-1} \tilde{\mathbf{h}}^d}{(\tilde{\mathbf{h}}^d)^H \mathbf{\Phi}_{nn}^{-1} \tilde{\mathbf{h}}^d}}_{\text{MVDR}} \times \underbrace{\frac{\phi_{y_s y_s}}{\phi_{y_s y_s} + \mu \phi_{y_n y_n}}}_{\text{SDW-SWF}}$$

where, $\phi_{y_s y_s} = |h_0^d|^2 \phi_{s^d s^d}$ is the desired speech component at the MVDR output and $\phi_{y_n y_n}$ is the respective noise output.

Zelinski Postfilter [Zelinski, 1988]

Assumptions

- Distortionless beamformer $\phi_{y_s y_s} = \phi_{s^d s^d}$.
- Spatially white noise field, $\mathbf{\Phi}_{nn} = \phi_{nn}\mathbf{I}$ (no other interference sources).
- Hence, $\phi_{z_i z_j} = \phi_{s^d s^d}$; $i \neq j$ & $\phi_{z_i z_i} = \phi_{s^d s^d} + \phi_{nn}$.

Estimated Wiener Postfilter

- Recursive estimation of the auto- and cross-spectra: $\hat{\phi}_{z_i z_j}(\ell) = \alpha \hat{\phi}_{z_i z_j}(\ell - 1) + (1 - \alpha) z_i(\ell) z_j^*(\ell)$.
- Zelinski's postfilter:

$$w_{\mathrm{Zel}}(\ell, k) = \frac{\frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} \Re(\hat{\phi}_{z_i z_j}(\ell, k))}{\frac{1}{M} \sum_{i=0}^{M-1} \hat{\phi}_{z_i z_i}(\ell, k)}$$

- Combined with Spectral Subtraction [Meyer and Simmer, 1997].
- Further developed and analyzed [Marro et al., 1998].

McCowan & Bourlard Postfilter [McCowan and Bourlard, 2003]

Further Assumptions

- Noise field with known and isotropic coherence function, $\phi_{n_i n_j} = \phi_{nn}\Gamma_{n_i n_j}$ (no other interference sources).
- Hence, $\phi_{z_i z_j} = \phi_{s^d s^d} + \phi_{nn}\Gamma_{n_i n_j}$; $i \neq j$ & $\phi_{z_i z_i} = \phi_{z_j z_j} = \phi_{s^d s^d} + \phi_{nn}$.
- Diffuse noise field is usually assumed $\Gamma_{n_i n_j}(\omega) = \mathrm{Sinc}(\frac{\omega d_{ij}}{c})$.

Estimated Wiener Postfilter

- McCowan & Bourlard postfilter:

$$\hat{\phi}_{s_i^d s_j^d}(\ell, k) = \frac{\Re(\hat{\phi}_{z_i z_j}) - 0.5\Re(\Gamma_{n_i n_j})(\hat{\phi}_{z_i z_i} + \hat{\phi}_{z_j z_j})}{1 - \Re(\Gamma_{n_i n_j})}$$

$$w_{\mathrm{MB}}(\ell, k) = \frac{\frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} \hat{\phi}_{s_i^d s_j^d}}{\frac{1}{M} \sum_{i=0}^{M-1} \hat{\phi}_{z_i z_i}} \triangleq \frac{\hat{\phi}_{s^d s^d}}{\frac{1}{M} \sum_{i=0}^{M-1} \hat{\phi}_{z_i z_i}}$$

Improved Noise PSD Estimation

Noise Over-estimation

Both postfilters [Zelinski, 1988] and [McCowan and Bourlard, 2003] use over-estimated noise PSD, since they use the input signals rather than the beamformer output.

Noise PSD at beamformer output [Leukimmiatis et al., 2006]

Replace the denominator by:

$$\hat{\phi}_{n_i n_j}(\ell, k) = \frac{0.5(\hat{\phi}_{z_i z_i} + \hat{\phi}_{z_j z_j}) - \Re(\hat{\phi}_{z_i z_j})}{1 - \Re(\Gamma_{n_i n_j})}$$

$$\hat{\phi}_{nn}(\ell, k) = \frac{2}{M(M-1)} \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} \hat{\phi}_{n_i n_j}(\ell, k)$$

$$w_{\mathrm{Leuk}}(\ell, k) = \frac{\hat{\phi}_{s^d s^d}}{\hat{\phi}_{s^d s^d} + \hat{\phi}_{nn} \mathbf{w}_{\mathrm{MVDR}}^H \mathbf{\Gamma}_{nn} \mathbf{w}_{\mathrm{MVDR}}}$$

Nonlinear Postfilter [Balan and Rosca, 2002] I

Motivation

- Nonlinear processing has many advantages in speech enhancement.
- A plethora of nonlinear algorithms for single microphone speech enhancement exist.
- An extension to the multichannel case can derived.

Sufficient Statistics

- Conditional p.d.f.:

$$P_r(\mathbf{z}|s^d; \phi_{s^d s^d}, \mathbf{\Phi}_{nn}, \mathbf{h}^d) = \frac{1}{\pi \mathbf{\Phi}_{nn}} \exp\{-(\mathbf{z} - \mathbf{h}^d s^d)^H \mathbf{\Phi}_{nn}^{-1} (\mathbf{z} - \mathbf{h}^d s^d)\}$$

- MVDR output is sufficient statistics for s_d: $T(\mathbf{z}) = \frac{(\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{z}}{(\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d}$
- $P_r(\rho(s^d)|\mathbf{z}) = P_r(\rho(s^d)|T(\mathbf{z}))$

Nonlinear Postfilter [Balan and Rosca, 2002] II

Log Spectral Amplitude Estimator extending [Ephraim and Malah, 1985]

- Beamformer output: $y = s + \frac{(\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1}}{(\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d} \mathbf{n}$.
- LSA criterion:

$$|\hat{s}^d| = \exp\{E\{\log(|s^d|)|\mathbf{z}\}\} = \exp\{E\{\log(|s^d|)|\,T(\mathbf{z})\}\}$$

- Estimator:

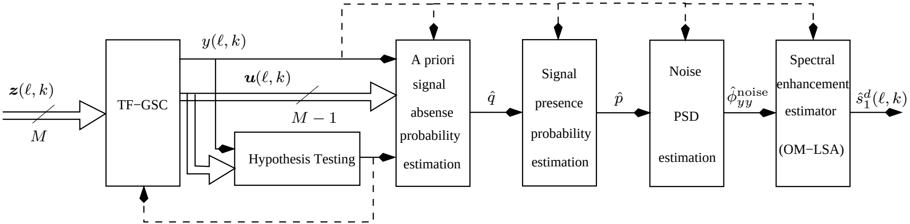
$$|\hat{s}^d| = \frac{\xi}{1+\xi} \exp\left\{\frac{1}{2}\int_v^\infty \frac{e^{-t}}{t} dt\right\} |y|$$

where $\xi \triangleq \phi_{s^d s^d} (\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d$ is the a priori SNR,
$\gamma \triangleq |y|^2 (\mathbf{h}^d)^H \mathbf{\Phi}_{nn}^{-1} \mathbf{h}^d$ is the a posteriori SNR and $v = \frac{\xi\gamma}{1+\xi}$.

