Signal Processing 105 (2014) 338–349

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Multi-channel post-filtering based on spatial coherence measure

Jwu-Sheng Hu *, Ming-Tang Lee

Department of Electrical Engineering, National Chiao-Tung University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form
15 March 2014
Accepted 17 April 2014
Available online 15 May 2014

Keywords:
Coherence
Multi-channel post-filtering
Microphone array
Multi-rank signal model

ABSTRACT

A multi-channel post-filtering algorithm using the proposed spatial coherence measure is derived. The spatial coherence measure evaluates the similarity between the measured signal fields using power spectral density matrices. In the proposed post-filter, the assumption of homogeneous sound fields is relaxed. Besides, multi-rank signal models can be easily adopted. Under this measure, the bias term due to the similarity of the desired signal field and the noise field is further investigated and a solution based on bias compensation is proposed. It can be shown that the compensated solution is equivalent to the optimal Wiener filter if the bias or the noise power spectral density matrix is perfectly measured. Simulations with incoherent, diffuse, and coherent noise fields and a local scattered desired source were conducted to evaluate the algorithms. The results demonstrate the superiority of the proposed bias compensated post-filter across different types of noise fields with a more accurate signal model.

1. Introduction

Multi-channel speech enhancement has attracted much attention in recent years. In the real world, desired speech signals are often corrupted by background noises, speech interferences, and reverberation. For more than two microphones, there are two main categories of speech enhancement approach: beamforming and multi-channel post-filtering. Beamforming has been applied to several narrow- or wide-band signals processes, which can be defined by a *filter-and-sum* process [1] in the conventional sense. A well-known designing strategy is to preserve the signal from the direction of interest while attenuating others, which can be achieved by the minimum variance distortionless response (MVDR) algorithm [2,3]. The MVDR beamforming is optimal in the mean square error (MSE) sense when the interference-plus-noise power spectral density (PSD) matrix can be obtained and there is no mismatch on the presumed steering vector. Typically, adaptive filtering techniques are applied to estimate the PSD matrix and additional training processes or *a priori* information of signal presence is needed for offline or online implementation [1–4]. On the other hand, the multi-channel post-filtering, which considers both the spatial information and the signal-to-noise ratio (SNR), can be designed in a more general way. Simmer et al. [5] show that the optimal minimum mean square error (MMSE) solution can be decomposed into an MVDR beamformer followed by a single-channel Wiener filter. This solution is also called a *multi-channel Wiener filter*.

Most post-filtering algorithms aim to enhance the single-channel Wiener filter by a more accurate estimation of SNR. The SNR estimation for speech enhancement can be implemented based on the minimum statistics for the stationary noise [6–8], or the spatially pre-processed power [9]. Most of them are energy-based. Alternatively, the phase information of a microphone pair has already

* Corresponding author.
E-mail addresses: jshu@cn.nctu.edu.tw (J.-S. Hu), lhoney.ece97g@g2.nctu.edu.tw (M.-T. Lee).

http://dx.doi.org/10.1016/j.sigpro.2014.04.020
0165-1684/

been used in blind source separation (BSS) [10] as well as the computational auditory scene analysis (CASA) [11]. Aarabi et al. [12–14] provide a different view of the SNR from the phase error perspective for the dual-channel case. In their work, the relationship between the phase error and the SNR was derived [12]. However, the idea of phase error can only be applied to the case of two-microphone. In addition to the SNR estimation, some post-filtering algorithms directly estimate the spectral densities [15–17]. Like the case of phase error, the cross-spectral density is usually defined between two microphones. For more than two microphones, the common practice is to perform average among all distinct microphone pairs [15,16]. Although this might enhance the robustness of the estimation, there is still no formal proof regarding its effectiveness. In particular, it does not consider the spatial arrangement of microphones, i.e., the advantages of using more than two microphones is not fully explored. In this paper, a new spatial measure is defined on a microphone array which leads to a novel post-filtering algorithm (named *spatial coherence based post-filter*, SCPF). The post-filter belongs to the class of spectral densities estimators (which is inherent in the estimation of the input PSD matrix), while it is guaranteed to lie in the range of [0, 1]. Further, the proposed spatial coherence measure can be easily extended to multi-rank signal models encompassing incoherently scattered source, etc. Multi-rank signal models or rank relaxation has been widely used in sensor array localization [18–21], beamforming [22–25], or quadratic optimization problems [25,26]. It is more convenient to consider various design requirements than previous methods using microphone array.

However, a bias term due to the similarity of the desired signal field and the noise field deteriorates the noise reduction performance. As a result, a bias compensated method is proposed (called *bias compensated spatial coherence based post-filter*, BC-SCPF). It can be shown that the BC-SCPF is equivalent to the optimal Wiener filter if the bias or the noise PSD matrix is perfectly measured. Three kinds of noise fields were used with a local scattered source for analysis: incoherent, diffuse, and coherent. Three ITU-T standards were computed to evaluate the perceptual quality and the noise reduction performance. The simulation results show the superiority of the proposed BC-SCPF with a more accurate signal model in all noise fields comparing with various methods proposed before.

The paper is organized as follows. Section 2 states the objective and reviews some related works. In Section 3, a trace inequality is introduced and a coherence measure is defined based on it. The SCPF and BC-SCPF are proposed in Section 4. The simulation setup and results with three noise fields are presented in Section 5, and Section 6 gives the conclusion.

2. Problem formulation and prior works

2.1. Problem formulation

Consider a linear array with M omni-directional microphones. The observation vector is given by

$$\mathbf{x}(t)=\mathbf{s}(t)+\mathbf{n}(t) \tag{1}$$

where $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are the desired signal and noise. Both of them can be multi-dimensional. By assuming locally time-invariant transfer functions and applying the short-time Fourier transform (STFT), the observations are divided in time into overlapping frames by the application of a window function and analyzed in the time–frequency domain as,

$$\mathbf{x}(\omega,k)=\mathbf{s}(\omega,k)+\mathbf{n}(\omega,k) \tag{2}$$

where ω and k are discrete frequency and frame indices respectively.

A beamforming method aims to find a spatial filter $\mathbf{w}$ to estimate the desired source by

$$y(\omega,k)=\mathbf{w}^H(\omega,k)\mathbf{x}(\omega,k) \tag{3}$$

A post-filtering method aims to find a gain function (or mask) to suppress the undesired noise, which can be multiplied on the beamformer output as

$$\hat{s}(\omega,k)=G(\omega,k)\cdot y(\omega,k) \tag{4}$$

2.2. Multi-channel post-filtering based on noise field coherence

McCowan et al. [16] proposed a multi-channel post-filter as a modification of the Zelinski post-filter [15]. In their systems, the microphones have to pass a time alignment module to adjust the propagation of the desired source between microphones before the post-filter estimation, which is equivalent to the information in the presumed steering vector $\mathbf{a}_s(\omega)$. That is, the pre-processed input vector $\tilde{\mathbf{x}}(\omega,k)$ after the time alignment module can be written as

$$\tilde{\mathbf{x}}(\omega,k)=\mathbf{x}(\omega,k)\circ\mathbf{a}_s(\omega) \tag{5}$$

where $\circ$ denotes the Schur–Hadamard (elementwise) matrix product. If the desired signal is a point source, the presumed steering vector $\boldsymbol{a}_s(\omega)$ can be equivalent to the truncated impulse response $\boldsymbol{h}(\omega)$ if the magnitudes and time delays of the source to the microphones are exactly measured. However, the steering vector is not sufficient to describe general cases, which will be discussed in detail in Section 4.1.

Compared to the Zelinski post-filter, the work in [16] considered a generalized coherence function to describe the characteristics of the noise field on the aligned inputs. Noises between sensors can be coherent (or correlated). The noise coherence function of the time aligned inputs is defined as

$$\tilde{\Gamma}_{n_in_j}(\omega)=\tilde{\phi}_{n_in_j}(\omega)/\sqrt{\tilde{\phi}_{n_in_i}(\omega)\cdot\tilde{\phi}_{n_jn_j}(\omega)} \tag{6}$$

where $\tilde{\phi}_{n_in_j}(\omega)$ is the cross-spectral density between the noises at the i-th and j-th microphones. Note that the diagonal terms of $\tilde{\boldsymbol{\Gamma}}_n(\omega)$ are 1 and its trace equals to M. In their works, the homogeneous sound fields are assumed. That is, the sources have the same power spectrum at each sensor. Based on this assumption, the spectral densities of the aligned inputs are expressed as [16]

$$\tilde{\phi}_{x_ix_i}(\omega)=\tilde{\phi}_s(\omega)+\tilde{\phi}_n(\omega) \tag{7}$$

$$\tilde{\phi}_{x_jx_j}(\omega)=\tilde{\phi}_s(\omega)+\tilde{\phi}_n(\omega) \tag{8}$$

$$\tilde{\phi}_{x_i x_j}(\omega) = \tilde{\phi}_s(\omega) + \tilde{\Gamma}_{n_i n_j}(\omega)\tilde{\phi}_n(\omega) \tag{9}$$

where $\tilde{\phi}_s(\omega)$, $\tilde{\phi}_n(\omega)$ are the aligned power spectral densities of the desired signal and noise. For each microphone pair, according to (7)–(9), the signal power spectral density can be estimated as

$$\tilde{\phi}_s^{(i,j)}(\omega) = \frac{\Re\{\tilde{\phi}_{x_i x_j}(\omega)\} - (1/2)\Re\{\tilde{\Gamma}_{n_i n_j}(\omega)\}(\tilde{\phi}_{x_i x_i}(\omega) + \tilde{\phi}_{x_j x_j}(\omega))}{(1 - \Re\{\tilde{\Gamma}_{n_i n_j}(\omega)\})} \tag{10}$$

where $\tilde{\phi}_{x_i x_j}(\omega)$ is the cross-spectral density between the i-th and j-th aligned inputs and $\Re\{\cdot\}$ is the real operator. The spectral densities can be estimated using a first-order recursive filter. Eq. (10) can be explained as removing the highly coherent part of the cross-spectral density and then compensating the residual.

