# Microphone Array Post-filter for Target Speech Enhancement Without a Prior Information of Point Interferers

*Guanjun Li*[1,2], *Shan Liang*[1], *Shuai Nie*[1], *Wenju Liu*[1], *Zhanlei Yang*[3], *Longshuai Xiao*[3]

[1]NLPR, Institute of Automation, Chinese Academy of Sciences, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
[3]Huawei Technologies, China

{guanjun.li,sliang,shuai.nie,lwj}@nlpr.ia.ac.cn, {yangzhanlei1,xiaolongshuai}@huawei.com

## Abstract

The post-filter for microphone array speech enhancement can effectively suppress noise including point interferers. However, the suppression of point interferers relies on the accurate estimation of the number and directions of point interferers, which is a difficult task in practical situations. In this paper, we propose a post-filtering algorithm, which is independent of the number and directions of point interferers. Specifically, we assume that the point interferers are continuously distributed at each direction of the plane but the probability of the interferer occurring at each direction is different in order to calculate the spatial covariance matrix of the point interferers. Moreover, we assume that the noise is additive and uncorrelated with the target signal to obtain the power spectral densities (PSDs) of the target signal and noise. Finally, the proposed post-filter is calculated using the estimated PSDs. Experimental results prove that the proposed post-filtering algorithm is superior to the comparative algorithms in the scenarios where the number and directions of point interferers are not accurately estimated.

**Index Terms**: microphone array, post-filter, spatial covariance matrix, probabilistic model

## 1. Introduction

Microphone arrays play an important role in speech enhancement and robust speech recognition. The minimum variance distortionless response (MVDR) beamformer takes advantage of the spatial discriminability of the array, but its noise reduction is not sufficient because MVDR is not an optimal solution in the minimum mean square error (MMSE) sense. The optimal solution is called multichannel Wiener filter (MWF), which can be decomposed into an MVDR beamformer followed by a single-channel post-filter [1]. Adding a post-filter can significantly improve the noise reduction performance compared to using MVDR alone [2, 3, 4, 5, 6, 7].

The key to implementing the post-filter is to obtain reliable power spectral densities (PSDs) of the target signal and noise [8]. [2] estimates these PSDs under the assumption of a spatially-white noise field. [3] and [4] further extend the hypothesis to the diffuse noise field to improve the noise reduction performance. However, [2, 3, 4] neglect the effect of the point interferers on the post-filter estimation.

In recent years, many smart voice devices have often encountered point interferers, such as music and interfering speakers etc. Therefore, there is a demand to design a post-filter that can suppress point interferers to improve the quality of the target speech. To deal with this problem, many methods have been proposed, which can be divided into deep neural network (DNN)-based methods and signal processing-based methods. The DNN-based methods generally use DNN to estimate a mask as a post-filter [9, 10, 11]. However, the DNN-based methods exhibit limited robustness to unseen acoustic environments and the complex network structures may cause undesirable processing delay when running on low-power chips. The signal processing-based methods can overcome the shortcomings of the DNN-based methods by modeling the law of the signal. To model the point interferers, [5] use spatial clustering to obtain the PSD for each source. [6] estimate the PSDs in a beamspace. [7] extend [2, 3, 4] to model the spatial covariance matrix of the point interferers into the signal model. Although [5, 6, 7] can handle point interferers, they are not practical because they all require the number and directions of point interferers in advance, which is difficult to estimate accurately in practical scenarios.

In this paper, we extend [7] and propose a post-filter, which is independent of the number and directions of point interferers. More specifically, we assume that the point interferers are continuously distributed at each direction of the plane and emit the same power, but the probability of the interferer occurring at each direction is different. The occurrence probability of the point interferer near the target direction is low, while the occurrence probability of the point interferer far from the target direction is high. We utilize the notched distribution [12, 13] to describe this probability. From the above assumptions, we use the integral to obtain the interferers' spatial covariance matrix that only depends on the target direction. Further, we use the relationship among the spatial covariance matrices of different signals and utilize the least squares (LS) algorithm to obtain the PSDs of the target signal, diffuse noise and point interferers respectively. Experimental results show that the proposed post-filter shows great advantages when the number and directions of point interferers cannot be accurately estimated in practice.