- Final estimator is obtained by $\hat{s}^d = |\hat{s}^d| e^{\angle(y)}$.
- Gives motivation to the algorithm presented next.

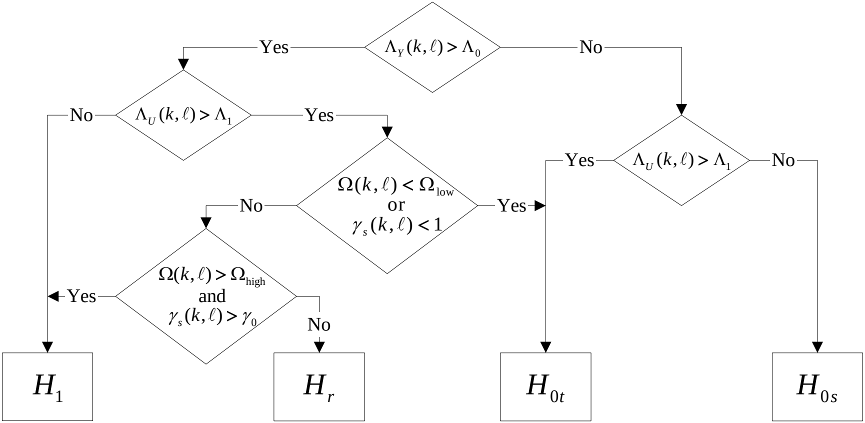
GSC & Speech Presence Probability based Postfiltering

[Cohen et al., 2003]; [Gannot and Cohen, 2004]



- Use main output and reference noise signals to update the speech presence probability.
- Feed backward the decision to update GSC parameters.
- Use the speech presence probability to update the OM-LSA [Cohen and Berdugo, 2001] algorithm for residual noise reduction.

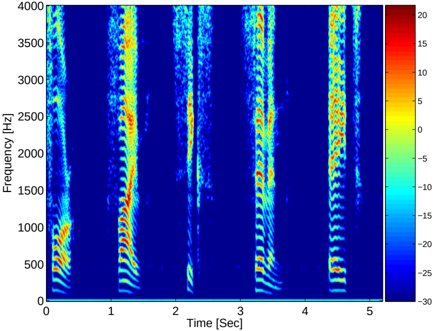
Hypothesis Test



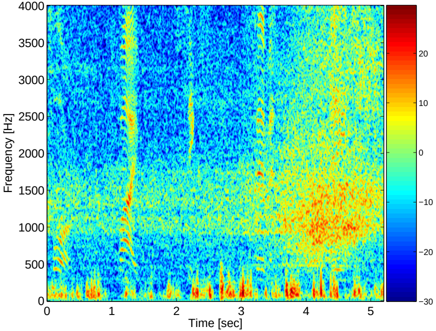
- Λ_Y - local non-stationarity at beamformer output.
- Λ_U - local non-stationarity at noise reference signals.
- Ω - The transient beam-to-reference ratio (TBRR).
- γ_s - a posteriori SNR at the beamformer output.

Experimental Study I

Car Scenario



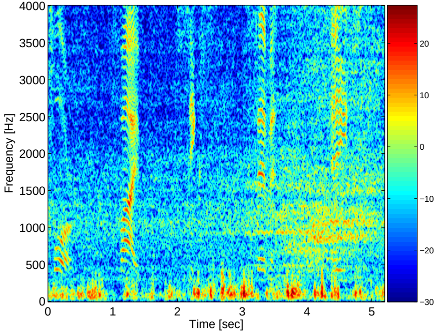
(a) Clean speech



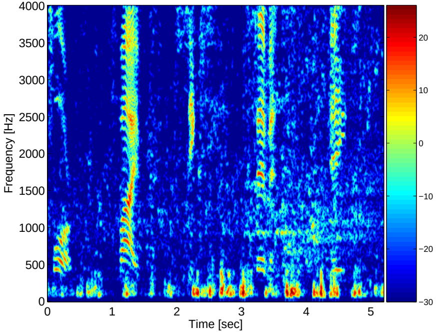
(b) Noisy at mic. #1

Experimental Study II

Car Scenario



(c) TF-GSC output



(d) Multichannel postfiltering

Figure: Speech utterance: "Dial: One, Two, Three, Four, Five, Six, Seven, Eight". Car with open windows equipped with 4 microphones.

The Convolutive TF-GSC [Talmon et al., 2009a] I

Motivation

The GSC [Griffiths and Jim, 1982]

Implemented in time-domain and assumes delay-only propagation. Hence speech distortion is expected.

The TF-GSC [Gannot et al., 2001]

- The RTFs are incorporated into the GSC beamformer.
- Adaptation to reverberant environments obtained by time-frequency implementation.
- For high T_{60}:
  - The RIRs and the respective relative RIRs become very long.
  - Multiplication in frequency-domain (MTF approximation) is only valid if the time frames are significantly larger than the relative RIR.
  - In practice, short frames are used, resulting in inaccurate representation of the RTF and hence performance degradation.

The Convolutive TF-GSC [Talmon et al., 2009a] II

Motivation

Time-Domain MVDR [Chen et al., 2008]

- Full relative RIR is taken into account.
- Theoretically, optimal MVDR in reverberant environment.
- The full-length RTF estimation requires:
  - Very long observations, limiting the ability to work in dynamic environments and to track time-variations.
  - Large computational complexity.
- In practice, the speech source RIRs are modelled as shorter filters.

STFT Implementation [Talmon et al., 2009a] Enables:

- Short frames.
- Long relative RIRs.

CTF-GSC

Objectives

In the STFT Domain:

- Formulate the problem using system representation in the STFT domain [Avargel and Cohen, 2007].
- Build a GSC scheme (a TF-GSC extension).
- Suggest practical solutions using approximations. Specifically, show solutions under the MTF and CTF approximations.
- Incorporate the RTF identification based on the CTF model [Talmon et al., 2009b] and compare experimental results with the TF-GSC.
- Currently, applicable only to single desired source.

