Directional Gain Based Noise Covariance Matrix Estimation for MVDR Beamforming

F. Zhang, C. Pan, J. Benesty and J. Chen

ICASSP 2024

April. 26. 2024. (Fri)

Yuseon Choi

SAPL
Speech and Audio Processing Lab.

Contents

1. Introduction

Introduction

- Microphone array beamforming is a widely used technique in practical applications for enhancing the desired speech signal from the noisy multichannel observations.
- Numerous beamforming algorithms have been developed and studied over the past few decades, which can be broadly categorized into fixed and adaptive beamformers.
  - **Fixed beamformer**
    - It is easy to implement, but its performance is in general suboptimal in practical applications.
    - It depends on array topology and pre-specified look direction
  - **Adaptive beamformer**
    - It has better performance, but its performance depends highly on the accuracy of estimated statistics.
    - Inaccurate estimation may lead to great performance degradation and signal self cancellation.
    - Accurate statistics estimation can be very challenging in real acoustic environments, especially with nonstationary noise.

Introduction

- To deal with this issue, the mask-based beamforming approaches were developed.
  - **Mask-based beamformer**
    - The time-frequency masks are used to estimate the second-order statistics for computing the beamformer coefficients.
    - The masks represent the probability or proportion of the desired speech component at the time-frequency bins.
    - The masks can be estimated from the observations by using either neural network (NN) or spatial filtering (SC).
    - ✓ **NN based estimator**
      - It needs a training process prior to application.
      - Its performance may degrade in real acoustic conditions if there exists mismatch between the training and test data.
    - ✓ **SC based estimator**
      - It is basically unsupervised learning scheme.
      - It generally suffers less from the generalization problem.
      - But, it may be less efficient as compared to the NN based methods and may suffer from the permutation problem.

Introduction

- In this paper, a novel time-frequency mask estimator for estimating the noise covariance matrix (NCM) for the minimum variance distortionless response (MVDR) beamformer is presented.
- By utilizing the direction information of the desired source, the proposed mask estimator employs the recent directional gain framework [19], which was developed as a postfilter to suppress both interference and ambient noise.
- Since it does not need a training stage, the proposed mask estimator does not suffer from the generalization problem.
- Furthermore, since it can be written in a closed form, the developed method does not require an iteration computation process as in the SC-based estimators.
- The experimental results in noise-plus-interference environments show that the MVDR beamformer powered by the developed NCM estimation yields better speech enhancement performance than two commonly used beamformers.

2. Signal Model and Problem Formulation

Signal Model and Problem Formulation

- Assumption
  - M-element uniform linear microphone array (ULMA) to capture the far-field speech signal in some noise field
  - Ignore the multipath effect in signal model
  - Interelement spacing: δ
  - Azimuth angle that the signal arrives: θ_{d}

- The array observation vector can be expressed in the short-time Fourier transform (STFT) domain as

$$\mathbf{y}(k,t) = \mathbf{x}(k,t) + \mathbf{v}(k,t) = X_1(k,t)\mathbf{d}_{\theta_{\mathrm{d}}}(k) + \mathbf{v}(k,t)$$

$$\mathbf{d}_{\theta_{\mathrm{d}}}(k) = \left[1 \; e^{-\jmath\omega_k\tau_0\cos\theta_{\mathrm{d}}} \cdots e^{-\jmath(M-1)\omega_k\tau_0\cos\theta_{\mathrm{d}}}\right]^T$$

k: the frequency-bin index
t: the time-frame index, which satisfies $1 \le t \le T$, with T being the number of frames
f_k: the temporal frequency at the kth frequency bin

$\omega_k = 2\pi f_k$
$\tau_0 = \delta/c$ with $c = 340$ m/s

Signal Model and Problem Formulation

- Assuming that $\mathbf{x}(k,t)$ and $\mathbf{v}(k,t)$ are zero-mean random vectors and uncorrelated,
  - the covariance matrix of $\mathbf{y}(k,t)$ as $\mathbf{\Phi_y}(k,t) = \phi_X(k,t)\, \mathbf{d}_{\theta_\mathrm{d}}(k)\, \mathbf{d}_{\theta_\mathrm{d}}^H(k) + \mathbf{\Phi_v}(k,t)$
    - $\phi_X(k,t)$: the variance of the desired signal $X_1(k,t)$
    - $\mathbf{\Phi_v}(k,t)$: the covariance matrix of $\mathbf{v}(k,t)$, which is assumed to be full rank
  - Authors attempt to estimate the desired signal $X_1(k,t)$ from $\mathbf{y}(k,t)$ with a beamforming techniques

Signal Model and Problem Formulation

□ MVDR beamformer

– It is designed to minimize the variance of the residual noise without distorting the signal arriving from the desired look direction.

- Criterion: $\mathbf{h}_{\mathrm{MVDR}} = \underset{\mathbf{h}}{\operatorname{argmin}}\left\{\mathbf{h}^H \boldsymbol{\Phi}_{\mathbf{v}}(k,t)\mathbf{h} \quad \mathrm{s.t.} \quad \mathbf{d}_{\theta_{\mathrm{d}}}^H(k)\mathbf{h} = 1\right\}$

- Solution: $\mathbf{h}_{\mathrm{MVDR}}(k,t) = \frac{\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(k,t)\mathbf{d}_{\theta_{\mathrm{d}}}(k)}{\mathbf{d}_{\theta_{\mathrm{d}}}^H(k)\boldsymbol{\Phi}_{\mathbf{v}}^{-1}(k,t)\mathbf{d}_{\theta_{\mathrm{d}}}(k)}$

- Beamforming output: $\hat{X}_1(k,t) = \mathbf{h}_{\mathrm{MVDR}}^H(k,t)\mathbf{y}(k,t)$

– Authors assume that the desired signal direction θ_{d} is known or has been estimated in advance.

– To implement the MVDR beamformer, one only needs to estimate the noise covariance matrix (NCM) $\boldsymbol{\Phi}_{\mathbf{v}}(k,t)$.

3. A Framework of Directional Gain Beamforming

A Framework of Directional Gain Beamforming

– The beamforming output in the STFT domain is obtained by $Z(\omega,t) = \mathbf{h}^H(\omega,t)\mathbf{y}(\omega,t)$

$\mathbf{h}(\omega,t) \triangleq [H_0(\omega,t)\ \ H_1(\omega,t) \cdots H_{M-1}(\omega,t)]^T$: The beamforming filter of length M consisting of complex weighting coefficients

– Three important measures are often used to evaluate the performance of a beamformer.