## 2. Problem formulation

We consider a situation where a microphone array containing $M$ microphones captures the target signal in a noisy environment. The $M \times 1$ observation vector in the short-time Fourier transform (STFT) domain is given by

$$\mathbf{y}(t,f) = \mathbf{s}(t,f) + \mathbf{v}(t,f) + \mathbf{u}(t,f), \tag{1}$$

where $t$ and $f$ denote the time and frequency indices respectively, and $\mathbf{s}(t,f)$, $\mathbf{v}(t,f)$ and $\mathbf{u}(t,f)$ are the vectors of the target signal, point interferers and diffuse noise as received to the microphone array. In order to obtain an estimate of the target signal $\hat{s}(t,f)$, the MWF can be performed by applying a linear filter $\mathbf{w}_{\text{mwf}}(t,f)$ to the observation vector $\mathbf{y}(t,f)$, i.e.,

$$\hat{s}(t,f) = \mathbf{w}_{\text{mwf}}^{H}(t,f)\mathbf{y}(t,f), \tag{2}$$

http://dx.doi.org/10.21437/Interspeech.2020-1351

where superscript $\hat{}$ denotes an estimated value and $(\cdot)^H$ denotes the conjugate transposition. In Eq. (2), $\mathbf{w}_{\text{mwf}}(t,f)$ can be factorized into an MVDR beamformer $\mathbf{w}_{\text{mvdr}}(t,f)$ followed by a single-channel post-filter $w_{\text{post}}(t,f)$ [1]:

$$\mathbf{w}_{\text{mwf}}(t,f) = \mathbf{w}_{\text{mvdr}}(t,f)\underbrace{\frac{\sigma_s^2(t,f)}{\sigma_s^2(t,f)+\sigma_\psi^2(t,f)}}_{w_{\text{post}}(t,f)}, \tag{3}$$

where $\sigma_s^2(t,f)$ is the target signal PSD and $\sigma_\psi^2(t,f)$ is the noise PSD at the output of the beamformer, defined as

$$\sigma_\psi^2(t,f) = \mathbf{w}_{\text{mvdr}}^H(t,f)\left[\mathbf{R}_v(t,f)+\mathbf{R}_u(t,f)\right]\mathbf{w}_{\text{mvdr}}(t,f), \tag{4}$$

where $\mathbf{R}_v(t,f) = E\{\mathbf{v}(t,f)\mathbf{v}(t,f)^H\}$ and $\mathbf{R}_u(t,f) = E\{\mathbf{u}(t,f)\mathbf{u}(t,f)^H\}$ are the spatial covariance matrices of the point interferers and diffuse noise respectively, and $E\{\cdot\}$ denotes the mathematical expectation. If there are $N$ discrete independent point interferers, $\mathbf{R}_v(t,f)$ can be written as

$$\mathbf{R}_v(t,f) = \sum_{n=1}^{N}\mathbf{R}_{v_n}(t,f), \tag{5}$$

$$= \sum_{n=1}^{N}\sigma_{v_n}^2(t,f)\mathbf{i}_n(f)\mathbf{i}_n^H(f), \tag{6}$$

where $\mathbf{R}_{v_n}(t,f)$, $\sigma_{v_n}^2(t,f)$ and $\mathbf{i}_n(f)$ are the spatial covariance matrix, PSD and steer vector of the $n$-th point interferer.

According to Eq. (3)– Eq. (6), in order to construct the post-filter $w_{\text{post}}(t,f)$, the number of the point interferers $N$ is required in advance. Besides, the directions of the point interferers are also required when we construct the steer vector in Eq. (6) [7]. However, estimating the number and directions of point interferers accurately in real-world scenarios remains a challenging task.

## 3. Proposed post-filter

In this section, we introduce the proposed post-filtering algorithm, which is independent of the number and directions of point interferers. We first model the spatial covariance matrix $\mathbf{R}_v(t,f)$ using a probabilistic model and then give a method for estimating the PSDs required to construct the post-filter.