The estimation can be improved by averaging the solutions over all sensor combinations, resulting in the post-filter

$$G_{\mathrm{McCowan}}(\omega) = \frac{\frac{2}{M(M-1)}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\tilde{\phi}_s^{(i,j)}(\omega)}{\frac{1}{M}\sum_{i=1}^{M}\tilde{\phi}_{x_i x_i}}(\omega) \tag{11}$$

This technique significantly improves the noise reduction in the diffuse noise field, and can be applied to any noise field by modeling the complex coherence function. When the noise field is incoherent, it reduces to the Zelinski post-filter as

$$G_{\mathrm{Zelinski}}(\omega) = \frac{\frac{2}{M(M-1)}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\Re\{\tilde{\phi}_{x_i x_j}(\omega)\}}{\frac{1}{M}\sum_{i=1}^{M}\tilde{\phi}_{x_i x_i}}(\omega) \tag{12}$$

3. A trace inequality and its induced coherence measure

3.1. Spatial coherence measure

It is known that the trace of the power spectral density (PSD) matrix, obtained from a sensor array, is the summation of the signal powers. This motivates us to use the trace operation to design a coherence measure between two PSD matrices. Let matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{M\times M}$ be positive semi-definite (which also ensures Hermitian), the trace inequality is established as [27]

$$\mathrm{tr}(\mathbf{AB})^p \le \{\mathrm{tr}(\mathbf{A})^{2p}\mathrm{tr}(\mathbf{B})^{2p}\}^{1/2} \tag{13}$$

where tr($\mathbf{A}$) denotes the trace of matrix $\boldsymbol{A}$, and p is an integer. Considering the special case when $p=1$, we have

$$\mathrm{tr}(\mathbf{AB}) \le \mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B}) \tag{14}$$

Based on (14), the *spatial coherence measure* between PSD matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathscr{F}(\mathbf{A},\mathbf{B}) \equiv \frac{\mathrm{tr}(\mathbf{AB})}{\mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})} \tag{15}$$

According to the Frobenius inner product and Kronecker product, (15) can be written as

$$\mathscr{F}(\mathbf{A},\mathbf{B}) = \frac{\langle \mathbf{A},\mathbf{B}\rangle}{\mathrm{tr}(\mathbf{A}\otimes\mathbf{B})} \tag{16}$$

where $\langle \mathbf{A},\mathbf{B}\rangle$ denotes the Frobenius inner product of PSD matrices $\mathbf{A}$ and $\mathbf{B}$, and $\otimes$ denotes the Kronecker product. The inner product measures the similarity among the bases in the matrices, and the trace of the Kronecker product gives the normalization. From the positive semi-definite property of the matrices and the inequality given by (14), the *spatial coherence measure* $\mathscr{F}(\mathbf{A},\mathbf{B})$ is guaranteed to be mapped in the interval [0, 1]. Since the PSD matrix represents the signal field measured by the sensor array (in the second-order statistics), the proposed *spatial coherence measure* in (15) gives the "closeness" between two *measured signal fields* (named MSF hereafter).

3.2. Properties of proposed spatial coherence measure

The PSD matrices can be decomposed as,

$$\mathbf{A} = \sum_{i=1}^{M}\sigma_i^2(\mathbf{A})\mathbf{u}_i(\mathbf{A})\mathbf{u}_i^H(\mathbf{A}) \quad \text{and} \quad \mathbf{B} = \sum_{j=1}^{M}\sigma_j^2(\mathbf{B})\mathbf{u}_j(\mathbf{B})\mathbf{u}_j^H(\mathbf{B}) \tag{17}$$

where $\sigma_i^2(\mathbf{A})$ and $\mathbf{u}_i(\mathbf{A})$ denote the i-th eigenvalue and eigenvector of the PSD matrix $\boldsymbol{A}$, respectively. By (17), the *spatial coherence measure* can be rewritten as

$$\mathscr{F}(\mathbf{A},\mathbf{B}) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\sigma_i^2(\mathbf{A})\sigma_j^2(\mathbf{B})|\mathbf{u}_i^H(\mathbf{A})\mathbf{u}_j(\mathbf{B})|^2}{\sum_{i=1}^{M}\sigma_i^2(\mathbf{A})\cdot\sum_{j=1}^{M}\sigma_j^2(\mathbf{B})} \tag{18}$$

It can be seen that the coherence measure is the weighted similarity of the bases, and the eigenvalues give the weighting on each basis. When two MSFs belong to the same 1-dimensional subspace, the *spatial coherence measure* gives a measure of unity. As one of the MSF's dimension increases, the *spatial coherence measure* decreases according to the normalization of eigenvalues. Therefore, given PSD matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, several properties of the proposed *spatial coherence measure* can be listed below:

Property 1. *If $\boldsymbol{A}$ belongs to the null-space of $\boldsymbol{B}$, then* $\mathscr{F}(\boldsymbol{A},\boldsymbol{B})=0$.

Property 2. *If $\boldsymbol{A}$ is 1-dimensional subspace, the self-coherence measure* $\mathscr{F}(\mathbf{A},\mathbf{A})=1$. *As the eigenvalue spread of $\boldsymbol{A}$ increases, the* $\mathscr{F}(\mathbf{A},\mathbf{A})$ *decreases to* $1/M$ *until the eigenvalue spread is uniform (i.e., incoherent field, $\boldsymbol{A}=\sigma^2\boldsymbol{I}$ where σ^2 is the signal power).*

Property 3. *If $\boldsymbol{A}$ or $\boldsymbol{B}$ is an incoherent field, then the* spatial coherence measure *equals to a constant value of* $\mathscr{F}(\boldsymbol{A},\boldsymbol{B})=1/M$ *(It can be easily observed from (15)).*

From Property 1, consider $\mathbf{A}$ as the PSD matrix of the desired MSF and $\mathbf{B}$ as the PSD matrix measured by the microphone array. Then $\mathscr{F}(\mathbf{A},\mathbf{B})=0$ could be interpreted as the signals of the microphones do not contain the target source information. Thus, if a multiplicative gain of a post-filter is designed, the gain should be zero. For Property 2,

the self-coherence measure $\mathcal{F}(\mathbf{A},\mathbf{A})$ is derived from (18) as

$$\mathcal{F}(\mathbf{A},\mathbf{A})=\frac{\sum_{i=1}^{M}\sigma_i^4(\mathbf{A})}{\left(\sum_{i=1}^{M}\sigma_i^2(\mathbf{A})\right)^2} \tag{19}$$

where $\mathcal{F}(\mathbf{A},\mathbf{A})$ is purely determined by the eigenvalues of $\boldsymbol{A}$. According to the natural of coherent speech sources, the eigenvalue spread of the desired MSF typically condenses on some low-dimensional subspace. Therefore, if $\mathcal{F}(\mathbf{A},\mathbf{B})$ is used as a multiplicative post-filtering gain, the gain approaches to unity when there is only a desired signal.

In the next section, the general-rank signal models used in array signal processing are introduced, and a novel post-filter is proposed based on the *spatial coherence measure*.

4. Multi-channel post-filtering based on spatial coherence measure

4.1. General-rank signal models in array signal processing

For multi-channel speech enhancement, the *spatial coherence measure* defined in Section 3 can be used to evaluate the similarity between the desired MSF and the input MSF. One commonly used desired MSF is a point source in a homogeneous sound field [9]. Assuming that there is no mismatch between microphones, the desired MSF using the PSD matrix is

$$\boldsymbol{\Phi}_s(\omega)=\phi_s(\omega)\cdot\boldsymbol{\Gamma}_s(\omega) \tag{20}$$

where $\boldsymbol{\Phi}_s(\omega)=E_k[\mathbf{s}(\omega,k)\ \mathbf{s}^H(\omega,k)]$, $\boldsymbol{\Gamma}_s(\omega)=\mathbf{a}_s(\omega)\mathbf{a}_s^H(\omega)$, $\mathbf{a}_s(\omega)$, and $\phi_s(\omega)$ are the PSD matrix, coherence matrix, steering vector, and power spectral density of the desired signal respectively. A single point source is usually referred as the rank-1 signal model.