- **Beampattern**

$$\mathcal{B}(\omega,\varphi) \triangleq \mathbf{h}^H(\omega,t)\mathbf{d}(\omega,\varphi)$$

where $\varphi \triangleq [\theta\ \phi]^T$ with $\theta \in [0,\pi]$ and $\phi \in [0,2\pi]$

- **White-Noise-Gain (WNG)**

$$\mathrm{WNG}(\omega) \triangleq \frac{|\mathbf{h}^H(\omega,t)\mathbf{d}(\omega,\varphi_d)|^2}{\mathbf{h}^H(\omega,t)\mathbf{h}(\omega,t)} = \frac{1}{|\mathbf{h}^H(\omega,t)\mathbf{h}(\omega,t)|}$$

$$\mathcal{S} = \frac{1}{\mathrm{WNG}} = \|\mathbf{h}\|^2$$

- **Directivity Factor (DF)**

$$\mathrm{DF}(\omega) \triangleq \frac{|\mathcal{B}(\omega,\varphi_d)|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|\mathcal{B}(\omega,\varphi)|^2 \sin\theta d\theta d\phi}$$

$$= \frac{|\mathbf{h}^H(\omega,t)\mathbf{d}(\omega,\varphi_d)|^2}{\mathbf{h}^H(\omega,t)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}(\omega,t)}$$

$$\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega) \triangleq \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\mathbf{d}(\omega,\varphi)\mathbf{d}^H(\omega,\varphi)\sin\theta d\theta d\phi$$

A Framework of Directional Gain Beamforming

- Fixed beamformers are easy and robust to implement though the performance in terms of noise and interference suppression is generally sub-optimal.
- Adaptive beamformers can achieve better noise and interference suppression, but designing such beamformers that can achieve optimal performance and are robust in practical application is a challenging issue.
- Generally, an adaptive beamformer can be decomposed as the product of a **fixed beamformer** and a **post filter.**

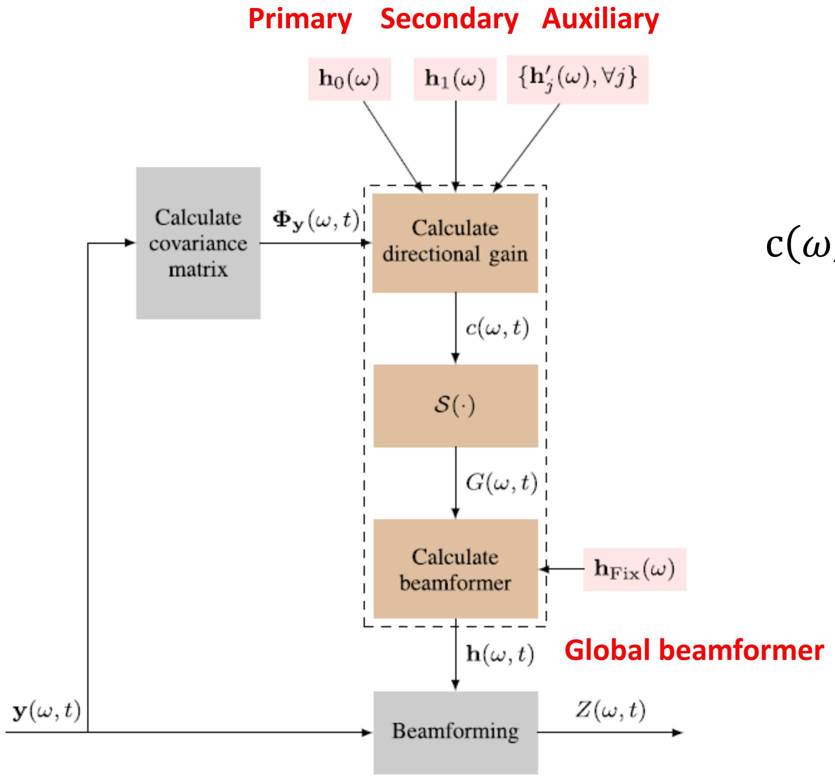
$$\mathbf{h}(\omega, t) = \mathbf{h}_{\mathrm{Fix}}(\omega, t)G(\omega, t)$$

- The problem of adaptive beamforming is transformed into one of designing the post filter $G(\omega, t)$ to achieve the highest possible array gain in terms of noise and interference suppression.
- The post filter is a function of the source incidence angle so the resulting gain is called a **directional gain.**

A Framework of Directional Gain Beamforming

– Authors design the directional gain $G(\omega, t)$ in a multistage manner, which combines sequentially a number of fixed beamformers.

– Assuming that $J + 2$ beamformers are designed, the directional gain $G(\omega, t)$ is computed as



$$G(\omega, t) \triangleq S[c(\omega, t)],$$

$$c(\omega, t) \triangleq \frac{\mathbf{h}_0^H(\omega)\mathbf{\Phi}_\mathbf{y}(\omega, t)\mathbf{h}_1(\omega)}{\sum_{j=0}^{J-1} \mathbf{h}_j'^H(\omega)\mathbf{\Phi}_\mathbf{y}(\omega, t)\mathbf{h}_j'(\omega)}, \quad S_{L,H,a}(x) = \begin{cases} H, & \Re(x) - a \geq H \\ \Re(x) - a, & L \leq \Re(x) - a < H \\ L, & \Re(x) - a < L \end{cases}$$

$\mathbf{\Phi}_\mathbf{y}(\omega, t) = E[\mathbf{y}(\omega, t)\mathbf{y}^H(\omega, t)]$: the covariance matrix of the array observation signal

Since $\boldsymbol{G(\omega, t)}$ is **signal-dependent,**

it is clear that the resulting global beamformer is **adaptive.**

A Framework of Directional Gain Beamforming

□ Principle of the Beamformers

- In the presented multistage approach, each beamformer serves a different purpose.
  - **The primary beamformer** $\mathbf{h}_0(\omega)$
    - Preserves the signal from the desired look direction
    - Shares the same principles as the classical beamformers such as DS, superdirective and differential beamformers
  - **The secondary beamformer** $\mathbf{h}_1(\omega)$
    - Attempts to increase the DF so the directional gain can be used to achieve maximum possible noise and interference suppression
  - **The auxiliary beamformers** $\mathbf{h}_j'(\omega), j = 0, 1, \dots, J-1$
    - Help maintain the robustness and shape of the overall beampattern
    - Designed with the following constraints: $\sum_{j=0}^{J-1} \left| \mathcal{B}_j'(\omega, \varphi) \right|^2 = 1, \forall \varphi$
- Clearly, the beampattern $\mathcal{B}_{\mathrm{DG}}(\omega, \varphi)$ can be made sharper than $\mathcal{B}_0(\omega, \varphi)$.

$$\mathcal{B}_{\mathrm{DG}}(\omega, \varphi) \triangleq \mathrm{c}(\omega, t) \Big|_{\text{point source}}$$

$$= \frac{\mathcal{B}_0(\omega, \varphi) \mathcal{B}_1^*(\omega, \varphi)}{\sum_{j=0}^{J-1} \left| \mathcal{B}_j'(\omega, \varphi) \right|^2}$$

$$\mathcal{B}_{\mathrm{DG}}(\omega, \varphi) = \mathcal{B}_0(\omega, \varphi) \mathcal{B}_1^*(\omega, \varphi)$$