Signal Model I

Time Domain

$$z_m(t) = s^d(t) * h_m^d(t) + n_m(t) = \tilde{s}^d(t) * \tilde{h}_m^d(t) + n_m(t)$$

- $\tilde{s}^d(t) = s^d(t) * h_1^d(t)$ - Desired signal component at microphone #1.
- $\tilde{h}_m^d(t)$ - relative RIR between microphone #1 and microphone $\#m$.

STFT Domain

$$z_m(\ell, k) = \sum_{k'=0}^{N_{\mathrm{FFT}}-1} \sum_{p'} \tilde{h}_m(\ell', k', k)\tilde{s}_d(\ell - \ell', k') + n_m(\ell, k)$$

Concatenating successive signal frames:

$$\mathbf{z}_m(k) = \sum_{k'=0}^{N_{\mathrm{FFT}}-1} \tilde{\mathbf{H}}_m(k', k)\tilde{\mathbf{s}}_d(k') + \mathbf{n}_m(k) \overset{\mathrm{CTF}}{\approx} \tilde{\mathbf{H}}_m(k)\tilde{\mathbf{s}}_d(k) + \mathbf{n}_m(k)$$

Signal Model II

Beamforming in the STFT Domain

$$\hat{\hat{\mathbf{s}}}_d(k) = \sum_{m=1}^{M} \sum_{k'=0}^{N_{\mathrm{FFT}}-1} \mathbf{W}_m^H(k',k)\mathbf{z}_m(k') \overset{\mathrm{CTF}}{\approx} \sum_{m=1}^{M} \mathbf{W}_m^H(k)\mathbf{z}_m(k)$$

MVDR & GSC

- Constrained power minimization (MVDR) can be defined.
- GSC structure exists (# of constraints $<$ # of measurements).
- Similarly to the TF-GSC, $\tilde{\mathbf{H}}_m(k', k)$ can be identified [Talmon et al., 2009b].

Setup

Comparing the proposed method to the TF-GSC:

- Image method ([Allen and Berkley, 1979], implemented by [Habets, 2006]).
- Array of 5 microphones.
- Reverberation time $T_{60} = 0.5$s.
- TF-GSC:
  - Frame length - $N = 512$.
  - RTF length - 500.
  - Noise Canceller length - 450.
- CTF-GSC:
  - In FBF and BM - $N = 512, 50\%$ overlap.
  - In adaptive NC - $N = 512, 75\%$ overlap.
  - RTF length - 5 frames.

Signal Blocking

The signal blocking factor (SBF) is defined by:

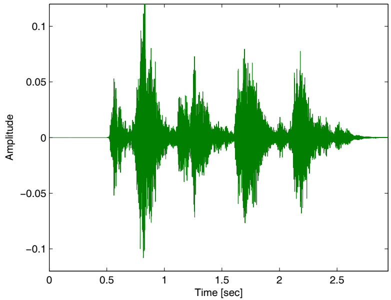
$$\mathbf{SBF} = 10 \log_{10} \frac{E\left\{(\tilde{s}^d(t))^2\right\}}{\mathrm{Mean}_m E\left\{u_m^2(t)\right\}}$$

where $u_m(t)$; $m = 2, \ldots, M$ are the blocking matrix outputs.

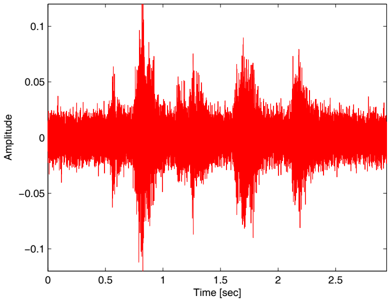
The blocking ability [dB] (known RTF)

	mic. 2	mic. 3	mic. 4	mic. 5
TF	17	12	14	8
CTF	22	17	18	13

Known RTF, Input SNR=0dB I

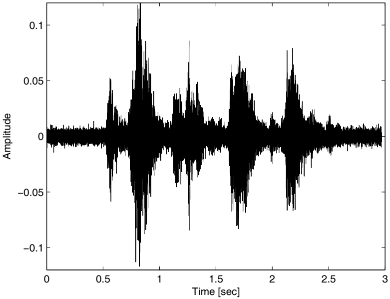


(a) Reverberated speech at microphone #1.

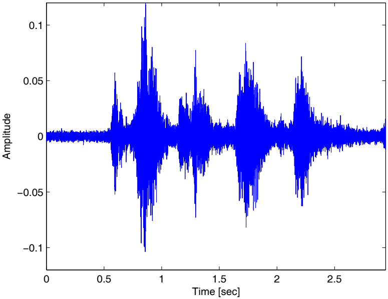


(b) Noisy signal at microphone #1.

Known RTF, Input SNR=0dB II



(c) TF-GSC output.

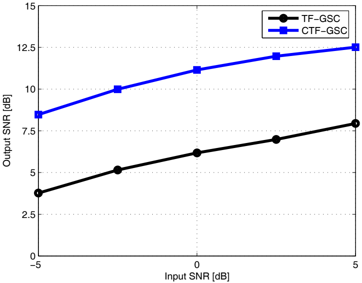


(d) CTF-GSC output.

Summary

Output SNR and Noise Reduction [dB] for known RTF

In SNR	SNR		NR	
	TF-GSC	CTF-GSC	TF-GSC	CTF-GSC
-5	3.8	8.5	-5.9	-10.9
-2.5	5.2	10.0	-6.2	-10.9
0	6.2	11.2	-6.2	-10.9
2.5	7.0	12.0	-6.7	-10.9
5	7.9	12.5	-6.1	-10.9



Estimated RTF

Signal Blocking

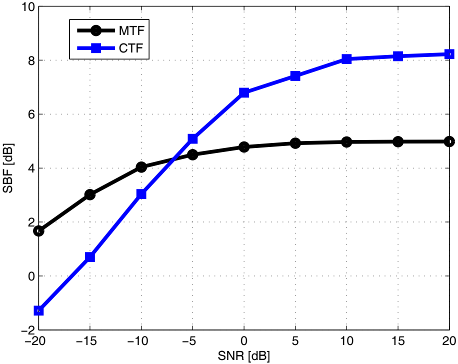
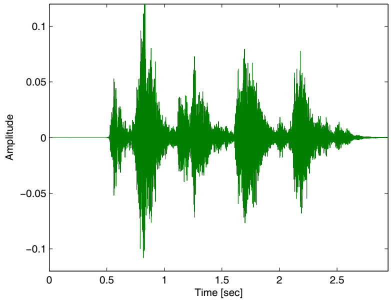
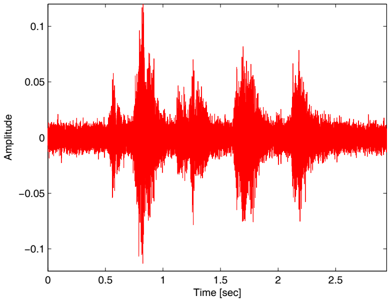


Figure: SBF curves obtained by the RTF identification method based on the MTF and CTF models.