However, in practice, the rank of signal model is usually greater than 1. Typical examples are incoherently scattered signal source or signals with random fluctuating wavefronts in wireless communication, sonar, and microphone array [18–20]. Further, environmental reverberation also increases the rank. For example, in the case of incoherently scattered source, the desired MSF using the PSD matrix can be expressed by [18–20]

$$\boldsymbol{\Phi}_s(\omega)=\phi_s(\omega)\int_{-\pi/2}^{\pi/2}\rho(\theta,\omega)\mathbf{a}(\theta,\omega)\mathbf{a}^H(\theta,\omega)d\theta \tag{21}$$

where $\rho(\theta,\omega)$ is the normalized angular power density function ($\int_{-\pi/2}^{\pi/2}\rho(\theta,\omega)d\theta=1$), and $\mathbf{a}(\theta,\omega)$ is the steering vector at direction θ. In the case of randomly fluctuating wavefronts, the PSD matrix can be expressed by [22]

$$\boldsymbol{\Phi}_s(\omega)=\phi_s(\omega)\mathbf{B}\circ\{\mathbf{a}_s(\omega)\mathbf{a}_s^H(\omega)\} \tag{22}$$

where $\boldsymbol{B}$ is the M-by-M coherence loss matrix, and $\circ$ is the Schur–Hadamard (elementwise) matrix product. Two commonly used models for the coherence loss matrix are

$$[\mathbf{B}]_{i,j}=\exp\{-(i-j)^2\zeta\} \tag{23}$$

$$[\mathbf{B}]_{i,j}=\exp\{-|i-j|\zeta\} \tag{24}$$

where ζ is the coherence loss parameter. Note that both the signal models in (21) and (22) are multi-rank.

In practice, the desired MSF $\boldsymbol{\Phi}_s(\omega)$ can be estimated empirically from the clean signal recordings of the microphone array. It is worth to note that the usage of $\phi_s(\omega)$ in (20) is not crucial since it is canceled during the normalization of the *spatial coherence measure*.

4.2. The proposed spatial coherence based post-filter

The proposed post-filter is designed by comparing the input PSD matrix $\boldsymbol{\Phi}_x(\omega)$ with a desired one $\boldsymbol{\Phi}_s(\omega)$ as,

$$G_{\mathrm{SCPF}}(\omega)=\frac{\mathrm{tr}(\boldsymbol{\Phi}_s(\omega)\boldsymbol{\Phi}_x(\omega))}{\mathrm{tr}(\boldsymbol{\Phi}_s(\omega))\cdot\mathrm{tr}(\boldsymbol{\Phi}_x(\omega))} \tag{25}$$

where

$$\begin{aligned}\boldsymbol{\Phi}_x(\omega)&=E_k[\mathbf{x}(\omega,k)\mathbf{x}^H(\omega,k)]\\&=\begin{bmatrix}\phi_{x_0x_0}(\omega) & \phi_{x_0x_1}(\omega) & \cdots & \phi_{x_0x_{M-1}}(\omega)\\ \phi_{x_1x_0}(\omega) & \phi_{x_1x_1}(\omega) & \cdots & \phi_{x_1x_{M-1}}(\omega)\\ \vdots & \vdots & \ddots & \vdots\\ \phi_{x_{M-1}x_0}(\omega) & \phi_{x_{M-1}x_1}(\omega) & \cdots & \phi_{x_{M-1}x_{M-1}}(\omega)\end{bmatrix}\end{aligned} \tag{26}$$

and $\phi_{x_ix_j}(\omega)$ is the cross-spectral density between the inputs at the i-th and j-th microphones. The post-filter uses the measure directly as the gain function and is called *spatial coherence based post-filter* (SCPF).

In order to compare with the previous algorithms, we consider the special case as follows:

(1) The desired MSF is assumed to be a point source.
(2) The sound fields are assumed to be homogeneous [9].

According to the these conditions, the theoretical PSD matrix can be expressed as

$$\boldsymbol{\Phi}_x(\omega)=\phi_s(\omega)\mathbf{a}_s(\omega)\mathbf{a}_s^H(\omega)+\phi_n(\omega)\boldsymbol{\Gamma}_n(\omega) \tag{27}$$

where $\phi_s(\omega)$, $\phi_n(\omega)$ are the power spectral densities of the desired signal and noise; and $\boldsymbol{\Gamma}_n(\omega)$ denotes the coherence matrix of the noise field. The manifold vector is usually selected such that $||\mathbf{a}_s(\omega)||^2=M$. Note that $\mathrm{tr}(\boldsymbol{\Gamma}_n(\omega))=M$. With the desired MSF and the theoretical PSD matrix, the SCPF can be expressed by

$$\begin{aligned}G_{\mathrm{SCPF}}(\omega)&=\frac{\phi_s(\omega)\mathbf{a}_s^H(\omega)\boldsymbol{\Phi}_x(\omega)\mathbf{a}_s(\omega)}{\phi_s(\omega)||\mathbf{a}_s(\omega)||^2\cdot\mathrm{tr}(\boldsymbol{\Phi}_x(\omega))}\\&=\frac{\phi_s^2(\omega)||\mathbf{a}_s(\omega)||^4+\phi_s(\omega)\phi_n(\omega)\mathbf{a}_s^H(\omega)\boldsymbol{\Gamma}_n(\omega)\mathbf{a}_s(\omega)}{\phi_s(\omega)||\mathbf{a}_s(\omega)||^2\cdot(\phi_s(\omega)||\mathbf{a}_s(\omega)||^2+\phi_n(\omega)\mathrm{tr}(\boldsymbol{\Gamma}_n(\omega)))}\\&=G_{\mathrm{Wiener}}(\omega)+\frac{c(\omega)}{M^2(\xi(\omega)+1)}\end{aligned} \tag{28}$$

where $G_{\mathrm{Wiener}}(\omega)$ is the optimal Wiener filter as,

$$G_{\mathrm{Wiener}}(\omega)=\frac{\xi(\omega)}{\xi(\omega)+1} \tag{29}$$

and $\xi(\omega)=\phi_s(\omega)/\phi_n(\omega)$ denotes the SNR. The term $c(\omega)$ denotes the Frobenius inner product of the coherence matrices of the desired signal field and the noise field.

$$c(\omega)=\mathbf{a}_s^H(\omega)\boldsymbol{\Gamma}_n(\omega)\mathbf{a}_s(\omega)=\mathrm{tr}(\boldsymbol{\Gamma}_n(\omega)\boldsymbol{\Gamma}_s(\omega))=\langle\boldsymbol{\Gamma}_s(\omega),\boldsymbol{\Gamma}_n(\omega)\rangle \tag{30}$$

This can be treated as a bias term to the optimal Wiener filter as shown in (28). Note that $c(\omega)$ lies in the following range for all kinds of noise fields when the desired MSF is chosen as (20)

$$0 \le c(\omega) < M^2 \tag{31}$$

The lower bound happens when the noise subspace lies in the null-space of the desired MSF, while the supremum happens when the noise MSF is identical to the desired MSF under the rank-1 signal model. Obviously, the SCPF is a function of the SNR and it reduces to the Wiener filter when $c(\omega)=0$.

4.3. Mean square error analysis of proposed SCPF

The mean square error (MSE) corresponding to the desired signal in the reference channel can be defined as

$$\mathrm{MSE}(\omega) = E_k[|\hat{s}(\omega,k)-s(\omega,k)|^2] \tag{32}$$

where $\hat{s}(\omega,k)$ is the enhanced signal given by a beamformer or a post-filter. Applying the SCPF on the reference microphone (microphone 1) results in the following MSE:

$$\begin{aligned}\mathrm{MSE}_{\mathrm{SCPF}}(\omega) &= E_k[|G_{\mathrm{SCPF}}(\omega)x_1(\omega,k)-s(\omega,k)|^2]\\ &= E_k[|(G_{\mathrm{SCPF}}(\omega)-1)s(\omega,k)+G_{\mathrm{SCPF}}(\omega)n_1(\omega,k)|^2]\\ &= |G_{\mathrm{SCPF}}(\omega)-1|^2\phi_s(\omega)+G^2_{\mathrm{SCPF}}(\omega)\phi_n(\omega)\end{aligned} \tag{33}$$

By substituting (28) into (33), we have

$$\mathrm{MSE}_{\mathrm{SCPF}}(\omega) = G_{\mathrm{Wiener}}(\omega)\phi_n(\omega)+\frac{c^2(\omega)}{M^4(\xi(\omega)+1)}\phi_n(\omega) \tag{34}$$

It can be shown that the Zelinski post-filter [15] is related to SCPF as (see Appendix A)

$$G_{\mathrm{Zelinski}}(\omega) = \frac{G_{\mathrm{SCPF}}(\omega)-1/M}{1-1/M} \tag{35}$$

By substituting the SCPF in (28) into (35), we have

$$\begin{aligned}G_{\mathrm{Zelinski}}(\omega) &= \frac{M[(M-1)\xi(\omega)-1]+c(\omega)}{M(M-1)(\xi(\omega)+1)}\\ &= G_{\mathrm{Wiener}}(\omega)-\frac{M-c(\omega)}{M(M-1)(\xi(\omega)+1)}\end{aligned} \tag{36}$$