### 3.1. Probabilistic model for $\mathbf{R}_v(t,f)$

We consider a situation where the number and directions of point interferers are unknown. In this case, we make the following assumptions:

(1) The point interferers are continuously distributed at each direction of the plane and they emit the same power $\sigma_v^2(t,f)$. The occurrence probability of the point interferer at direction $\theta$ is $p_v(\theta,\theta_s)$, where $\theta_s$ is the target signal's direction.

(2) When $\theta$ tends to $\theta_s$, $p_v(\theta,\theta_s)$ tends to zero. When $\theta$ gradually away from $\theta_s$, $p_v(\theta,\theta_s)$ gradually increases[1].

Although the point interferers may not satisfy the above assumptions in practice, we experimentally found that these assumptions can help the post-filter to effectively suppress the point interferers. To satisfy the assumption (2), we use the notched distribution [12, 13] to describe $p_v(\theta,\theta_s)$,

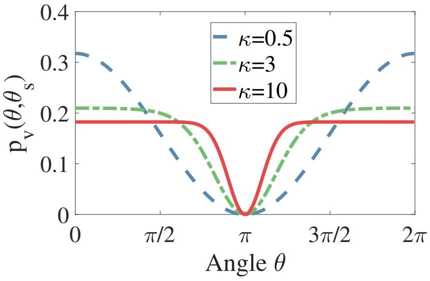

Figure 1: *The notched distribution for a given direction $\theta_s = \pi$.*

$$p_v(\theta,\theta_s) = \frac{e^{\kappa} - e^{\kappa\cos(\theta-\theta_s)}}{2\pi\left[e^{\kappa} - I_0(\kappa)\right]}, \tag{7}$$

where $I_0$ is the modified Bessel function of the first kind and $\kappa$ controls the shape of the distribution. The notched distribution for different values of $\kappa$ is illustrated in Figure 1.

According to assumption (1) and (2), Eq. (6) can be extended into a continuous form by integrating over a 2D plane:

$$\mathbf{R}_v(t,f) = \sigma_v^2(t,f)\int_0^{2\pi} p_v(\theta,\theta_s)\mathbf{i}_\theta(f)\mathbf{i}_\theta^H(f)\,d\theta \tag{8}$$

$$= \sigma_v^2(t,f)\mathbf{R}_i(f), \tag{9}$$

where $\mathbf{R}_i(f)$ can be viewed as the spatial correlation matrix of all point interferers and $\mathbf{i}_\theta(f)$ is the steer vector corresponding to the point interferer from direction $\theta$. Under the far-field assumption [14], $\mathbf{i}_\theta(f)$ can be expressed as

$$\mathbf{i}_\theta(f) = \left[1, e^{jA_f(\mathbf{r}_2-\mathbf{r}_1)^T\mathbf{q}(\theta)}, \ldots, e^{jA_f(\mathbf{r}_M-\mathbf{r}_1)^T\mathbf{q}(\theta)}\right]^T, \tag{10}$$

where $(\cdot)^T$ denotes the transposition, $j = \sqrt{-1}$ is the imaginary unit, $A_f = \frac{2\pi k_f}{c}$, $k_f$ is the frequency in Hz corresponding to the frequency index $f$, $\mathbf{r}_m$ is the 2D location vector of the $m$-th microphone and $\mathbf{q}(\theta) = [\cos\theta, \sin\theta]^T$ is a direction vector.

Denoting the $(a,b)$-th element of $\mathbf{R}_i(f)$ as $[\mathbf{R}_i(f)]_{ab}$, using Eq. (10), $\mathbf{R}_i(f)$ in Eq. (9) can be rewritten as

$$[\mathbf{R}_i(f)]_{ab} = \int_0^{2\pi} p_v(\theta,\theta_s)e^{jA_f(\mathbf{r}_a-\mathbf{r}_b)^T\mathbf{q}(\theta)}\,d\theta, \tag{11}$$

where $\mathbf{r}_a$ and $\mathbf{r}_b$ are the 2D location vectors of the $a$-th microphone and the $b$-th microphone respectively. Substituting Eq. (7) into Eq. (11), we can obtain