A Framework of Directional Gain Beamforming

□ Design of the Directional Gain

- **The primary beamformer $\mathbf{h}_0(\omega)$**
  - Authors take the WNG constrained superdirective beamformer as the primary beamformer.

$$\min_{\mathbf{h}_0(\omega)} \mathbf{h}_0^H(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_0(\omega) \text{ s.t. } \mathbf{h}_0^H(\omega)\mathbf{d}(\omega,\varphi_\mathrm{d}) = 1,\ \mathbf{h}_0^H(\omega)\mathbf{h}_0(\omega) \le \rho_{\mathrm{WNG}},$$

where φ_d is the incidence angle of the desired source, and $\rho_{\mathrm{WNG}} > 0$ is applied to control the WNG

- **The secondary beamformer $\mathbf{h}_1(\omega)$**
  - It is designed to make the global beampattern as sharp as possible.
  - The cost function of $\mathbf{h}_1(\omega)$ can be expressed as: $\mathcal{J}(\omega) \triangleq \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|\mathcal{B}_0(\omega,\varphi)\mathcal{B}_1^*(\omega,\varphi)|^2 \sin\theta d\theta d\phi = \mathbf{h}_1^H(\omega)\boldsymbol{\Psi}(\omega)\mathbf{h}_1(\omega)$

where $\boldsymbol{\Psi}(\omega)$ is a matrix of size $M \times M$, and its (m,n)th element is equal to $[\boldsymbol{\Psi}(\omega)]_{m,n} = \mathbf{h}_0^H(\omega)\boldsymbol{\Upsilon}_{m,n}(\omega)\mathbf{h}_0(\omega)$ with $\left[\boldsymbol{\Upsilon}_{m,n}(\omega)\right]_{i,j} = \sin(\omega/c\left\|\zeta_m - \zeta_n + \zeta_i - \zeta_j\right\|)$

$\boldsymbol{\Psi}(\omega)$ is a non-negative Hermitian matrix, and $\boldsymbol{\Upsilon}_{m,n}(\omega) = \boldsymbol{\Gamma}_{\mathrm{dn}}(\omega), \forall m = 0, 1, \dots, M-1.$

$$\min_{\mathbf{h}_1(\omega)} \mathbf{h}_1^H(\omega)\boldsymbol{\Psi}(\omega)\mathbf{h}_1(\omega) \text{ s.t. } \mathbf{h}_1^H(\omega)\mathbf{d}(\omega,\varphi_\mathrm{d}) = 1,\ \mathbf{h}_1^H(\omega)\mathbf{h}_1(\omega) \le \rho_{\mathrm{WNG}}.$$

A Framework of Directional Gain Beamforming

Design of the Directional Gain

- **The auxiliary beamformers** $\mathbf{h}'_j(\omega), j = 0, 1, 2, \ldots, J-1$
  - In this work, first two auxiliary beamformers are $\mathbf{h}'_0(\omega) \propto \mathbf{h}_0(\omega)$ and $\mathbf{h}'_1(\omega) \propto \mathbf{h}_1(\omega)$.
  - The rest auxiliary beamformers $\mathbf{h}'_j(\omega)$ $(j = 2, 3, \ldots, J-1)$ are applied to capture the signal blocked by $\mathbf{h}_0(\omega)$ and $\mathbf{h}_1(\omega)$; they are designed, iteratively.
  - The desired beampattern of the jth auxiliary beamformer is expressed as:

$$\mathcal{B}'_{\mathrm{d},j}(\omega,\varphi) = \sqrt{\mathcal{B}'_{\max,j}(\omega) - \sum_{i=0}^{j-1} \left|\mathcal{B}'_i(\omega,\varphi)\right|^2}, \text{ where } \mathcal{B}'_{\max,j}(\omega) \triangleq \max_{\varphi} \sum_{i=0}^{j-1} \left|\mathcal{B}'_i(\omega,\varphi)\right|^2$$

  - If the last auxiliary beamformer $\mathbf{h}'_{j-1}(\omega)$ is determined, authors normalize them according to $\mathbf{h}'_j(\omega) \leftarrow \frac{1}{\sqrt{\mathcal{B}'_{\max,j}(\omega)}} \mathbf{h}'_j(\omega)$.
  - This normalization is necessary due to the constraint on the auxiliary beamformers, $\sum_{j=0}^{J-1} \left|\mathcal{B}'_j(\omega,\varphi)\right|^2 = 1, \forall \varphi$.

A Framework of Directional Gain Beamforming

□ Design of the Directional Gain

– **The auxiliary beamformers $\mathbf{h}'_j(\omega), j = 0, 1, 2, \dots, J-1$**

- By discretizing the variable φ, i,e., $\{\varphi_\ell, \forall \ell = 0,1,2, \dots, L-1\}$, one can form the following linear system, i.e.,

$$\mathbf{h}'^H_j(\omega)\mathbf{D}(\omega) = \mathbf{b}^T_{\mathrm{d},j}(\omega),$$