In (36), it reveals that the Zelinski post-filter gives a negative gain $-1/(M-1)$ when $c(\omega)=0$ and $\xi(\omega)=0$. The negative gain will introduce unwanted phase flips and leaves some noisy time–frequency blocks in the post-filter output. Similarly, the MSE of the Zelinski post-filter can be derived by following the derivation in (33) as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{Zelinski}}(\omega) &= |G_{\mathrm{Zelinski}}(\omega)-1|^2\phi_s(\omega)+G^2_{\mathrm{Zelinski}}(\omega)\phi_n(\omega)\\ &= G_{\mathrm{Wiener}}(\omega)\phi_n(\omega)+\frac{(M-c(\omega))^2}{M^2(M-1)^2(\xi(\omega)+1)}\phi_n(\omega)\end{aligned} \tag{37}$$

The bias terms in (34) and (37) reveals interesting differences between the proposed SCPF and the Zelinski post-filter for different noise fields:

(1) $c(\omega)=M$: the noise field is incoherent, i.e., $\mathbf{\Gamma}_n(\omega)=\mathbf{I}$. In this case, the Zelinski post-filter reduces to the optimal Wiener filter and the proposed SCPF has additional term as $\phi_n(\omega)/[M^2(\xi(\omega)+1)]$.
(2) $c(\omega)=0$: the noise field belongs to the null-space of $\mathbf{a}_s(\omega)$. In this case, on the contrary, the proposed SCPF reduces to the optimal Wiener filter and the Zelinski post-filter has additional term as $\phi_n(\omega)/[(M-1)^2(\xi(\omega)+1)]$.

It is worth to note that when the rank-1 desired MSF is chosen, the proposed SCPF is a special case of the post-filter algorithm [5]. In this case, the proposed SCPF can be explained as the ratio of the output power of the delay-and-sum (DS) beamformer to the sum of the input power. Note that the Zelinski and McCowan post-filters also belong to the same family.

It is also interesting to analyze the MSE of the delay-and-sum (DS) beamformer. Given the DS beamformer as $\mathbf{w}_{\mathrm{DS}}(\omega)=\mathbf{a}_s(\omega)/\|\mathbf{a}_s(\omega)\|$, which introduces no distortion on the desired signal and the MSE thereof is derived as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{DS}}(\omega) &= E_k p\left[|\mathbf{w}^H_{\mathrm{DS}}(\omega)\mathbf{x}(\omega,k)-s(\omega,k)|^2\right]\\ &= E_k\left[\left|\frac{\mathbf{a}^H_s(\omega)\mathbf{n}(\omega,k)}{M}\right|^2\right] = \phi_n\cdot\frac{c(\omega)}{M^2}\end{aligned} \tag{38}$$

It can be seen that the MSE of the DS beamformer is independent of the desired signal.

Before ending this section, we give an illustration to show the difference between the DS beamformer, the Zelinski post-filter, and the SCPF in the MSE sense.

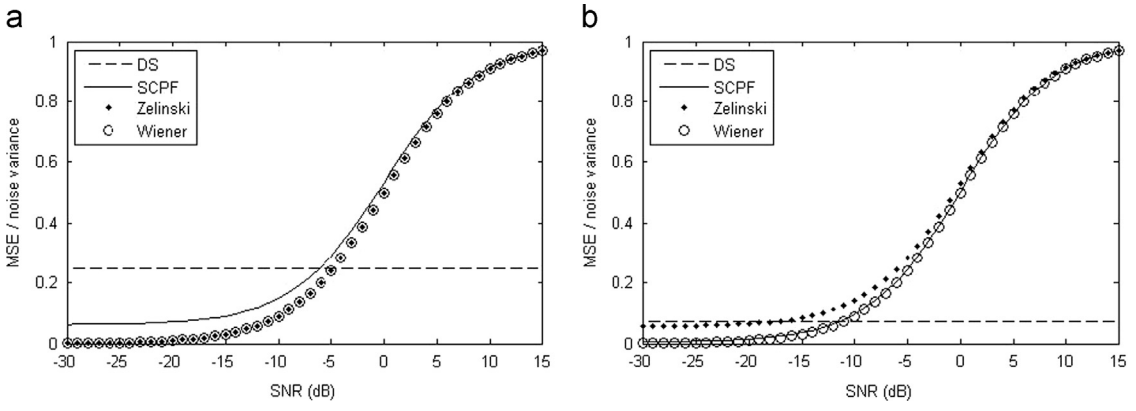

Fig. 1. Comparison among the delay-and-sum (DS) beamformer, the Zelinski post-filter, and the proposed SCPF in the MSE sense: (a) incoherent noise field and (b) a coherent noise impinged from $\theta=45^\circ$.

Consider a uniformly distributed linear array (ULA) with $M=4$ sensors spaced at half-wavelength distance. The $\mathbf{\Gamma}_s(\omega)$ described in (20) was steered at $\theta=0^\circ$. Two different noise fields were analyzed: a coherent interference impinged into the array at $\theta=45^\circ$ and the incoherent white sensor noise. In Fig. 1, compared to the DS beamformer, it can be seen that both the post-filters attenuate more noise component at low SNRs and preserves more noise at high SNRs. Since speeches are highly nonstationary signal, the post-filters are able to give aggressive noise reduction at low SNRs, especially in the case of incoherent noise fields.

4.4. The bias compensated SCPF

Recall from (28), the noise reduction ability of the proposed SCPF is limited due to the additional term $c(\omega)/[M^2(\xi(\omega)+1)]$. Since $c(\omega)$ is the inner product of the coherence matrices of the desired signal field and the noise field, its effect becomes significant at low frequencies where the similarity between coherence matrices is high due to the insufficient spatial sampling. This can happen both in beamforming and multi-channel post-filtering techniques. When the desired signal is absent in the data, the optimal Wiener filter gives a zero gain, which completely removes the noise. However, the SCPF gives a gain of

$$G_{\mathrm{SCPF}}(\omega)\big|_{\xi(\omega)=0}=\frac{\mathrm{tr}(\mathbf{\Phi}_s(\omega)\mathbf{\Phi}_n(\omega))}{\mathrm{tr}(\mathbf{\Phi}_s(\omega))\cdot\mathrm{tr}(\mathbf{\Phi}_n(\omega))}\equiv\beta_\omega \tag{39}$$

where $\mathbf{\Phi}_n(\omega)$ is the noise PSD matrix. Under the assumptions of homogeneous sound fields and point source model, the bias β_ω can be expressed as

$$\beta_\omega=\frac{\mathbf{a}_s^H(\omega)\mathbf{\Gamma}_n(\omega)\mathbf{a}_s(\omega)}{M^2}=\frac{c(\omega)}{M^2} \tag{40}$$

Since $\mathbf{\Phi}_s(\omega)$ is designed *a priori*, the bias term β_ω only depends on the noise PSD matrix $\mathbf{\Phi}_n(\omega)$.

To decrease the effect of the bias β_ω, an intuitive way is to remove the bias and compensate the gain to map the value in the range of [0,1]. The result is called biased-compensated SCPF (BC-SCPF) as follows:

$$G_{\mathrm{BC-SCPF}}(\omega)=\frac{G_{\mathrm{SCPF}}(\omega)-\beta_\omega}{1-\beta_\omega} \tag{41}$$

Note that the bias β_ω lies in the following range for all kinds of noise fields according to the range of $c(\omega)$ given in (31)

$$1/M\le\beta_\omega<1 \tag{42}$$

By substituting (28) and (40) into (41), we have

$$\begin{aligned}G_{\mathrm{BC-SCPF}}(\omega)&=\left[\frac{M^2\xi(\omega)+c(\omega)}{M^2(\xi(\omega)+1)}-\frac{c(\omega)}{M^2}\right]\Bigg/\left(1-\frac{c(\omega)}{M^2}\right)\\&=\left[\frac{\xi(\omega)(M^2-c(\omega))}{M^2(\xi(\omega)+1)}\right]\Bigg/\left(\frac{(M^2-c(\omega))}{M^2}\right)\\&=G_{\mathrm{Wiener}}(\omega)\end{aligned} \tag{43}$$

This gives the optimal Wiener filter if the noise field coherence is perfectly measured. In essence, the BC-SCPF amplifies the small spatial deviation at low frequencies. It is also worth to note that the Zelinski post-filter is a special case of the proposed BC-SCPF with $\beta_\omega=1/M$ according to (35).

For the bias estimation, (39) can be used if the PSD matrices of the desired signal and the noise, $\mathbf{\Phi}_s(\omega)$ and $\mathbf{\Phi}_n(\omega)$, can be obtained in the training process. For the special case of the homogeneous sound fields, the information given by the noise PSD matrix $\mathbf{\Phi}_n(\omega)$ equals to that of the noise coherence matrix $\mathbf{\Gamma}_n(\omega)$. Furthermore, on-line implementation of the bias estimation can be achieved since the bias is the smallest gain of proposed SCPF at each discrete frequency if the noise field does not change. Thus, the minimum tracking skills [6–8] can be conducted and implemented on-line.