$$[\mathbf{R}_i(f)]_{ab} = \frac{\frac{e^{\kappa}}{2\pi}\int_0^{2\pi} e^{jA_f r_{ab}\cos(\theta-\theta_{ab})}\,d\theta}{e^{\kappa}-I_0(\kappa)} - \frac{I_0(\kappa)\int_0^{2\pi} p_s(\theta,\theta_s)e^{jA_f(\mathbf{r}_a-\mathbf{r}_b)^T\mathbf{q}(\theta)}\,d\theta}{e^{\kappa}-I_0(\kappa)}, \tag{12}$$

where $r_{ab} = |\mathbf{r}_a - \mathbf{r}_b|$, $\theta_{ab} \in [-\pi,\pi]$ is the angle between the vector $(\mathbf{r}_a - \mathbf{r}_b)$ and the positive direction of the $x$-axis and $p_s(\theta,\theta_s)$ is the von Mises distribution [15] defined as $p_s(\theta,\theta_s) = \frac{e^{\kappa\cos(\theta-\theta_s)}}{2\pi I_0(\kappa)}$. After some algebraic manipulation, we can get the analytical solution of Eq. (12), i.e.,

$$[\mathbf{R}_i(f)]_{ab} = \frac{e^{\kappa}J_0(A_f r_{ab}) - J_0(z)}{e^{\kappa}-I_0(\kappa)}, \tag{13}$$

[1] Here we only consider the situation where the point interferers and the target are not close to each other. It is difficult to separate them by using direction cues if they come from the same direction.

where $z = \sqrt{A_f^2 r_{ab}^2 - \kappa^2 - 2j\kappa A_f r_{ab}\cos(\theta_{ab} - \theta_s)}$ is a complex number and $J_0$ is the zero-order Bessel function of the first kind. We recommend readers to refer to [16] for the detailed derivation of the integral on the numerator of the second term of Eq. (12).

As can be seen from Eq. (9) and Eq. (13), the spatial covariance matrix $\mathbf{R}_v(t, f)$ no longer depends on the number and directions of point interferers, but only the target signal's direction $\theta_s$. We assume that the target signal's direction is known. This assumption can be realized by using the surveillance video.

### 3.2. Estimation of PSDs

In this subsection, we will utilize $\mathbf{R}_i(t, f)$ derived in Eq. (13) to estimate the PSDs required for $w_{\text{post}}(t, f)$ in Eq. (3).

Assuming that different signals in Eq. (1) are mutually uncorrelated, the spatial covariance matrices are related as

$$\mathbf{R}_y(t, f) = \mathbf{R}_s(t, f) + \mathbf{R}_v(t, f) + \mathbf{R}_u(t, f), \tag{14}$$

where $\mathbf{R}_s(t, f) = E\{\mathbf{s}(t, f)\mathbf{s}(t, f)^H\}$ and $\mathbf{R}_y(t, f) = E\{\mathbf{y}(t, f)\mathbf{y}(t, f)^H\}$. In practice, $\mathbf{R}_y(t, f)$ can be obtained using recursive averaging. Specifically,

$$\mathbf{R}_y(t, f) = \alpha\mathbf{R}_y(t-1, f) + (1-\alpha)\mathbf{y}(t, f)\mathbf{y}^H(t, f), \tag{15}$$

where $0 < \alpha < 1$ is a smoothing factor.

Next, we first decompose the spatial covariance matrix of the target signal as follows:

$$\mathbf{R}_s(t, f) = \sigma_s^2(t, f)\mathbf{R}_h(f), \tag{16}$$

where

$$\mathbf{R}_h(f) = \mathbf{h}(f)\mathbf{h}^H(f) \tag{17}$$

is the spatial correlation matrix of the target signal and $\mathbf{h}(f)$ is the steer vector of the target signal,

$$\mathbf{h}(f) = \left[1, e^{jA_f(\mathbf{r}_2-\mathbf{r}_1)^T\mathbf{q}(\theta_s)}, \ldots, e^{jA_f(\mathbf{r}_M-\mathbf{r}_1)^T\mathbf{q}(\theta_s)}\right]^T. \tag{18}$$