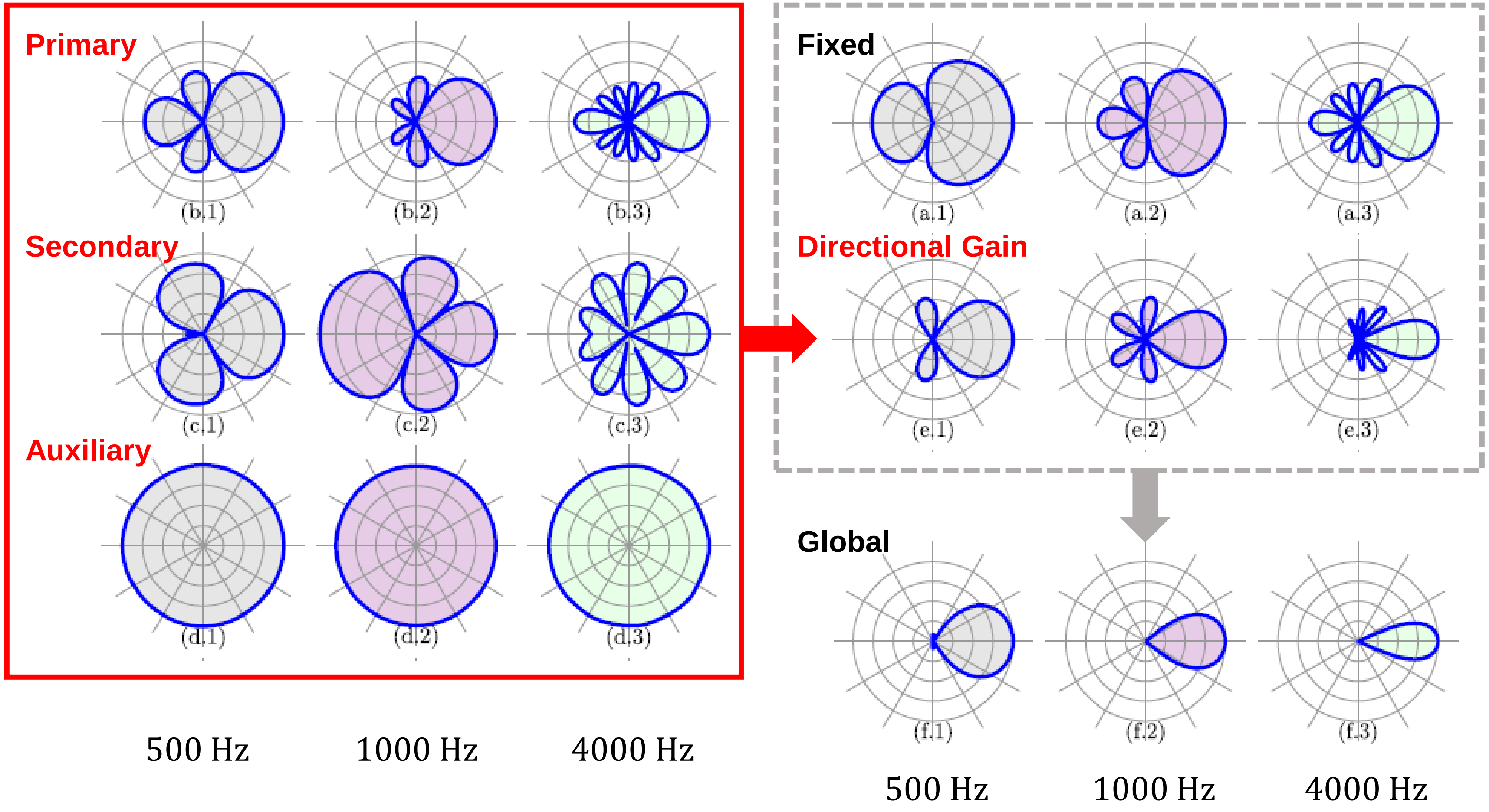
where $\mathbf{D}(\omega) \triangleq [\mathbf{d}(\omega, \varphi_0)\ \mathbf{d}(\omega, \varphi_1)\ \cdots\ \mathbf{d}(\omega, \varphi_{L-1})]$, $\mathbf{b}_{\mathrm{d},j}(\omega) \triangleq \left[\mathcal{B}'_{\mathrm{d},j}(\omega, \varphi_0)\ \mathcal{B}'_{\mathrm{d},j}(\omega, \varphi_1)\ \cdots\ \mathcal{B}'_{\mathrm{d},j}(\omega, \varphi_{L-1})\right]^T$

- To prevent the additive white noise from amplification, $\mathbf{h}'^H_j(\omega)\mathbf{h}'_j(\omega) \le \rho_{\mathrm{WNG}}$.
- Therefore, the optimization problem for the $\mathbf{h}'_j(\omega)$ can finally be expressed as:

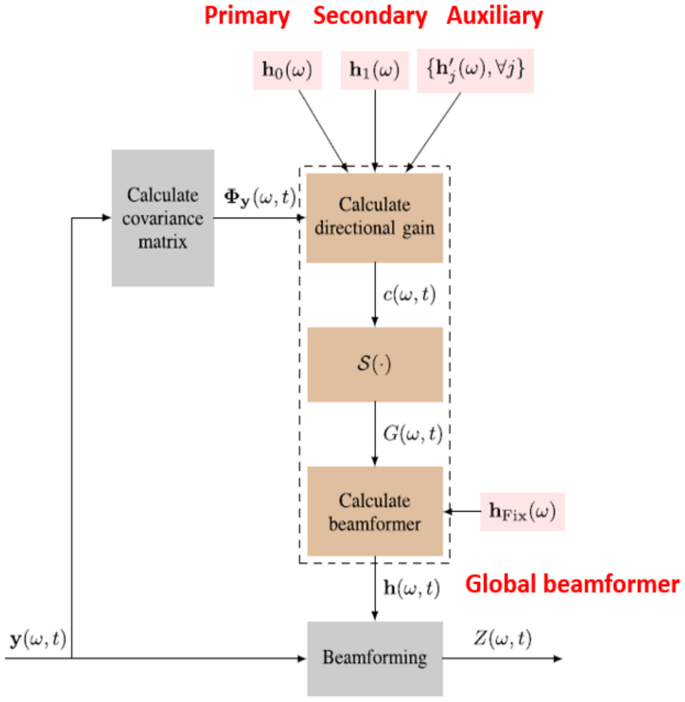
$$\min_{\mathbf{h}'_j(\omega)} \left\|\mathbf{h}'^H_j(\omega)\mathbf{D}(\omega) - \mathbf{b}^T_{\mathrm{d},j}(\omega)\right\|^2 \text{ s.t. } \mathbf{h}'^H_j(\omega)\mathbf{h}'_j(\omega) \le \rho_{\mathrm{WNG}}.$$

A Framework of Directional Gain Beamforming

□ **Beampatterns of the proposed approach**



$$G(\omega,t) \triangleq S\left[\frac{\mathbf{h}_0^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega,t)\mathbf{h}_1(\omega)}{\sum_{j=0}^{J-1}\mathbf{h}_j'^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega,t)\mathbf{h}_j'(\omega)}\right]$$



A Framework of Directional Gain Beamforming

□ Insights of the Directional Gain

– To understand the meaning of the directional gain, consider the general noise environment.

$$\boldsymbol{\Phi}_{\mathbf{y}}(\omega,t) = \underbrace{\phi_{\mathrm{ss}}(\omega,t)\mathbf{d}(\omega,\varphi_{\mathrm{ss}})\mathbf{d}^H(\omega,\varphi_{\mathrm{ss}})}_{\text{The desired source}} + \underbrace{\sum_{i=0}^{I-1}\phi_{\mathrm{in},i}(\omega,t)\mathbf{d}(\omega,\varphi_{\mathrm{in},i})\mathbf{d}^H(\omega,\varphi_{\mathrm{in},i})}_{\text{Multiple interference sources}} + \underbrace{\phi_{\mathrm{dn}}(\omega,t)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)}_{\text{Diffuse noise}} + \underbrace{\phi_{\mathrm{wn}}(\omega,t)\mathbf{I}}_{\text{White noise}}$$

- φ_{ss}: the direction of the source / $\varphi_{\mathrm{in},i}$: the direction of the ith interference source