4.5. Comparison between BC-SCPF and McCowan post-filter

Under the assumption of homogeneous sound field and rank-1 signal model, the McCowan post-filter has been derived as [16]

$$G_{\mathrm{McCowan}}(\omega)=G_{\mathrm{Wiener}}(\omega)+\underbrace{\frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)}\left[\frac{2}{M(M-1)}\Re\left\{\frac{\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\tilde{\Gamma}_{n_in_j}(\omega)-\hat{\tilde{\Gamma}}_{n_in_j}(\omega)}{1-\hat{\tilde{\Gamma}}_{n_in_j}(\omega)}\right\}\right]}_{e_1} \tag{44}$$

where $\tilde{\Gamma}_{n_in_j}(\omega)$ and $\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$ are the actual and estimated noise coherence matrices of the aligned inputs. From (44) it can be easily seen that the McCowan post-filter reduces to the Wiener filter when the noise coherence matrix is perfectly measured.

Similarly, the proposed BC-SCPF can be expressed with $\tilde{\Gamma}_{n_in_j}(\omega)$ and $\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$ as (see Appendix B)

$$G_{\mathrm{BC-SCPF}}(\omega)=G_{\mathrm{Wiener}}(\omega)+\underbrace{\frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)}\left[\frac{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}(\tilde{\Gamma}_{n_in_j}(\omega)-\hat{\tilde{\Gamma}}_{n_in_j}(\omega))}{\sum_{i=1}^{M}\sum_{j=1}^{M}(1-\hat{\tilde{\Gamma}}_{n_in_j}(\omega))}\right]}_{e_2} \tag{45}$$

By comparing (44) and (45), it can be observed that the error term e_1 in the McCowan post-filter is the average of the ratios $(\tilde{\Gamma}_{n_in_j}(\omega)-\hat{\tilde{\Gamma}}_{n_in_j}(\omega))/(1-\hat{\tilde{\Gamma}}_{n_in_j}(\omega))$ for each microphone pair, while the error term e_2 in the proposed BC-SCPF is the ratio of averaged $\tilde{\Gamma}_{n_in_j}(\omega)-\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$ and $1-\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$. It is known that the averaging before division may be robust to the estimation errors. In other words, the error term e_1 is sensitive to the cases such as one of the estimated $\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$ approaches to unity or is significantly different from the true noise coherence matrix. The effects are alleviated after the averaging in the proposed BC-SCPF. Listed below are potential advantages of the proposed BC-SCPF comparing with Zelinski and McCowan post-filters.

(1) Multi-rank signal models can be directly adopted in the proposed method.
(2) The assumption of homogeneous sound fields used in Zelinski and McCowan post-filters can be relaxed in the proposed method. It becomes crucial when the near-field effect or the sound attenuation is taken into account.
(3) Compared to the estimation of the noise field coherence under each microphone pairs, the proposed

method merged those as one bias term, which can be designed in many ways.

(4) The proposed BC-SCPF is less sensitive to the individual estimation error of the noise coherence function (or the noise PSD function).

4.6. Estimation of the PSD matrices

In practice, $\mathbf{\Phi}_s(\omega)$ can be estimated empirically from the clean signal recordings of the microphone array. Typically, statistically white or sweep sinusoid signals can be chosen as the training signals. The training signal is uttered from a sound device and recorded by the microphone array to model the desired sound propagation. The recorded training signals are analyzed using STFT, and then $\mathbf{\Phi}_s(\omega)$ can be computed using sample mean of the input vectors under each discrete frequency ω as

$$\hat{\mathbf{\Phi}}_s(\omega)=\frac{1}{N}\sum_{k=1}^{N}\mathbf{s}(\omega,k)\mathbf{s}^H(\omega,k) \tag{46}$$

where N is the sample size. It is worth to note that the spectral density of the training signal $\phi_s(\omega)$ in (20) is not crucial since it is canceled during the normalization of the *spatial coherence measure*.

Next, the PSD matrix $\mathbf{\Phi}_x(\omega)$ can be estimated using a first-order recursive update formula

$$\hat{\mathbf{\Phi}}_x(\omega,k)=\alpha\hat{\mathbf{\Phi}}_x(\omega,k-1)+(1-\alpha)\mathbf{x}(\omega,k)\mathbf{x}^H(\omega,k) \tag{47}$$

where α is the forgetting factor close to unity. Note that the PSD matrices estimated by (46) and (47) are guaranteed to be semi-positive definite.

5. Simulation results

In this section, the comparison between beamformers and post-filter algorithms is shown first. Second, the sensitivity to the array imperfection for the proposed BC-SCPF and the McCowan post-filter is analyzed. Finally, the performances with different number of microphones are investigated.

5.1. Comparisons of algorithms

In this section, we use three different noise fields and several SNR conditions to evaluate the proposed algorithms. The simulations were generated by the room impulse response generator [29] with reverberation corresponding to the reverberation time $RT_{60}=503$ ms using Sabin–Franklin's formula. There are three noise field conditions: (1) stochastically white noise where noises between microphones are uncorrelated (i.e., incoherent noise field); (2) babble noises which were uttered from four corners of the room to simulate a diffuse noise field; (3) speech interference which is a coherent source impinged into the array at the direction of 45°. The desired source impinged into the array at the direction 0° with local scattering as described in Fig. 2. The angular power density $\rho(\theta,\omega)$ used in (21) is a Gaussian function with standard deviation set as 10. A uniformly distributed linear array (ULA) with eight omni-directional microphones with 5 cm spacing was used. The simulation environment is illustrated in Fig. 3. The sampling rate and the fast Fourier transform (FFT) size were 8 kHz and 256, respectively. A female voice and a male voice were used as the desired source and the interference respectively. The white and babble noise signals were taken from the NOISEX-92 database [30]. All the recordings were 60 s in duration and combined into different SNR conditions. The speech quality was evaluated by ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) [31]. Higher PESQ score indicates better speech quality. For noise reduction performance, SNRI (Signal-to-Noise Ratio Improvement) from ITU-T G.160 [32] was computed.

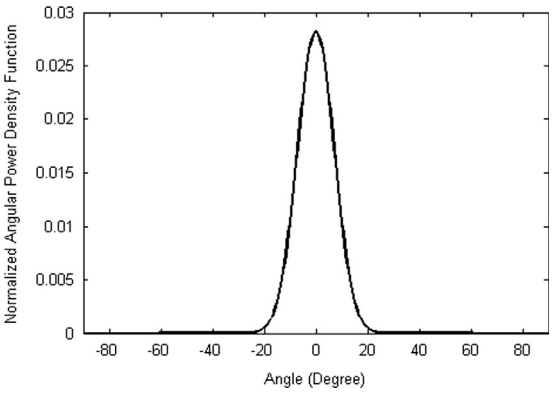

Fig. 2. Normalized angular distribution function $\rho(\theta)$ of the scattered source for all frequencies.

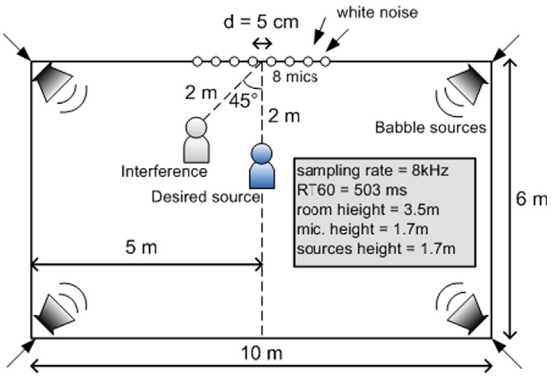

Fig. 3. Simulation environment. A ULA with eight omni-directional microphones, a desired speech source, an unwanted speech interference, and four babble sources were set up.

The detailed parameter settings and abbreviations of the algorithms are listed below:

(1) DS: Delay-and-sum beamformer.

(2) SD: Super-directive beamformer [28]. The ratio of $(\sigma_w^2/\phi_n(\omega))$ was chosen as -20 dB.

(3) Zelinski: Zelinski post-filter [15]. The post-filter was implemented using (12), where the spectral densities were estimated using a first-order recursive filter with the forgetting factor $\alpha=0.875$.

(4) McCowan: McCowan post-filter based on noise field coherence [16]. The post-filter was implemented using (10) and (11) with the noise coherence matrices

estimated by (6). The coherence matrices were trained with 1875 noise-only frames (30 s) for each case.

(5) SCPF: The proposed method was implemented using (25) and (47). The forgetting factor $\alpha=0.875$ for estimating the PSD matrices was used for the proposed methods, which is the same as the factor used in Zelinski and McCowan post-filters.

(6) BC-SCPF (Rank-1): The proposed method was implemented using (41) with the same training noise data as the McCowan et al. The biases were then computed using (39). For comparison, $\mathbf{\Phi}_s(\omega)$ was implemented by the point source model in (20) without training.

(7) BC-SCPF(Multi-rank): The proposed method was implemented using (39) and (41). $\mathbf{\Phi}_s(\omega)$ was trained using (46).

All the post-filters are processed on the output of the DS beamformer.