We then decompose the spatial covariance matrix of the diffuse noise as follows:

$$\mathbf{R}_u(t, f) = \sigma_u^2(t, f)\mathbf{\Gamma}_u(f), \tag{19}$$

where $\sigma_u^2(t, f)$ and $\mathbf{\Gamma}_u(f)$ are the PSD and spatial correlation matrix of the diffuse noise. The $(a, b)$-th element of $\mathbf{\Gamma}_u(f)$ can be calculated as [17]:

$$[\mathbf{\Gamma}_u(f)]_{ab} = \frac{\sin(A_f r_{ab})}{A_f r_{ab}}. \tag{20}$$

Substituting Eq. (9), Eq. (16) and Eq. (19) into Eq. (14), we can obtain

$$\mathbf{R}_y(t, f) = \sigma_s^2(t, f)\mathbf{R}_h(f) + \sigma_v^2(t, f)\mathbf{R}_i(f) + \sigma_u^2(t, f)\mathbf{\Gamma}_u(f). \tag{21}$$

Considering that $\mathbf{R}_y(t, f)$, $\mathbf{R}_h(f)$, $\mathbf{R}_i(f)$ and $\mathbf{\Gamma}_u(f)$ are Hermitian, we can rewrite Eq. (21) into the following form:

$$\boldsymbol{\phi}(t, f) = \mathbf{\Omega}(f)\boldsymbol{\chi}(t, f), \tag{22}$$

where

$$\boldsymbol{\phi}(t, f) = \text{triv}\{\mathbf{R}_y(t, f)\}, \tag{23}$$

$$\mathbf{\Omega}(f) = [\text{triv}\{\mathbf{R}_h(f)\}, \text{triv}\{\mathbf{R}_i(f)\}, \text{triv}\{\mathbf{\Gamma}_u(f)\}], \tag{24}$$

$$\boldsymbol{\chi}(t, f) = \left[\sigma_s^2(t, f), \sigma_v^2(t, f), \sigma_u^2(t, f)\right]^T, \tag{25}$$

where triv$\{\mathbf{R}\}$ represents a $\frac{M(M+1)}{2} \times 1$ column vector formed by stacking the upper triangular parts of an $M \times M$ matrix $\mathbf{R}$. The LS solution for $\boldsymbol{\chi}(t, f)$ is given by

$$\hat{\boldsymbol{\chi}}(t, f) = \Re\{\mathbf{\Omega}^\dagger(f)\boldsymbol{\phi}(t, f)\}, \tag{26}$$

where superscript † denotes pseudo-inverse and $\Re\{.\}$ denotes real part. In Eq. (26), we use $\Re\{.\}$ to avoid complex results in a heuristic manner, and we experimentally found that it outperformed the absolute (ABS) operation. Since PSDs can only be positive-valued, we lower-bound the PSD estimates by 0, i.e., $\sigma_s^2(t, f), \sigma_v^2(t, f), \sigma_u^2(t, f) \geq 0$.

**Algorithm 1** Proposed post-filtering algorithm per subband.

```
Input: $\mathbf{y}(t, f)$ and $\mathbf{w}_{\text{mvdr}}(t, f)$, $t = 1, 2, 3, \ldots$
Compute $\mathbf{R}_i(f)$, $\mathbf{R}_h(f)$ and $\mathbf{\Gamma}_u(f)$ using (13), (17) and (20)
$\mathbf{\Omega}(f) \leftarrow (\mathbf{R}_h(f), \mathbf{R}_i(f), \mathbf{\Gamma}_u(f))$ using (24)
for $t = 1, 2, 3, \cdots$ do
    Compute $\mathbf{R}_y(t, f)$ using (15)
    $\boldsymbol{\phi}(t, f) \leftarrow \mathbf{R}_y(t, f)$ using (23)
    Obtain $\sigma_s^2(t, f)$,$\sigma_v^2(t, f)$ and $\sigma_u^2(t, f)$ using (26)
    Obtain $\mathbf{R}_v(t, f)$ and $\mathbf{R}_u(t, f)$ using (9) and (19)
    Obtain the post-filer $w_{\text{post}}(t, f)$ using (3) and (4)
    $\mathbf{w}_{\text{mwf}}(t, f) = \mathbf{w}_{\text{mvdr}}(t, f)w_{\text{post}}(t, f)$
    Estimate the target signal using (2)
end for
```

Finally, we obtain the PSDs required to construct $w_{\text{post}}(t, f)$ in Eq. (3) through $\hat{\boldsymbol{\chi}}(t, f)$. The proposed post-filtering algorithm is listed in Algorithm 1. Although Eq. (26) in Algorithm 1 requires a computationally intensive pseudo-inverse operation, $\mathbf{\Omega}^\dagger(f)$ in Eq. (26) is independent of the time $t$. Therefore, $\mathbf{\Omega}^\dagger(f)$ for different target directions can be pre-calculated and loaded into the system, which makes the proposed algorithm suitable for real-time processing.