By considering that…

- $\sum_{j=0}^{J-1}\left|\mathcal{B}_j'(\omega,\varphi)\right|^2 = 1, \forall\varphi$
- $\sum_{j=0}^{J-1}\mathbf{h}_j'^H(\omega)\,\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_j'(\omega) = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\left[\sum_{j=0}^{J-1}\left|\mathcal{B}_j'(\omega,\varphi)\right|^2\right]\sin\theta d\theta d\phi = 1$
- $C_{\mathbf{h}} \triangleq \sum_{j=0}^{J-1}\mathbf{h}_j'^H\mathbf{h}_j'$

The denominator of $c(\omega,t)$

$$\sum_{j=0}^{J-1}\mathbf{h}_j'^H(\omega)\,\boldsymbol{\Phi}_{\mathbf{y}}(\omega)\mathbf{h}_j'(\omega) = \phi_{\mathrm{ss}}(\omega,t) + \sum_{i=0}^{I-1}\phi_{\mathrm{in},i}(\omega,t) + \phi_{\mathrm{dn}}(\omega,t) + C_{\mathbf{h}}\phi_{\mathrm{wn}}(\omega,t)$$

*** Directional gain**

$$\mathrm{c}(\omega,t) \triangleq \frac{\mathbf{h}_0^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega,t)\mathbf{h}_1(\omega)}{\sum_{j=0}^{J-1}\mathbf{h}_j'^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega,t)\mathbf{h}_j'(\omega)}$$

A Framework of Directional Gain Beamforming

□ Insights of the Directional Gain

By considering that…

- $\mathbf{h}_1^H \mathbf{d}(\omega, \varphi_{\mathrm{ss}}) = 1$ / $\mathbf{h}_0^H \mathbf{d}(\omega, \varphi_{\mathrm{ss}}) = 1$ **(distortionless constraint)**

The numerator of $c(\omega, t)$

$$\mathbf{h}_0^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega, t)\mathbf{h}_1(\omega) = \phi_{\mathrm{ss}}(\omega, t) + \phi_{\mathrm{ridw}}(\omega, t)$$

*** Directional gain**

$$\mathrm{c}(\omega, t) \triangleq \frac{\mathbf{h}_0^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega, t)\mathbf{h}_1(\omega)}{\sum_{j=0}^{J-1} \mathbf{h}_j'^H(\omega)\boldsymbol{\Phi}_{\mathbf{y}}(\omega, t)\mathbf{h}_j'(\omega)}$$

The total variance of the residual interference, diffuse noise, and spatially white noise is defined as

$$\phi_{\mathrm{ridw}}(\omega, t) = \sum_{i=0}^{I-1} \phi_{\mathrm{in},i}(\omega, t)\alpha_{\mathrm{in},i}(\omega) + \phi_{\mathrm{dn}}(\omega, t)\alpha_{\mathrm{dn}}(\omega) + \phi_{\mathrm{wn}}(\omega, t)\alpha_{\mathrm{wn}}(\omega)$$

- $\alpha_{\mathrm{in},i}(\omega) \triangleq \mathcal{B}_{\mathrm{DG}}\left(\omega, \varphi_{\mathrm{in},i}\right)$
- $\alpha_{\mathrm{dn}}(\omega) \triangleq \mathbf{h}_0^H(\omega, t)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_1(\omega)$
- $\alpha_{\mathrm{wn}}(\omega) \triangleq \mathbf{h}_0^H(\omega)\mathbf{h}_1(\omega)$

A Framework of Directional Gain Beamforming

□ Insights of the Directional Gain

$$\phi_{\mathrm{ridw}}(\omega,t) = \sum_{i=0}^{I-1} \phi_{\mathrm{in},i}(\omega,t)\alpha_{\mathrm{in},i}(\omega) + \phi_{\mathrm{dn}}(\omega,t)\alpha_{\mathrm{dn}}(\omega) + \phi_{\mathrm{wn}}(\omega,t)\alpha_{\mathrm{wn}}(\omega)$$

- $\alpha_{\mathrm{in},i}(\omega) \triangleq \mathcal{B}_{\mathrm{DG}}\left(\omega, \varphi_{\mathrm{in},i}\right)$
  - ✓ As long as the interference sources are not close to the source, one can design the beamformer $\mathbf{h}_1(\omega)$ such that $\left|\mathcal{B}_{\mathrm{DG}}\left(\omega, \varphi_{\mathrm{in},i}\right)\right| \ll 1$.
  - ✓ If some interference sources are close to the desired source signal, they will be treated as part of the source signal.
- $\alpha_{\mathrm{dn}}(\omega) \triangleq \mathbf{h}_0^H(\omega,t)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_1(\omega)$
  - ✓ Using the fact that $\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega) = \boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2}(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2}(\omega)$, $\boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2H}(\omega) = \boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2}(\omega)$, one can derive that $|\alpha_{\mathrm{dn}}(\omega)| = \left|\mathbf{h}_0^H(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2}(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}^{1/2}(\omega)\mathbf{h}_1(\omega)\right| \leq$ $\sqrt{|\mathbf{h}_0^H(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_0(\omega)|} \times \sqrt{|\mathbf{h}_1^H(\omega)\boldsymbol{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}_1(\omega)|}$.
- $\alpha_{\mathrm{wn}}(\omega) \triangleq \mathbf{h}_0^H(\omega)\mathbf{h}_1(\omega)$
  - ✓ One can derive that

$$|\alpha_{\mathrm{wn}}(\omega)| \leq \sqrt{[\mathbf{h}_0^H(\omega)\mathbf{h}_0(\omega)] \cdot [\mathbf{h}_1^H(\omega)\mathbf{h}_1(\omega)]}.$$

Two terms in the root are often bounded to guarantee the robustness of beamformers against the array self noise.

A Framework of Directional Gain Beamforming

□ Insights of the Directional Gain

– If we treat all the interference sources close to the desired source direction as part of the desired source, the directional gain can be expressed as:

$$c(\omega,t) \approx \frac{\phi_{\mathrm{ss}}(\omega,t)}{\phi_{\mathrm{ss}}(\omega,t) + \sum_{i=0}^{I-1} \phi_{\mathrm{in},i}(\omega,t) + \phi_{\mathrm{dn}}(\omega,t)}$$

– The directional gain $c(\omega,t)$ can be viewed as an approximation of the optimal Wiener gain at the current frequency bin and frame index.

– Unlike the post filter in the traditional adaptive beamformers, the post filter $c(\omega,t)$ is estimated from the covariance matrix $\mathbf{\Phi}_{\mathbf{y}}(\omega,t)$ and the pre-designed fixed beamformers $\mathbf{h}_0(\omega)$, $\mathbf{h}_1(\omega)$, and $\mathbf{h}_j'(\omega)$'s.