In the following discussion, the evaluation of the speech quality using PESQ score improvement is studied, as shown in Table 1. Consider the DS and SD beamformers. It is known that the delay-and-sum (DS) beamformer is optimal for the MVDR design in the incoherent noise field. Therefore, the DS beamformer is ensured to perform better than the SD beamformer in this case. The SD beamformer increases the directivity at low frequencies, or in other words, it amplifies the small deviations between microphones to obtain more noise reduction. As a result, the SD beamformer has better performance than the DS beamformer in the diffuse and coherent noise fields, where the insufficient spatial sampling has to be taken into account in these cases. However, it has some artifacts in the incoherent noise field due to the increased white noise gain [1]. Compared to the DS beamformer, the usages of post-filters give better performances. That means the post-filters followed by a DS beamformer have contributions to both the speech quality and noise reduction.

The Zelinski post-filter is a special case of the McCowan post-filter when the noise field is incoherent. Hence, it can be seen that in the incoherent noise field, the performances of the Zelinski and the McCowan post-filters are almost the same. While in other noise fields, the consideration of noise field coherence provides evident performance improvements. Likewise, the proposed BC-SCPF after the bias compensation has evident performance improvements compared to the proposed SCPF. However, unlike the relationship between the Zelinski and the McCowan post-filters in the incoherent noise field, the bias compensation still improves the performance in this scenario.

For comparison between the Zelinski post-filter and the proposed SCPF, the SCPF has better performances in the diffuse and coherent noise fields. While in the incoherent noise field, according to the theoretical analysis in Section 4.3, the Zelinski post-filter should always have better performance than the SCPF. However, at high input SNR conditions, the assumption of homogeneous sound fields leads to signal distortion and degrades the speech quality.

Next, the performance of the McCowan post-filter and the proposed BC-SCPF is discussed. Here, the BC-SCPFs were implemented using the rank-1 and multi-rank

Table 1
PESQ score improvement under different noise fields.

Input SNR→	PESQ score improvement			
Algorithm ↓	5 dB	10 dB	15 dB	20 dB
White noise (incoherent)				
Original noisy PESQ	1.77	2.05	2.39	2.72
DS	0.49	0.53	0.53	0.54
SD	0.03	0.05	0.05	0.04
DS+Zelinski	1.03	0.95	0.84	0.75
DS+SCPF	0.98	0.94	0.85	0.77
DS+McCowan	1.03	0.95	0.84	0.75
DS+BC-SCPF(Rank-1)	1.03	0.96	0.84	0.75
DS+BC-SCPF(Multi-Rank)	**1.08**	**1.01**	**0.89**	**0.79**
Babble noise (diffuse)				
Original noisy PESQ	2.05	2.35	2.65	2.95
DS	0.36	0.36	0.35	0.33
SD	**0.57**	**0.57**	**0.57**	**0.58**
DS+Zelinski	0.36	0.38	0.37	0.35
DS+SCPF	0.43	0.43	0.41	0.38
DS+McCowan	0.55	0.52	0.46	0.43
DS+BC-SCPF(Rank-1)	0.54	0.51	0.46	0.42
DS+BC-SCPF(Multi-Rank)	0.57	0.54	0.49	0.45
Speech interference (coherent)				
Original noisy PESQ	2.28	2.57	2.86	3.16
DS	0.28	0.30	0.30	0.29
SD	0.47	0.48	0.48	0.47
DS+Zelinski	0.34	0.37	0.36	0.34
DS+SCPF	0.37	0.39	0.38	0.36
DS+McCowan	0.35	0.45	0.48	0.45
DS+BC-SCPF(Rank-1)	0.66	0.62	0.55	0.47
DS+BC-SCPF(Multi-Rank)	**0.69**	**0.64**	**0.57**	**0.49**

Table 2
SNRI score obtained by different input SNRS.

Input SNR→	**SNRI** (dB)			
Algorithm ↓	5 dB	10 dB	15 dB	20 dB
White noise (incoherent)				
DS	9.38	9.34	9.07	8.51
SD	−0.02	−0.07	−0.27	−0.49
DS+Zelinski	26.12	23.46	20.44	17.16
DS+SCPF	20.69	19.33	17.43	15.63
DS+McCowan	26.15	**23.48**	**20.46**	**17.17**
DS+BC-SCPF(Rank-1)	26.14	22.73	19.63	16.28
DS+BC-SCPF(Multi-Rank)	**26.70**	22.73	19.63	16.28
Babble noise (diffuse)				
DS	3.55	3.36	3.11	2.82
SD	6.83	6.81	6.62	6.23
DS+Zelinski	5.10	5.03	4.55	3.95
DS+SCPF	4.63	4.42	3.98	3.43
DS+McCowan	**12.64**	**11.96**	**10.30**	**8.26**
DS+BC-SCPF(Rank-1)	11.85	11.60	9.94	8.03
DS+BC-SCPF(Multi-Rank)	11.73	11.48	9.81	7.93
Speech interference (coherent)				
DS	0.20	0.03	−0.13	−0.32
SD	2.33	2.31	2.18	1.92
DS+Zelinski	0.86	0.60	0.43	0.17
DS+SCPF	0.73	0.59	0.29	0.06
DS+McCowan	1.64	2.97	3.69	3.34
DS+BC-SCPF(Rank-1)	**10.96**	**9.80**	7.59	5.17
DS+BC-SCPF(Multi-Rank)	10.94	9.77	**7.60**	**5.24**

models for comparison. For the former, it uses the same presumed point source model as the McCowan post-filter. With the same information, it can be observed that there is no big difference between BC-SCPF(Rank-1) and McCowan post-filter under incoherent and diffuse noise fields. But the superiority becomes obvious in the coherent noise field. One of the possible reasons is the homogeneous assumption used in McCowan post-filter that breaks down its performance. For incoherent and diffuse noise fields, the noise PSDs in each microphone are similar. However, the coherent noise field is no more homogeneous due to the sound attenuation during propagation. When the multi-rank $\mathbf{\Phi}_s(\omega)$ is used, it shows that better speech quality can be achieved.

Second, consider the SNRI results given in Table 2. Compared to Table 1, typically higher speech quality corresponds to higher noise reduction performance. However, higher noise reduction performance does not guarantee a better speech quality since it may also cause signal distortion. In the incoherent and diffuse noise fields, the McCowan post-filter has slightly larger SNRIs than that of the BC-SCPF within 1 dB, but it has smaller PESQ score improvements. It indicates that the multi-rank signal model reduces signal distortion and leads to a better speech quality with comparable noise reduction. In addition, the coherence and bias estimation in McCowan post-filter and the BC-SCPF may be over-estimated which gives aggressive noise reduction for low SNR observations. It is the reason why the SNRI of the SD beamformer is smaller than the SNRIs of the above mentioned two post-filters, even though it has the highest PESQ scores.

In the last, one example of the signal spectrograms is given in Fig. 4 as reference. The noise condition is the speech interference at 10 dB SNR. The noise suppression of the DS beamformer is effective at high frequency bands but is limited at low frequency bands due to the