Identified RTF, Input SNR=5dB I

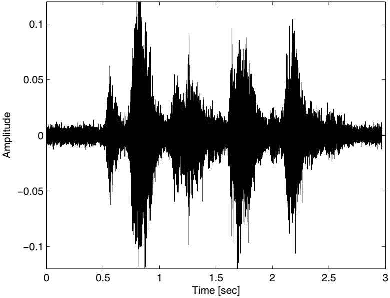


(a) Reverberated speech at microphone #1.

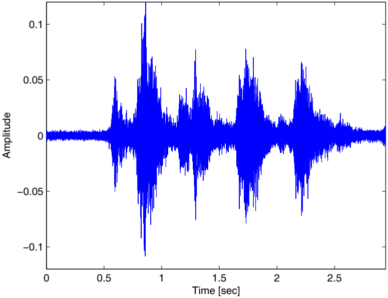


(b) Noisy signal at microphone #1.

Identified RTF, Input SNR=5dB II

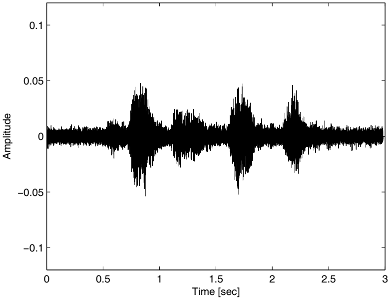


(c) TF-GSC output.

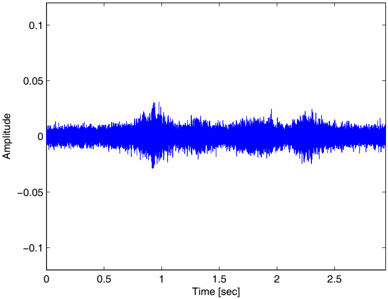


(d) CTF-GSC output.

Identified RTF, Input SNR=5dB III



(e) BM output TF-GSC.

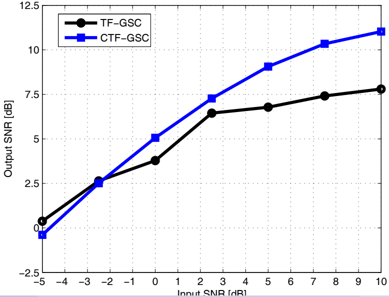


(f) BM output CTF-GSC.

Summary

Output SNR and Noise Reduction [dB] for estimated RTF

In SNR	SNR		NR	
	TF-GSC	CTF-GSC	TF-GSC	CTF-GSC
-5	0.9	-0.4	-2.2	-4.7
-2.5	2.6	2.5	-2.9	-5.3
0	3.8	5.1	-3.2	-6.0
2.5	6.5	7.3	-4.0	-6.8
5	6.8	9.1	-4.3	-7.7
7.5	7.4	10.3	-4.8	-8.4
10	7.8	11.0	-5.5	-9.1



Dynamic Scenario [Markovich-Golan et al., 2010]

Subspace tracking of Multiple Sources

Goal

Extract desired moving speakers from a mixture of speakers using the LCMV beamformer.

Working hypothesis

- Activity indicator for desired speech signals is available.
- Availability of time segments with nonconcurrent desired and interfering speakers.
- "Stable" subspaces represent static speakers with high probability.

Features

- Tracking ability using projection approximation subspace tracking deflation (PASTd) [Yang, 1995].
- Double talk within group allowed during estimation.
- "Expiry time" for outdated basis vectors.

LCMV beamformer

Definitions (Reminder)

$$\mathbf{w} = \mathbf{\Phi}_{zz}^{-1}\mathbf{C}\left(\mathbf{C}^{\dagger}\mathbf{\Phi}_{zz}^{-1}\mathbf{C}\right)^{-1}\mathbf{g}$$

Straightforward Constraints Set

$$\mathbf{C} = \left[\, \mathbf{H}^{d} \;\; \mathbf{H}^{i} \,\right] \qquad \mathbf{g} = \left[\, \mathbf{1}_{1\times P_d} \;\; \mathbf{0}_{1\times P_i} \,\right]^{T}$$

Modified Constraints Set

$$\tilde{\mathbf{C}} = \left[\, \mathbf{Q}^{d} \;\; \mathbf{Q}^{i} \,\right] \qquad \tilde{\mathbf{g}} = \left[\, \left(Q^{d}_{1,1}\right)^{*} \cdots \left(Q^{d}_{P_d,1}\right)^{*} \;\; \mathbf{0}_{1\times P_i} \,\right]^{T}$$

where $\mathbf{Q}^{d}$, $\mathbf{Q}^{i}$ - bases for desired and interfering subspaces.

Output

$$y(\ell, k) = \sum_{j=1}^{P_d} h^{d}_{j,1}(\ell, k) s^{d}_{j}(\ell, k) + \text{residual noise}$$

Tracking Scheme I

Forgetting Factor Consideration

- Tracking $\mathbf{Q}^d$ and $\mathbf{Q}^i$ is a variant of the PASTd algorithm [Yang, 1995] with pre-whitening.
- Forgetting factor β controls the adaptation, with $N_\beta = \frac{1}{1-\beta}$ the algorithm's memory length.
- Standard PASTd suffers from contradicting requirements for β:
  - Fast adaptation $\Rightarrow$ small β.
  - Long memory $\Rightarrow$ large β.
- The contradicting requirements can be mitigated by combined tracking scheme.