– As a result, it is a direction sensitive gain.

4. Noise Covariance Matrix Estimation with Directional Gain

Noise Covariance Matrix Estimation with Directional Gain

□ Mask-Based NCM Estimation

- Estimated time-frequency masks are used to compute the second-order statistics required in the calculation of the beamformer coefficients.
- $\widehat{\boldsymbol{\Phi}}_{\mathbf{v}}(k,t) = \alpha_v \widehat{\boldsymbol{\Phi}}_{\mathbf{v}}(k,t-1) + (1-\alpha_v)\mathcal{M}_0(k,t)\mathbf{y}(k,t)\mathbf{y}^H(k,t)$
  - α_v: smoothing factor, $0 < \alpha_v < 1$
  - $\mathcal{M}_0(k,t)$: the time-frequency mask for the noise

□ An optimal mask

- $\mathcal{M}_0(k,t) = |G_0(k,t)|^2$
  - $G_0(k,t)$: a complex gain to estimate the noise
  - $\widehat{\boldsymbol{\Phi}}_{\mathbf{v}}(k,t) = \alpha_v \widehat{\boldsymbol{\Phi}}_{\mathbf{v}}(k,t-1) + (1-\alpha_v)\hat{\mathbf{v}}(k,t)\hat{\mathbf{v}}^H(k,t)$ where $\hat{\mathbf{v}}(k,t) = G_0(k,t)\mathbf{y}(k,t)$

Noise Covariance Matrix Estimation with Directional Gain

– The Wiener gain for noise estimation can be derived from the minimization of the mean-squared error (MSE) criterion

$$\mathcal{J}[G_0(k,t)] = E[\|\hat{\mathbf{v}}(k,t) - \mathbf{v}(k,t)\|^2]$$

– Authors deduce that $G_{0,\mathrm{W}}(k,t) = \frac{\phi_V(k,t)}{\phi_Y(k,t)}$

- The observation signal variance: $\phi_Y(k,t) = \phi_X(k,t) + \phi_V(k,t)$
- An optimal mask for NCM estimation: $\mathcal{M}_{0,\mathrm{opt}}(k,t) = G_{0,\mathrm{W}}^2(k,t) = \left[1 - G_{1,\mathrm{W}}(k,t)\right]^2$, where $G_{1,\mathrm{W}}(k,t) = \frac{\phi_X(k,t)}{\phi_Y(k,t)}$

Noise Covariance Matrix Estimation with Directional Gain

□ Directional Gain

- Assuming that two fixed beamformers $\mathbf{h}_{\mathrm{P}}(k)$ and $\mathbf{h}_{\mathrm{s}}(k)$ are properly designed, their outputs can be seen as reasonable estimates of the desired signal $X_1(k,t)$.
- The wiener gain $G_{1,\mathrm{W}}(k,t)$ can be approximated with directional gain $G_{\mathrm{dir}}(k,t)$.

$$G_{1,\mathrm{W}}(k,t) \approx G_{\mathrm{dir}}(k,t) = S\left[\frac{\mathbf{h}_{\mathrm{P}}^H(k)\mathbf{\Phi}_{\mathbf{y}}(k,t)\mathbf{h}_{\mathrm{s}}(k)}{\phi_Y(k,t)}\right]$$

$$S(x) = \begin{cases} 1, & \Re(x) > 1 \\ 0, & \Re(x) < 0 \\ \Re(x), & otherwise \end{cases}$$

- Authors design $\mathbf{h}_{\mathrm{P}}(k)$ and $\mathbf{h}_{\mathrm{s}}(k)$ in a heuristic way.

maximum directivity factor (mDF) beamformer

$$\mathbf{h}_{\mathrm{P}}(k) = \frac{[\mathbf{\Gamma}_{\mathrm{dn}}(k) + \epsilon(k)\mathbf{I}_M]^{-1}\mathbf{d}_{\theta_{\mathrm{d}}}(k)}{\mathbf{d}_{\theta_{\mathrm{d}}}^H(k)[\mathbf{\Gamma}_{\mathrm{dn}}(k) + \epsilon(k)\mathbf{I}_M]^{-1}\mathbf{d}_{\theta_{\mathrm{d}}}(k)}$$

where $\mathbf{\Gamma}_{\mathrm{dn}}(k) = \frac{1}{2}\int_0^{\pi} \mathbf{d}_{\theta_{\mathrm{d}}}(k)\mathbf{d}_{\theta_{\mathrm{d}}}^H(k)\sin\theta d\theta$

- $\mathbf{d}_{\theta_{\mathrm{d}}}(k)$: the steering vector at direction θ
- $\epsilon(k)$: the diagonal loading factor
- $\mathbf{I}_M$: the $M \times M$ identity matrix

$$\mathrm{DF}(\omega) \triangleq \frac{|\mathbf{h}^H(\omega,t)\mathbf{d}(\omega,\varphi_d)|^2}{\mathbf{h}^H(\omega,t)\mathbf{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}(\omega,t)}$$

delay-and-sum (DS) beamformer

$$\mathbf{h}_{\mathrm{s}}(k) = \frac{\mathbf{d}_{\theta_{\mathrm{d}}}(k)}{\mathbf{d}_{\theta_{\mathrm{d}}}^H(k)\mathbf{d}_{\theta_{\mathrm{d}}}(k)}$$

$$\mathrm{WNG}(\omega) \triangleq \frac{|\mathbf{h}^H(\omega,t)\mathbf{d}(\omega,\varphi_d)|^2}{\mathbf{h}^H(\omega,t)\mathbf{h}(\omega,t)}$$

- Known to maximize the white noise gain (WNG)

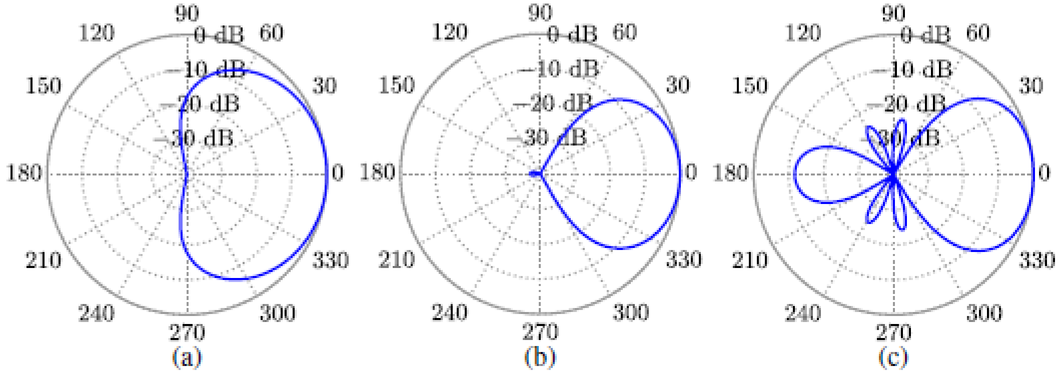
Noise Covariance Matrix Estimation with Directional Gain