## 4. Experimental results

### 4.1. Data and settings

To evaluate the proposed algorithm, we convolved the room impulse responses (RIRs) with the target source and point interferers to generate multi-channel mixed signals. The target source and point interferers were randomly selected from the TIMIT dataset [18]. Their average length was 25 s and the sampling rate was 16 kHz. We used the real RIRs measured in the speech & acoustic lab of the Faculty of Engineering at Bar-Ilan University [19]. The reverberation time $T_{60} \approx 360$ ms and the target source and point interferers were 2 m from the microphone array. The array contained 4 microphones with a distance of 3 cm between each microphone. We generated 6 types of mixed signals with a target source at 30°. The first 3 types of the mixed signals contained only one point interferer at 90° with the input signal-to-interference ratio (SIR) between 0 and 10 dB. The last 3 types of the mixed signals contained two point interferers at 90° and 150° respectively with the input SIR between 0 and 10 dB. These 6 types of the mixed signals all contained diffuse noise with an input signal-to-diffuse-noise ratio of 15 dB [20].

The STFT frame size was 32 ms with 50 % overlap. We experimentally set $\kappa = 20$ and $\alpha = 0.72$. In order to avoid signal cancellation problem in the adaptive implementation of MVDR beamformer, we chose to use the super-directive (SD)

Table 1: *Results ($\Delta$PESQ / $\Delta$oSINR) for the case of one point interferer with different input SIRs.*

| Method | 0 dB | 5 dB | 10 dB |
|---|---|---|---|
| SD + HPF | 0.44 / 4.74 | 0.37 / 3.82 | 0.35 / 2.52 |
| SD | 0.16 / 2.01 | 0.14 / 1.96 | 0.13 / 1.17 |
| SD + LPF | 0.31 / 3.33 | 0.29 / 2.42 | 0.28 / 1.58 |
| SD + Proposed | **0.43 / 4.55** | **0.34 / 3.09** | **0.32 / 1.82** |

Table 2: *Results ($\Delta$PESQ / $\Delta$oSINR) for the case of two point interferers with different input SIRs.*

| Method | 0 dB | 5 dB | 10 dB |
|---|---|---|---|
| SD + HPF | 0.51 / 4.27 | 0.44 / 4.12 | 0.43 / 3.09 |
| SD | 0.24 / 1.88 | 0.17 / 2.15 | 0.17 / 1.77 |
| SD + LPF | 0.30 / 3.17 | 0.27 / 2.90 | 0.29 / 1.99 |
| SD + HPF_1 | 0.48 / 4.13 | 0.33 / 3.66 | 0.32 / 2.05 |
| SD + Proposed | **0.53 / 4.55** | **0.46 / 4.21** | **0.36 / 2.88** |

beamformer [21] for front-end processing which uses the time-invariant diffuse noise field [17]. We implemented two comparative post-filters: one is Leukimmiatis's post-filter [4] (LPF), which only considers diffuse noise, and the other one is Huang's post-filter [7] (HPF), which depends on the oracle number and directions of point interferers. HPF uses Eq. (6) instead of Eq. (8) to obtain PSDs. We did not implement the algorithm proposed in [5] because it uses the expectation-maximization (EM) algorithm, which is not suitable for real-time processing compared to the above algorithms. For performance evaluation, we focused on two objective metrics: the perceptual evaluation of speech quality (PESQ) [17] and the output signal-to-interference-and-noise ratio (oSINR) [22].