Tracking Scheme II

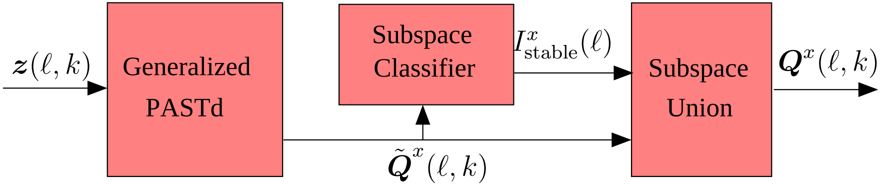
Short & Long Memory

- Use short memory PASTd for fast adaptation of the instantaneous subspace of the xth group of signals, $\tilde{\mathbf{Q}}^x(\ell, k)$.
- Declare stable subspaces, $\mathbf{Q}^x(\ell, k)$, if the basis is valid for more than pre-defined number of frames.
  $I^x_{\mathrm{stable}}(\ell)$ - Indicator for stable subspace of the xth group.
- Subspace union of the valid stable subspaces and the instantaneous subspace using QRD.
- Attribute an expiration time for each stable subspace.

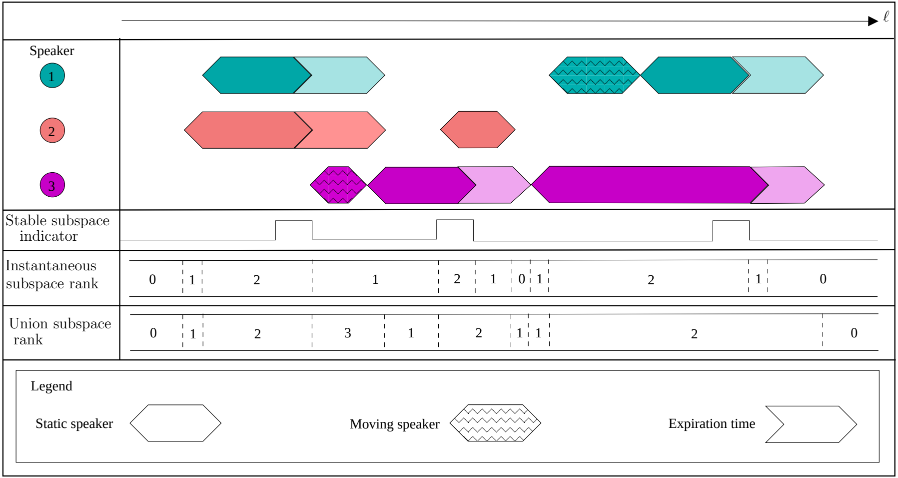
Tracking Scheme III

Classification of Subspace Stability

- The energy of the projected signals onto the instantaneous subspace $\tilde{\mathbf{Q}}^x(\ell, k)$ (integrated over past N_β frames) consists of most of the signals' energy.
- $I^x_{\mathrm{stable}}(\ell) = 1$ if the aggregated energy of the projected signals onto the instantaneous subspace (integrated over past $N_{\mathrm{stable}} \gg N_\beta$ frames) consists of most of the signals' energy.

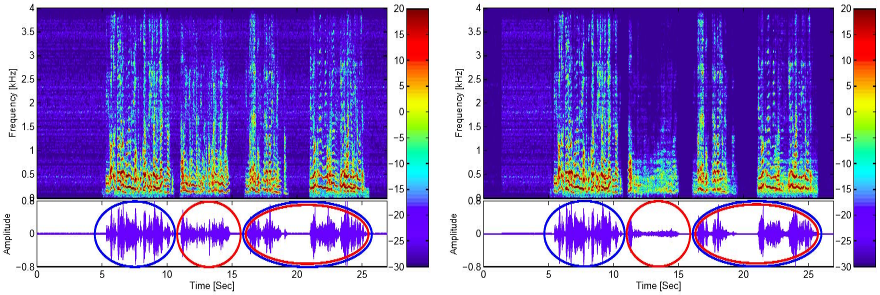


Tracking Example



Experimental Study

Results



(a) Noisy at mic. #1 (b) Enhanced signal

Figure: 2 concurrent desired speakers and 2 competing speakers. 8 microphones recorded at BIU acoustic lab set to $T_{60} = 300$ms.

Binaural LCMV Beamformer

Hadad, Gannot, Doclo, 2012

Motivation

- Duplicate the LCMV beamformer at both ears utilizing all microphones.
- The concept of RTF can be extended and used for preservation of binaural cues (ILD & ITD).
- Efficient implementation by block sharing.

Problem Formulation

Microphone Signals

$$\mathbf{z} = \mathbf{H}^d \mathbf{s}^d + \mathbf{H}^i \mathbf{s}^i + \mathbf{v}$$

Left & Right Reference Microphones

$$z_\ell = \mathbf{e}_\ell^H \mathbf{z}; \quad z_r = \mathbf{e}_r^H \mathbf{z}$$

where

$$\mathbf{e}_\ell = \begin{cases} 1 & m = m_\ell \\ 0 & \text{otherwise} \end{cases} \qquad \mathbf{e}_r = \begin{cases} 1 & m = m_r \\ 0 & \text{otherwise} \end{cases}$$

Binaural Spatial Filters

$$y_\ell = \mathbf{w}_\ell^H \mathbf{z}; \quad y_r = \mathbf{w}_r^H \mathbf{z}.$$

Double LCMV Criterion

Two BFs Utilizing All Microphones

$$\mathbf{w}_\ell = \mathrm{LCMV}(\mathbf{z}; \mathbf{C}, \mathbf{g}_\ell); \quad \mathbf{w}_r = \mathrm{LCMV}(\mathbf{z}; \mathbf{C}, \mathbf{g}_r)$$

Orthonormal Basis for the ATFs

$$\{\mathbf{H}_d = \mathbf{Q}_d \boldsymbol{\Theta}_d; \quad \mathbf{H}_i = \mathbf{Q}_i \boldsymbol{\Theta}_i\} \Rightarrow \mathbf{C} = \begin{bmatrix} \mathbf{Q}_d & \mathbf{Q}_i \end{bmatrix}$$

Left & Right Response Vectors

Apply dereverberation relaxation utilizing RTFs.