□ Beampattern

- Authors define the beampattern of the directional gain $G_{\mathrm{dir}}(k,t)$ as the response to a plane wave from direction θ.
- The mainlobe is along the $0°$ direction and the response at other direction is less than one.
- The goal is to achieve a beampattern with the mainlobe as narrow as possible and sidelobes as small as possible.

$$\mathcal{B}_\theta[G_{\mathrm{dir}}(k,t)] \triangleq S\{\mathcal{B}_\theta[\mathbf{h}_{\mathrm{P}}(k)]\mathcal{B}_\theta^*[\mathbf{h}_{\mathrm{S}}(k)]\}$$

where $\mathcal{B}_\theta[\mathbf{h}(k)] \triangleq \mathbf{h}^H(k)\,\mathbf{d}_\theta(k)$



$f_k = 1$ kHz | $f_k = 4$ kHz | $f_k = 7$ kHz

Conditions: $M = 4, \delta = 2$ cm, $\theta_d = 0°$, and $\epsilon(k) = 0.01$

Noise Covariance Matrix Estimation with Directional Gain

- The parameter $\mathbf{\Phi}_{\mathbf{y}}(k,t)$ and $\phi_Y(k,t)$ can be estimated from the observations directly.

$$\widehat{\mathbf{\Phi}}_{\mathbf{y}}(k,t) = \alpha_y \widehat{\mathbf{\Phi}}_{\mathbf{y}}(k,t-1) + \left(1-\alpha_y\right)\mathbf{y}(k,t)\mathbf{y}^H(k,t)$$

where $0 < \alpha_y < 1$ is a smoothing factor

$$\hat{\phi}_Y(k,t) = \mathrm{tr}\left[\widehat{\mathbf{\Phi}}_{\mathbf{y}}(k,t)\right]/M$$

- With given beamformers and obtained parameters $\widehat{\mathbf{\Phi}}_{\mathbf{y}}(k,t)$ and $\hat{\phi}_Y(k,t)$, the directional gain $G_{\mathrm{dir}}(k,t)$ is computed.
- The gain is then used to obtain the estimate of the Wiener gain, i.e., $\hat{G}_{1,\mathrm{W}}(k,t)$ for the optimal mask $\mathcal{M}_{0,\mathrm{opt}}(k,t)$, which is subsequently applied to estimate the NCM $\widehat{\mathbf{\Phi}}_{\mathbf{v}}(k,t)$.

$$G_{1,\mathrm{W}}(k,t) \approx G_{\mathrm{dir}}(k,t) = S\left[\frac{\mathbf{h}_{\mathrm{P}}^H(k)\mathbf{\Phi}_{\mathbf{y}}(k,t)\mathbf{h}_{\mathrm{S}}(k)}{\phi_Y(k,t)}\right]$$

$$\mathcal{M}_{0,\mathrm{opt}}(k,t) = G_{0,\mathrm{W}}^2(k,t) = \left[1 - G_{1,\mathrm{W}}(k,t)\right]^2$$

$$\widehat{\mathbf{\Phi}}_{\mathbf{v}}(k,t) = \alpha_v \widehat{\mathbf{\Phi}}_{\mathbf{v}}(k,t-1) + (1-\alpha_v)\mathcal{M}_0(k,t)\mathbf{y}(k,t)\mathbf{y}^H(k,t)$$

4. Simulations

Simulations

□ Setup

- ULMA of 4 sensors with the interelement spacing being $2\ \mathrm{cm}$
- Room size: $6\ \mathrm{m} \times 5\ \mathrm{m} \times 4\mathrm{m}$
- A desired source
  - $1.5\ \mathrm{m}$ away from the array center
  - The endfire direction (denoted as $0°$) of the array
- An interference source
  - $2.0\ \mathrm{m}$ away from the array center
  - Incident angle: θ_{i}

Simulations

□ Setup

- The desired and interference sources are two clean speech signals sampled at 16 kHz.
- The duration of the signal: $30\ \mathrm{s}$
- The impulse responses from the sources to the sensors are generated with the image model method.
- The reflections coefficients of the walls, floor and ceiling are assumed to be the same and are controlled.
  - The reverberation time T_{60}: $165\ \mathrm{ms}$
- The noise signals received at the sensors are composed of the interference signal (point-source noise) and the ambient diffuse noise.
  - The interference-to-diffuse-noise power ratio (IDNR): $15\ \mathrm{dB}$

Simulations

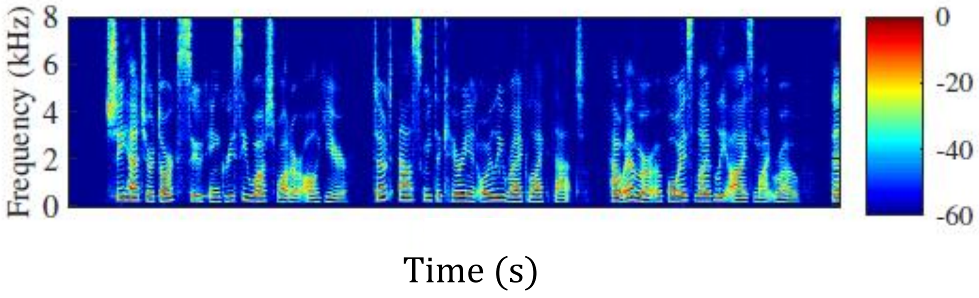
□ Setup

- Frame length: $16\ \mathrm{ms}$ (Hamming window)
- Overlapping factor: 75%
- FFT length: 256
- To implement the primary beamformer,
  - $\epsilon(k) = 0.01$
  - Forgetting factors α_y and α_v are set, respectively to 0.85 and 0.99.
- The baseline algorithms: mDF beamformer, MPDR beamformer