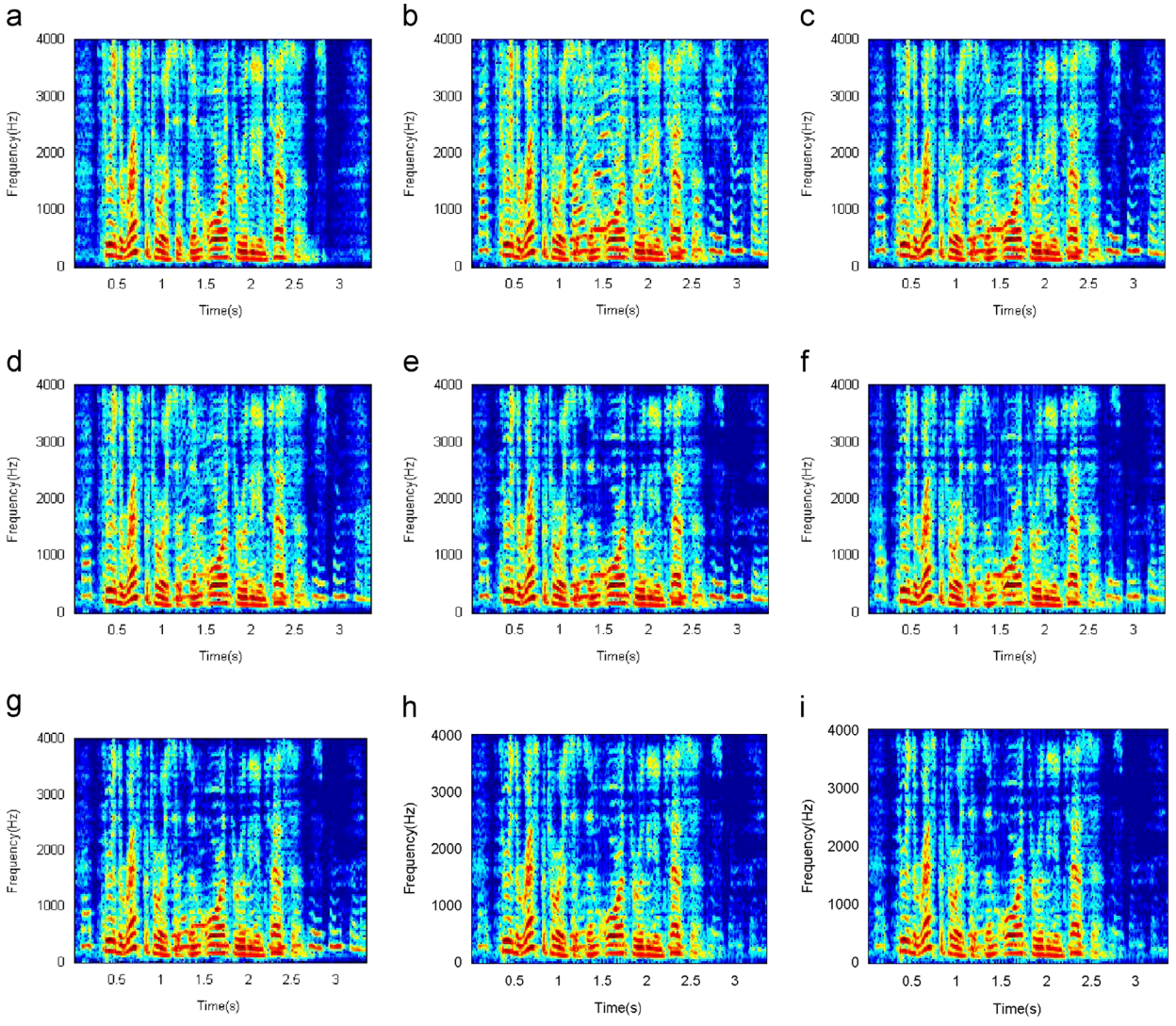

Fig. 4. Signal spectrograms (speech interference at 10 dB SNR). (a) Clean signal, (b) noisy input, (c) DS, (d) SD, (e) DS+Zelinski, (f) DS+McCowan, (g) DS+SCPF, (h) DS+BC-SCPF(Rank-1), and (i) DS+BC-SCPF(Multi-Rank).

Table 3
PESQ score improvement with microphone #8 gain mismatches (under speech interference).

Input SNR→	PESQ Score Improvement			
Algorithm ↓	5 dB	10 dB	15 dB	20 dB
Original noisy PESQ	2.28	2.57	2.86	3.16
DS+McCowan				
No mismatch	0.36	0.45	0.48	0.45
Mismatch=1 dB	0.34	0.40	0.38	0.28
Mismatch=2 dB	−0.14	−0.22	−0.38	−0.61
Mismatch=3 dB	−0.68	−0.83	−1.05	−1.31
DS+BC-SCPF(Multi-Rank)				
No mismatch	0.69	0.64	0.57	0.49
Mismatch=1 dB	0.69	0.64	0.57	0.49
Mismatch=2 dB	0.68	0.64	0.57	0.49
Mismatch=3 dB	0.64	0.60	0.52	0.42

Table 4
PESQ score improvement with different number of microphones (with speech interference, input SNR=5 db).

Number of microphones →	PESQ score improvement			
Algorithm ↓	4	6	8	12
DS+McCowan	0.20	0.26	0.36	0.45
DS+BC-SCPF(Multi-Rank)	0.51	0.60	0.69	0.76

insufficient spatial sampling. Compared to DS beamformer, the SD beamformer forms a null at the speech interference thus it provides better noise suppression and speech quality. Next, consider the post-filters. The performances of the Zelinski post-filter and SCPF are similar in this case. Compared to the Zelinski post-filter, the McCowan post-filter can be observed that it has a little bit better noise suppression at low frequency bands with the information of noise coherence matrix. Finally, both the BC-SCPF with rank-1 and multi-rank $\mathbf{\Phi}_s(\omega)$ models give evident noise suppression compared to the McCowan post-filter.

5.2. Sensitivity to array imperfection

Typically, microphone mismatch can easily happen in the implementation. Here, microphone #8 is assumed to have 1–3 dB gain mismatches. In Tables 3 and 4, it is shown that the McCowan post-filter is very sensitive to array imperfections while the proposed BC-SCPF has little performance degradation. The results coincide with the inference mentioned in Section 4.5.

5.3. Number of microphones

The effect of number of microphones is investigated with the speech interference at 5 dB input SNR. The PESQ improvement increases with the number of microphones, where the improvement of the BC-SCPF is larger than the McCowan post-filter about 0.2–0.3. This shows the feasibility and superiority of the BC-SCPF with different number of microphones in the non-homogeneous field. For homogeneous noise fields, the performance difference may be small, as shown in Table 1.

6. Conclusion

This paper has presented a multi-channel post-filter based on the spatial coherence measure, and a bias compensated solution. The bias compensated solution gives the optimal Wiener filter theoretically, as the McCowan post-filter did. In the coherent noise field, the proposed BC-SCPF with the multi-rank signal model provides better speech quality than the McCowan post-filter since the sound field is not homogeneous due to air attenuation. As for the homogeneous fields such as incoherent and diffuse noise, the performance improvement is limited. Besides, the proposed BC-SCPF is less sensitive to the array uncertainties such as microphone mismatch.

Furthermore, the similarity between the coherence matrices of the desired sound field and the noise field can be merged into a single real-valued bias. Several noise level estimation skills can be adopted to estimate the bias. Compared to the estimation of the noise coherence function, the bias estimation has fewer variables to be estimated. Besides, the noise level estimation can be carried out in the presence of the desired signal, while the estimation of noise coherence function is carried out during noise-only period. Finally, it is relatively easy to describe the similarity between the multi-rank signal models and the noise field using the proposed post-filters. This provides a more flexible design for the real-world environments.

It is worth to note that the multi-rank signal model is helpful when the directivity of the array [1] is large enough and the impact of local scattering or wavefront fluctuation is obvious. Otherwise, the performance of the proposed BC-SCPF may be similar to the McCowan post-filter.

Appendix A. Relationship between the proposed SCPF and the Zelinski post-filter

To compare with the Zelinski post-filter, the special case of the proposed SCPF in (28) is used. Let us denote the PSD matrix of the pre-processed input vector as $\tilde{\mathbf{\Phi}}_x(\omega)$. Using (5), the SCPF can be rewritten as

$$\begin{aligned} G_{\mathrm{SCPF}}(\omega) &= \frac{\phi_s(\omega)\mathbf{a}_s^H(\omega)\mathbf{\Phi}_x(\omega)\mathbf{a}_s(\omega)}{\phi_s(\omega)||\mathbf{a}_s(\omega)||^2\cdot \mathrm{tr}(\mathbf{\Phi}_x(\omega))} \\ &= \frac{\mathbf{1}^T\tilde{\mathbf{\Phi}}_x(\omega)\mathbf{1}}{M\cdot \mathrm{tr}(\tilde{\mathbf{\Phi}}_x}(\omega)) \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{\phi}_{x_ix_j}(\omega)\Big/\left(M\cdot\sum_{i=1}^{M}\tilde{\phi}_{x_ix_i}(\omega)\right) \\ &= \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{\phi}_{x_ix_j}(\omega)\Big/\left(\frac{1}{M}\cdot\sum_{i=1}^{M}\tilde{\phi}_{x_ix_i}(\omega)\right) \end{aligned} \tag{48}$$

where $\mathbf{1}$ is the all-one vector and $\mathrm{tr}(\tilde{\mathbf{\Phi}}_x(\omega))=\mathrm{tr}(\tilde{\mathbf{\Phi}}_x(\omega))$ if all the magnitudes of the elements in $\mathbf{a}_s$ are equal to unity. $\tilde{\phi}_{x_ix_j}(\omega)$ is the cross-spectral density between the pre-processed inputs at the i-th and j-th microphones as used in (10). In this case, the SCPF can be interpreted as the ratio

between the average of total spectral densities and the average of auto-spectral densities. Since the Zelinski post-filter is the ratio between the average of cross-spectral densities and the average of auto-spectral densities (as in (12)), the relationship between the SCPF and the Zelinski post-filter can be easily derived by

$$G_{\mathrm{SCPF}}(\omega) = \frac{(1/M^2)\left(\underbrace{2\sum_{i=1}^{M}\sum_{j=1}^{M}\Re\{\tilde{\phi}_{x_ix_j}(\omega)\}}_{\text{off-diagonal}} + \underbrace{\sum_{i=1}^{M}\tilde{\phi}_{x_ix_i}(\omega)}_{\text{diagonal}}\right)}{\frac{1}{M}\cdot\sum_{i=1}^{M}\tilde{\phi}_{x_ix_i}}(\omega)$$
$$= G_{\mathrm{Zelinski}}(\omega) + \frac{1}{M}(1 - G_{\mathrm{Zelinski}}(\omega)) \tag{49}$$

Or it can be written as

$$G_{\mathrm{Zelinski}}(\omega) = \frac{G_{\mathrm{SCPF}}(\omega) - 1/M}{1 - 1/M} \tag{50}$$

Note that the covariance matrix $\tilde{\mathbf{\Phi}}_x(\omega)$ is Hermitian, hence $\tilde{\phi}_{x_ix_j}(\omega) + \tilde{\phi}_{x_jx_i}(\omega) = 2\Re\{\tilde{\phi}_{x_ix_j}(\omega)\}$.