Cue Gain Factors:

Desired response $0 < \eta \approx 1$; Interference response $0 < \mu \ll 1$

Interaural Signal Ratio (ISR)

Input ISR

$$\mathrm{ISR}^{in} = \frac{z_\ell}{z_r} = \frac{\mathbf{e}_\ell^\dagger\left(\mathbf{H}_d\mathbf{s}_d + \mathbf{H}_i\mathbf{s}_i\right)}{\mathbf{e}_r^\dagger\left(\mathbf{H}_d\mathbf{s}_d + \mathbf{H}_i\mathbf{s}_i\right)}.$$

Output ISR (in our implementation)

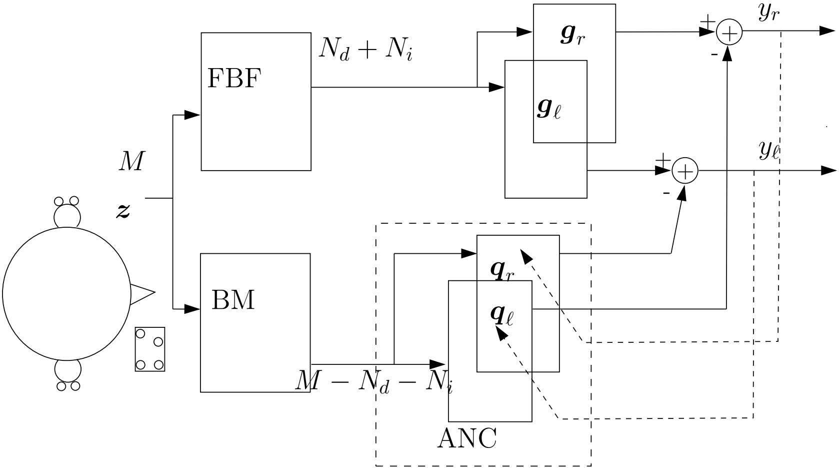
$$\mathrm{ISR}^{out} = \frac{y_\ell}{y_r} = \frac{\mathbf{e}_\ell^\dagger\left(\eta\mathbf{H}_d\mathbf{s}_d + \mu\mathbf{H}_i\mathbf{s}_i\right)}{\mathbf{e}_r^\dagger\left(\eta\mathbf{H}_d\mathbf{s}_d + \mu\mathbf{H}_i\mathbf{s}_i\right)}.$$

ISR vs. ITF

Properties

- Single source case: $\mathrm{ISR}^{out} = \mathrm{ISR}^{in}$ and ISR identifies with the ITF.
- Only one group is active $\Rightarrow$ spatial cues of the group maintained.
- Speech sparsity in STFT domain $\Rightarrow$ cues are preserved also for arbitrary activity pattern.
- Binaural cue preservation is only guaranteed for the constrained sources.
- Unconstrained stationary noise sources and residual (constrained) interference sources will "inherit" the input cues of the dominant source.
- $0 < \mu \ll 1$ will mask the artifacts resulting from leakage.

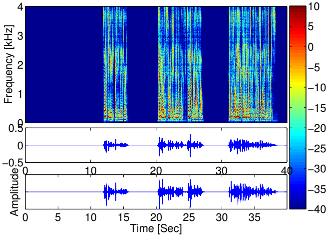
Block Diagram



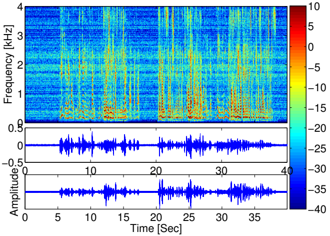
Setup

- Hearing device:
  - 2 hearing aid devices mounted on B&K HATS, with 2 microphones, 2cm inter-distance.
  - A 9×5 utility device with 4 mics. at the corners, average distance 3.5cm. The device placed on a table at a distance of 0.5m.
- Signals:
  - 1 desired speaker, $\theta_d = 30^o$, 1m (constrained).
  - 1 interference speaker at $\theta_i = -70^o$, 1m (constrained).
  - 1 directional stationary noise, $\theta_n = -40^o$, 2.5m (unconstrained).
  - SIR=0dB, SNR=14dB.
- Acoustic lab:
  - Dimensions $6 \times 6 \times 2.4$; Controllable reverb. time $T_{60} = 0.3s$.
- STFT:
  - Sampling frequency 8kHz, 4096 points, 75% overlap.
- Algorithm Cue gain factors:
  - Desired speech - $\eta = 1$.
  - Interference speech - $\mu = 0.1$ (20dB attenuation).

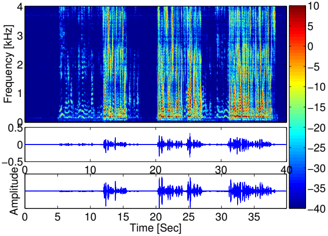
Sonograms



(a) Desired speaker, reference mic.

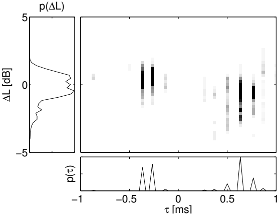


(b) Received reference microphones

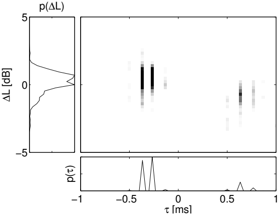


(c) BLCMV outputs

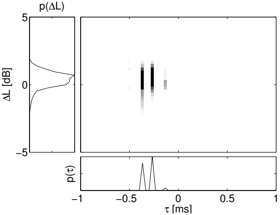
ILD & ITD Preservation (Faller and Merimaa, 2004) I



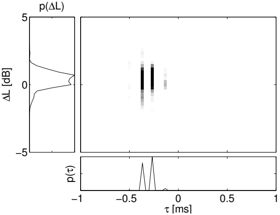
(a) Noisy input



(b) Enhanced output

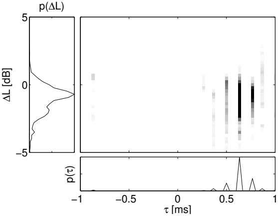


(c) Desired input

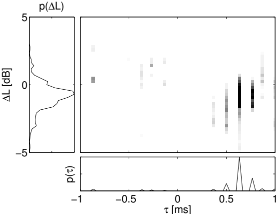


(d) Desired output

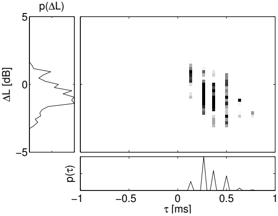
ILD & ITD Preservation (Faller and Merimaa, 2004) II



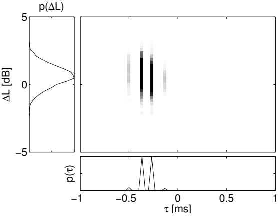
(e) Interference input



(f) Interference output



(g) Stationary input



(h) Stationary output

Audio Samples

Available at:

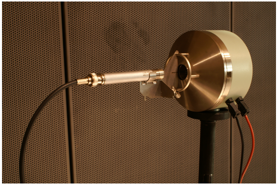
http://www.eng.biu.ac.il/gannot/speech-enhancement/

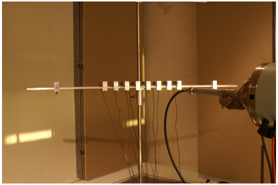
Acoustics Lab, Bar-Ilan University

Features

- Controlled and acoustically isolated environment.
- 60 double-sided panels control the reverberation time.
- Equipped with microphone arrays, loudspeakers, measurement and acquisition equipment.
- Enables fast testing, implementation and verification of algorithms.