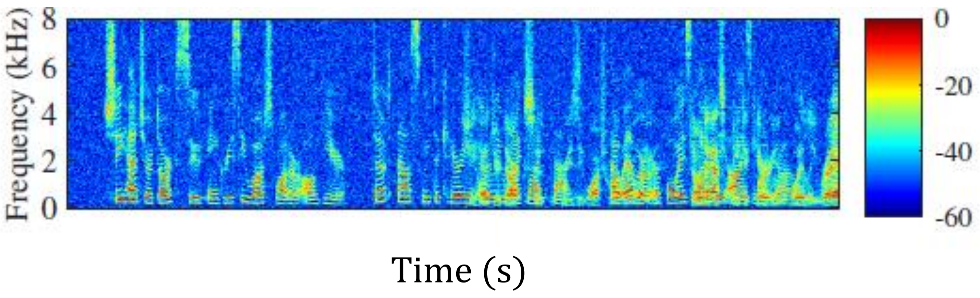
Simulations

□ Results

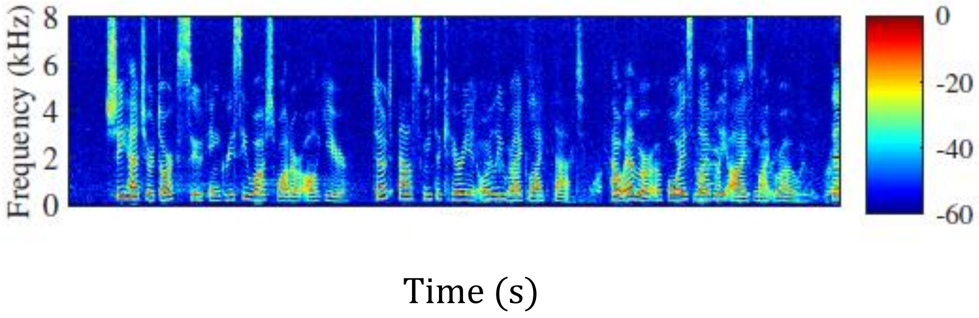
The desired speech signal



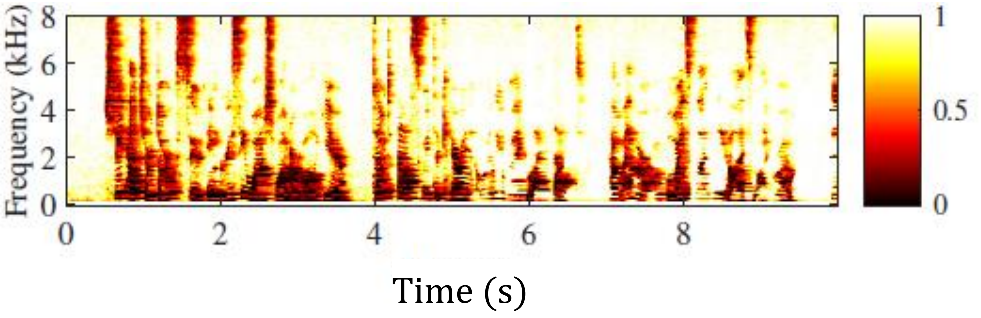
The observation signal at the first sensor



The enhanced signal by the MVDR beamformer



The estimated optimal mask $\mathcal{M}_{0,\mathrm{opt}}(k,t)$



– The input SNR is 0 dB and $\theta_{\mathrm{i}} = 120°$.

Simulations

□ Results

- The MPDR beamformer performs well at very low SNRs, whereas its performance degrades quickly as the SNR increases.

$$\mathbf{h}_{\mathrm{MPDR}}(k,t)=\frac{\mathbf{\Phi}_{\mathbf{y}}^{-1}(k,t)\mathbf{d}_{\theta_{\mathrm{d}}}(k)}{\mathbf{d}_{\theta_{\mathrm{d}}}^{H}(k)\mathbf{\Phi}_{\mathbf{y}}^{-1}(k,t)\mathbf{d}_{\theta_{\mathrm{d}}}(k)}$$

  - Probably due to the signal cancellation problem at high SNR conditions caused by the multipath propagation effect

- The mDF beamformer does not rely on any statistics estimation and has a consistent performance at different SNRs.

$$\mathbf{h}_{\mathrm{mDF}}(k)=\frac{[\mathbf{\Gamma}_{\mathrm{dn}}(k)+\epsilon(k)\mathbf{I}_{M}]^{-1}\mathbf{d}_{\theta_{\mathrm{d}}}(k)}{\mathbf{d}_{\theta_{\mathrm{d}}}^{H}(k)[\mathbf{\Gamma}_{\mathrm{dn}}(k)+\epsilon(k)\mathbf{I}_{M}]^{-1}\mathbf{d}_{\theta_{\mathrm{d}}}(k)}$$

where $\mathbf{\Gamma}_{\mathrm{dn}}(k)=\frac{1}{2}\int_{0}^{\pi}\mathbf{d}_{\theta_{\mathrm{d}}}(k)\mathbf{d}_{\theta_{\mathrm{d}}}^{H}(k)\sin\theta d\theta$

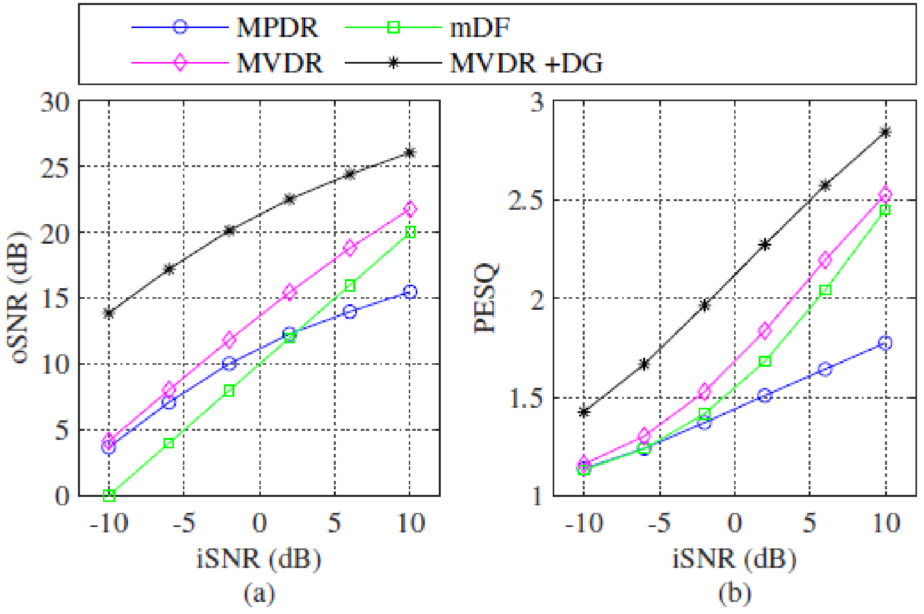


Fig. 3. (a) The output SNR and (b) PESQ of the considered enhancement solutions as a function of the input SNR.

Obtained by averaging over five different interference angles, i.e., θ_{i}: $\{60°, 90°, 120°, 150°, 180°\}$.

5. Conclusions

Conclusions

– This paper presented an NCM estimation method based on the principle of time-frequency masking.

– Authors presented an optimal mask formulation and employed the directional gain method which relies on the signal incident angle only.

– In comparison with existing methods, the presented mask-based NCM estimator does not suffer from the generalization and permutation problems.

– The simulation result showed that integrating the proposed NCM estimator into the MVDR beamformer leads to much better speech enhancement performance as compared to the widely used MPDR and mDF beamformers.

Thank you for listening

Q & A