Appendix B. Analysis of the proposed BC-SCPF

To compare with the McCowan post-filter, the assumptions of homogeneous sound fields and point source model are used. Assume the actual and estimated noise coherence matrices from the microphones are $\mathbf{\Gamma}_n(\omega)$ and $\hat{\mathbf{\Gamma}}_n(\omega)$. Then according to (40) and (43), we have

$$G_{\mathrm{BC-SCPF}}(\omega) = \left[\frac{M^2\xi(\omega) + \mathbf{a}_s^H(\omega)\mathbf{\Gamma}_n(\omega)\mathbf{a}_s(\omega)}{M^2(\xi(\omega)+1)} - \frac{\mathbf{a}_s^H(\omega)\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{a}_s(\omega)}{M^2}\right] \Big/ \left(1 - \frac{\mathbf{a}_s^H(\omega)\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{a}_s(\omega)}{M^2}\right)$$
$$= G_{\mathrm{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)} \times \frac{\mathbf{a}_s^H(\omega)\mathbf{\Gamma}_n(\omega)\mathbf{a}_s(\omega) - \mathbf{a}_s^H(\omega)\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{a}_s(\omega)}{(M^2 - \mathbf{a}_s^H(\omega)\hat{\mathbf{\Gamma}}_n}(\omega)\mathbf{a}_s(\omega)) \tag{51}$$

Using the time alignment expression in Appendix A, (51) can be rewritten as

$$G_{\mathrm{BC-SCPF}}(\omega) = G_{\mathrm{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)}\left[\frac{\mathbf{1}^T\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{1} - \mathbf{1}^T\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{1}}{(M^2 - \mathbf{1}^T\hat{\mathbf{\Gamma}}_n(\omega)\mathbf{1})}\right]$$
$$= G_{\mathrm{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)}\left[\frac{\sum_{i=1}^{M}\sum_{j=1}^{M}\tilde{\Gamma}_{n_in_j}(\omega) - \sum_{i=1}^{M}\sum_{j=1}^{M}\hat{\tilde{\Gamma}}_{n_in_j}(\omega)}{\left(M^2 - \sum_{i=1}^{M}\sum_{j=1}^{M}\hat{\tilde{\Gamma}}_{n_in_j}(\omega)\right)}\right]$$
$$= G_{\mathrm{Wiener}}(\omega) + \frac{\phi_s(\omega)}{\phi_s(\omega)+\phi_n(\omega)}\left[\frac{\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}(\tilde{\Gamma}_{n_in_j}(\omega) - \hat{\tilde{\Gamma}}_{n_in_j}(\omega))}{\sum_{i=1}^{M}\sum_{j=1}^{M}(1 - \hat{\tilde{\Gamma}}_{n_in_j}(\omega))}\right] \tag{52}$$

where $\tilde{\Gamma}_{n_in_j}(\omega)$ and $\hat{\tilde{\Gamma}}_{n_in_j}(\omega)$ are the actual and estimated noise coherence matrices of the aligned inputs.

References

[1] H.L. Van Trees, Optimum Array Processing, John Wiley & Sons, Inc., New York, 2002.
[2] H. Cox, R.M. Zeskind, M.M. Owen, Robust adaptive beamforming, IEEE Trans. Acoust. Speech Signal Process. 35 (1987) 1365–1375.
[3] S. Gannot, D. Burshtein, E. Weinstein, Signal enhancement using beamforming and nonstationarity with applications to speech, IEEE Trans. Signal Process. 49 (8) (2001) 1614–1626.
[4] Y.-H. Chen, C.-T. Chiang, Adaptive beamforming using the constrained Kalman filter, IEEE Trans. Antennas Propag. 41 (11) (1993) 1576–1580.
[5] K.U. Simmer, J. Bitzer, C. Marro, Post-filtering techniques, Microphone Arrays: Signal Process. Tech. Appl. 3 (2001) 39–60.
[6] R. Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Trans. Speech Audio Process. 9 (5) (2001) 504–512.
[7] I. Cohen, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging, IEEE Trans. Speech Audio Process. 11 (5) (2003) 466–475.
[8] S. Rangachari, P.C. Loizou, A noise-estimation algorithm for highly non-stationary environments, Speech Commun. 48 (2) (2006) 220–231.
[9] I. Cohen, Multichannel post-filtering in nonstationary noise environments, IEEE Trans. Signal Process. 52 (5) (2004) 1149–1160.
[10] O. Yilmaz, S. Rickard, Blind separation of speech mixtures via time–frequency masking, IEEE Trans. Signal Process. 52 (7) (2004) 1830–1847.
[11] N. Roman, D. Wang, G.J. Brown, Speech segregation based on sound localization, in: Proceedings of the IEEE Conference on Neural Networks, vol. 4, 2001, pp. 2861–2866.
[12] P. Aarabi, S. Guangji, Phase-based dual-microphone robust speech enhancement, IEEE Trans. Syst. Man Cybern., Part B: Cybern. 34 (4) (2004) 1763–1773.
[13] S. Guangji, P. Aarabi, H. Jiang, Phase-based dual-microphone speech enhancement using a prior speech model, IEEE Trans. Audio Speech Lang. Process. 15 (1) (2007) 109–118.
[14] P. Aarabi, Phase-Based Speech Processing, World Scientific Publishing, Singapore, 2006.
[15] R. Zelinski, A microphone array with adaptive post-filtering for noise reduction in reverberant rooms, in: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, vol. 5, 11–14 April 1988, pp. 2578–2581.
[16] I.A. McCowan, H. Bourlard, Microphone array post-filter based on noise field coherence, IEEE Trans. Speech Audio Process. 11 (6) (2003) 709–716.
[17] R.L. Bouquin-Jeannes, A.A. Azirani, G. Faucon, Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator, IEEE Trans. Speech Audio Process. 5 (5) (1997) 484–487.
[18] S. Valaee, B. Champagne, P. Kabal, Parametric localization of distributed sources, IEEE Trans. Signal Process. 43 (9) (1995) 2144–2153.
[19] A. Zoubir, Y. Wang, P. Charge, Efficient subspace-based estimator for localization of multiple incoherently distributed sources, IEEE Trans. Signal Process. 56 (2) (2008) 532–542.
[20] A. Zoubir, Y. Wang, Robust generalised Capon algorithm for estimating the angular parameters of multiple incoherently distributed sources, IET Signal Process. 2 (2) (2008) 163–168.
[21] H. Ge, I.P. Kirsteins, Multi-rank processing for passive ranging in underwater acoustic environments subject to spatial coherence loss, in: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, 22–27 May 2011, pp. 2692–2695.
[22] S. Shahbazpanahi, A.B. Gershman, Z.Q. Luo, K.M. Wong, Robust adaptive beamforming for general-rank signal models, IEEE Trans. Signal Process. 51 (9) (2003) 2257–2269.
[23] A. Pezeshki, B.D Van Veen, L.L. Scharf, H. Cox, M.L. Nordenvaad, Eigenvalue beamforming using a multirank MVDR beamformer and subspace selection, IEEE Trans. Signal Process. 56 (5) (2008) 1954–1967.
[24] L. Zhang, W. Liu, Robust forward backward based beamformer for a general-rank signal model with real-valued implementation, Signal Process. 92 (1) (2012) 163–169.
[25] A.B. Gershman, N.D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, B. Ottersten, Convex optimization-based beamforming, IEEE Signal Process. Mag. 27 (3) (2010) 62–75.
[26] Z.Q. Luo, W.K. Ma, A.M.C. So, Y. Ye, S. Zhang, Semidefinite relaxation of quadratic optimization problems, IEEE Signal Process. Mag. 27 (3) (2010) 20–34.
[27] X.M. Yang, X.Q. Yang, K.L. Teo, A matrix trace inequality, J. Math. Anal. Appl. 263 (1) (2001) 327–331.
[28] J. Bitzer, K.U. Simmer, K.-D. Kammeyer, Multimicrophone noise reduction techniques for hands-free speech recognition – a comparative study, Robust Methods Speech Recognit. Advers. Cond. (ROBUST-99) (1999) 171–174.

[29] E. Habets, Room Impulse Response (RIR) Generator, July 2006. Available from: ⟨http://home.tiscali.nl/ehabets/rir_generator.html.⟩.
[30] A. Varga, H.J.M. Steeneken, Assessment for automatic speech recognition: II. NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems, Speech Commun. 12 (3) (1993) 247–251.
[31] Perceptual Evaluation of Speech Quality (PESQ), and Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, ITU, ITU-T Rec., 2000, p. 862.
[32] T. Rohdenburg, V. Hohmann, B. Kollmeir, Objective perceptual quality measures for the evaluation of noise reduction schemes, in: Proceedings of the 9th International Workshop Acoustic, Echo and Noise Control, 2005, ITU-T Rec. G. 160, 2008, pp. 169–172.