BIU Acoustics Lab: Picture Gallery





Thanks to my Collaborators

1. Shmulik Markovich-Golan
2. Prof. Israel Cohen
3. Prof. Ronen Talmon
4. David Levin
5. Prof. Emanuël Habets
6. Elior Hadad
7. Prof. Jacob Benesty
8. and many more...

References and Further Reading I

Affes, S. and Grenier, Y. (1997).
A signal subspace tracking algorithm for microphone array processing of speech.
IEEE transactions on Speech and Audio Processing, 5(5):425–437.

Allen, J. and Berkley, D. (1979).
Image method for efficiently simulating small-room acoustics.
J. Acoustical Society of America, 65(4):943–950.

Avargel, Y. and Cohen, I. (2007).
On multiplicative transfer function approximation in the short-time Fourier transform domain.
IEEE Signal Processing Letters, 14(5):337–340.

Balan, R. and Rosca, J. (2002).
Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase.
In *IEEE Workshop on Sensor Array and Multichannel Signal Processing*, pages 209–213, Rosslyn, Virginia, USA.

Benesty, J., Chen, J., and Huang, Y. (2008a).
Microphone array signal processing.
Springer.

Benesty, J., Huang, Y., and Sondhi, M., editors (2008b).
Springer handbook of speech processing.
Springer Verlag.

Benesty, J., Makino, S., and Chen, J., editors (2005).
Speech Enhancement.
Signals and Communication Technology. Springer, Berlin.

References and Further Reading II

Brandstein, M. S. and Ward, D. B., editors (2001).
Microphone Arrays: Signal Processing Techniques and Applications.
Springer-Verlag, Berlin.

Breed, B. R. and Strauss, J. (2002).
A short proof of the equivalence of lcmv and gsc beamforming.
IEEE Signal Processing Letters, 9(6):168–169.

Buchner, H., Aichner, R., and Kellermann, W. (2004).
TRINICON: A versatile framework for multichannel blind signal processing.
In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages iii–889, Montreal, Canda.

Chen, J., Benesty, J., and Huang, Y. (2008).
A minimum distortion noise reduction algorithm with multiple microphones.
IEEE Transactions on Audio, Speech, and Language Processing, 16(3):481–493.

Cohen, I. (2004).
Relative transfer function identification using speech signals.
IEEE Transactions on Speech and Audio Processing, 12(5):451–459.

Cohen, I., Benesty, J., and Gannot, S., editors (2010).
Speech processing in modern communication: Challenges and perspectives.
Topics in signal processing. Springer.

Cohen, I. and Berdugo, B. (2001).
Speech enhancement for non-stationary noise environments.
Signal processing, 81(11):2403–2418.

References and Further Reading III

Cohen, I. and Gannot, S. (2008).
Springer Handbook of Speech Processing and Speech Communication, chapter Spectral enhancement methods.
In [Benesty et al., 2008b].

Cohen, I., Gannot, S., and Berdugo, B. (2003).
An integrated real-time beamforming and postfiltering system for nonstationary noise environments.
EURASIP Journal on Applied Signal Processing, 2003:1064–1073.

Cox, H., Zeskind, R., and Kooij, T. (1986).
Practical supergain.
Acoustics, Speech and Signal Processing, IEEE Transactions on, 34(3):393–398.

Cox, H., Zeskind, R., and Owen, M. (1987).
Robust adaptive beamforming.
IEEE trans. on Acoustics, Speech and Signal Proc., 35(10):1365–1376.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (2010).
Handbook on Array Processing and Sensor Networks, chapter Acoustic beamforming for hearing aid applications.
Volume 63 of [Haykin and Liu, 2010].

Doclo, S. and Moonen, M. (2002a).
GSVD-based optimal filtering for single and multimicrophone speech enhancement.
IEEE Trans. on Signal Processin, 50(9):2230–2244.

Doclo, S. and Moonen, M. (2002b).
GSVD-based optimal filtering for single and multimicrophone speech enhancement.
IEEE trans. on Speech and Audio Processing, 50(9):2230–2244.

References and Further Reading IV

Doclo, S. and Moonen, M. (2003).
Design of far-field and near-field broadband beamformers using eigenfilters.
Signal Processing, 83(12):2641 – 2673.

Doclo, S., Spriet, A., Wouters, J., and Moonen, M. (2005).
Speech Enhancement, chapter Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction, pages 199–228.
In [Benesty et al., 2005].

Elko, G. W. (1996).
Microphone array systems for hands-free telecommunication.
Speech Communication, 20(3):229–240.

Ephraim, Y. and Malah, D. (1985).
Speech enhancement using a minimum mean-square error log-spectral amplitude estimator.
IEEE Transactions on Acoustics, Speech and Signal Processing, 33(2):443–445.

Er, M. and Cantoni, A. (1983).
Derivative constraints for broad-band element space antenna array processors.
IEEE Transactions on Acoustics, Speech and Signal Processing, 31(6):1378–1393.

Frost III, O. L. (1972).
An algorithm for linearly constrained adaptive array processing.
Proceedings of the IEEE, 60(8):926–935.

Gannot, S., Burshtein, D., and Weinstein, E. (2001).
Signal enhancement using beamforming and nonstationarity with applications to speech.
IEEE Transactions on Signal Processing, 49(8):1614–1626.

References and Further Reading V

Gannot, S. and Cohen, I. (2004).
Speech enhancement based on the general transfer function GSC and postfiltering.
IEEE Transactions on Speech and Audio Processing, 12(6):561–571.

Gannot, S. and Cohen, I. (2008).
Springer Handbook of Speech Processing and Speech Communication, chapter Adaptive beamforming and postfiltering.
In [Benesty et al., 2008b].

Gay, S. L. and Benesty, J., editors (2000).
Acoustic signal processing for telecommunication.
Kluwer Academic.

Gilbert, E. and Morgan, S. (1955).
Optimum design of directive antenna arrays subject to random variations.
Bell Syst. Tech. J, 34(3):637–663.

Griffiths, L. J. and Jim, C. W. (1982).
An alternative approach to linearly constrained adaptive beamforming.
IEEE Trans. on Antennas and Propagation, 30(1):27–34.

Habets, E., Benesty, J., Cohen, I., Gannot, S., and Dmochowski, J. (2010).
New insights into the MVDR beamformer in room acoustics.
IEEE Transactions on Audio, Speech, and Language Processing, 18(1):158–170.

Habets, E., Benesty, J., Gannot, S., Naylor, P., and Cohen, I. (2009).
On the application of the LCMV beamformer to speech enhancement.
In *The IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 141–144, New Paltz, New York, USA.

References and Further Reading VI

Habets, E. A. P. (2006).
Room impulse response (RIR) generator.
http://home.tiscali.nl/ehabets/rir_generator.html.

Haykin, S. and Liu, K. R., editors (2010).
Handbook on array processing and sensor networks, volume 63.
Wiley-IEEE Press.

Herbordt, W. (2005).
Sound capture for human/machine interfaces - Practical aspects of microphone array signal processing, volume 315 of *Lecture Notes in Control and Information Sciences*.
Springer, Heidelberg, Germany.

Hoshuyama, O., Sugiyama, A., and Hirano, A. (1999).
A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters.
IEEE trans. on Signal Proc., 47(10):2677–2684.

Jan, E. and Flanagan, J. (1996).
Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors.
In *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, pages 917–920, Atlanta, Georgia, USA.

Kaneda, Y. and Ohga, J. (1986).
Adaptive microphone-array system for noise reduction.
IEEE Trans. Acoustics, Speech, and Signal Processing, 34(6):1391–1400.

References and Further Reading VII

Leukimmiatis, S., Dimitriadis, D., and Maragos, P. (2006).
An optimum microphone array post-filter for speech applications.
In *Proc. Interspeech-ICSLP*, pages 2142–2145.

Makino, S., Lee, T.-W., and Sawada, H. (2007).
Blind speech separation.
Springer Heidelberg.

Markovich, S., Gannot, S., and Cohen, I. (2008).
A comparison between alternative beamforming strategies for interference cancelation in noisy and reverberant environment.
In *The 25th Convention of IEEE Israel (IEEEI)*, pages 203–207, Eilat, Israel.

Markovich, S., Gannot, S., and Cohen, I. (2009).
Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals.
IEEE Transactions on Audio, Speech, and Language Processing, 17(6):1071–1086.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2010).
Subspace tracking of multiple sources and its application to speakers extraction.
In *The IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 201–204, Dallas, Texas, USA.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2012a).
A sparse blocking matrix for multiple constraints GSC beamformer.
In *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 197–200, Kyoto, Japan.

References and Further Reading VIII

Markovich-Golan, S., Gannot, S., and Cohen, I. (2012b).
A weighted multichannel Wiener filter for multiple sources scenarios.
In *The IEEE 27th Convention of IEEE Israel (IEEEI)*, Eilat, Israel.
best student paper award.

Marro, C., Mahieux, Y., and Simmer, K. (1998).
Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering.
IEEE Trans. on Speech and Audio Proc., 6(3):240–259.

McCowan, I. and Bourlard, H. (2003).
Microphone array post-filter based on noise field coherence.
IEEE Trans. Speech and Audio Process., 11(6):709–716.

Meyer, J. and Simmer, K. U. (1997).
Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction.
In *IEEE Internat. Conf. Acoust. Speech Signal Process (ICASSP)*, pages 21–24, Munich, Germany.

Nordholm, S., Claesson, I., and Bengtsson, B. (1993).
Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars.
IEEE trans. on Vehicular tech., 42(4):514–518.

Parsons, A. T. (1987).
Maximum directivity proof for three-dimensional arrays.
The Journal of the Acoustical Society of America, 82:179.

Reuven, G., Gannot, S., and Cohen, I. (2008).
Dual-source transfer-function generalized sidelobe canceller.
IEEE Transactions on Audio, Speech, and Language Processing, 16(4):711–727.

References and Further Reading IX

Shalvi, O. and Weinstein, E. (1996).
System identification using nonstationary signals.
IEEE Trans. Signal Processing, 44(8):2055–2063.

Shynk, J. (1992).
Frequency-domain and multirate and adaptive filtering.
IEEE Signal Processing Magazine, 9(1):14–37.

Simmer, K. U., Bitzer, J., and Marro, C. (2001).
Post-Filtering Techniques, chapter 3, pages 39–60.
In [Brandstein and Ward, 2001].

Sondhi, M. and Elko, G. (1986).
Adaptive optimization of microphone arrays under a nonlinear constraint.
In *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, volume 11, pages 981–984, Tokyo, Japan.

Spriet, A., Moonen, M., and Wouters, J. (2004).
Spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction.
Signal Processing, 84(12):2367–2387.

Talmon, R., Cohen, I., and Gannot, S. (2009a).
Convolutive transfer function generalized sidelobe canceler.
IEEE Transactions on Audio, Speech, and Language Processing, 17(7):1420–1434.

Talmon, R., Cohen, I., and Gannot, S. (2009b).
Relative transfer function identification using convolutive transfer function approximation.
IEEE Transactions on Audio, Speech, and Language Processing, 17(4):546–555.

References and Further Reading X

Van Compernolle, D. (1990).
Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings.
In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 833–836, Albuquerque, New Mexico, USA. IEEE.

Van Trees, H. L. (2002).
Detection, Estimation, and Modulation Theory, volume IV, Optimum Array Processing.
Wiley, New York.

Van Veen, B. D. and Buckley, K. M. (1988).
Beamforming: A versatile approach to spatial filtering.
IEEE Acoustics, Speech and Signal Proc. magazine, pages 4–24.

Warsitz, E. and Haeb-Umbach, M. (2007).
Blind acoustic beamforming based on generalized eigenvalue decomposition.
IEEE Transactions on Audio, Speech, and Language Processing, 15(5):1529–1539.

Widrow, B., Jr., J. G., McCool, J., Kaunitz, J., Williams, C., Hearn, R., Zeider, J., Jr., E. D., and Goodlin, R. (1975).
Adaptive noise cancelling: Principals and applications.
Proceeding of the IEEE, 63(12):1692–1716.

Yang, B. (1995).
Projection Approximation Subspace Tracking.
IEEE transactions on Signal Processing, 43(1):95–107.

Zelinski, R. (1988).
A microphone array with adaptive post-filtering for noise reduction in reverberant rooms.
In *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, pages 2578–2581, New-York, USA.