### 4.2. Results

Table 1 shows the performance of the algorithms as a function of the input SIRs in a scenario with only one point interferer. For better visualization, only the improvements of the objective measures with respect to the unprocessed microphone signal (denoted by the $\Delta$ values) were illustrated. It can be seen from Table 1 that the proposed algorithm performs better than LPF and has a small gap with HPF, indicating that the proposed algorithm can effectively suppress point interferers although it does not utilize the number and directions of point interferers.

Table 2 shows the performance of the algorithms as a function of the input SIRs in a scenario with two point interferers. In order to simulate the case where the number of point interferers is not estimated accurately, we added a post-filter HPF_1. It is a modified HPF, which only considers the point interferer at 150°. Comparing HPF and HPF_1 in Table 2, we can see that the performance of HPF begins to decline when the number of point interferers is not estimated accurately. However, the proposed algorithm is better than HPF_1, and even outperforms HPF in low input SIR scenarios (0 dB and 5 dB). This may be because the distribution of point interferers is more in line with the assumptions of the proposed algorithm when the number of point interferers increases.

Figure 2 shows an example of post-filtering with two point interferers. Compared Figure 2(c) with Figure 2(d), especially the red box, we can see that the signal processed by HPF will have obvious residual interfering signal when the number of point interferers is not estimated accurately. However, the proposed algorithm can effectively suppress the point interferers without estimating the number of point interferers.

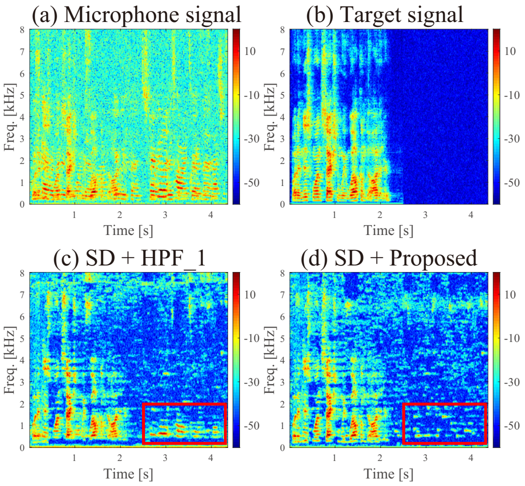

Figure 2: *An example of post-filtering for two point interferers with SIR=0 dB. HPF_1 was used to simulate the case where the number of point interferers is not accurately estimated.*

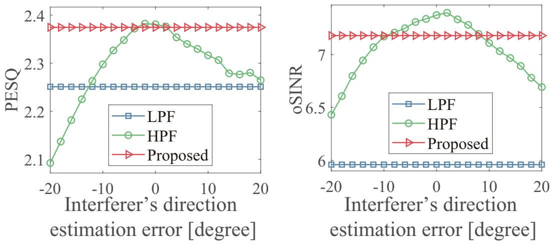

Figure 3: *PESQ (left) and oSINR (right) as a function of the interferer's direction estimation error.*

Next, we used the mixed signals of SIR = 0 dB with only one point interferer and considered a scenario where the direction of the point interferer is not estimated accurately. We controlled the direction estimation error of the point interferer between $-20°$ and $20°$. As can be seen from Figure 3, the performance of HPF begins to decline and even produces lower PESQ than LPF when the estimation of the interferer's direction is biased. Figure 2 and Figure 3 indicate that HPF is sensitive to estimation errors of the number and directions of point interferers. However, the proposed algorithm can effectively suppress interference while being independent of the number and directions of point interferers. This makes it show more potentialities in real applications.

## 5. Conclusions

In this paper, we proposed a microphone array post-filter that is independent of the number and directions of point interferers. The spatial covariance matrix of the point interferers that are required to obtain the post-filter are calculated by using a probabilistic model, which only demands the target signal's direction. The proposed post-filter shows more practical potentialities in the scenarios where the number and directions of point interferers cannot be accurately estimated.

## 6. Acknowledgments

This work was supported in part by the National Key R&D Plan of China (No. 2016YFB1001404) and the China National Nature Science Foundation (No. 61971419, No. 61573357, No. 61503382). We would like to thank Prof. Weimin Zhang from Tsinghua University for helpful writing suggestions.

# 7. References

[1] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone arrays*. Springer, 2001, pp. 39–60.

[2] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*. IEEE, 1988, pp. 2578–2581.

[3] I. A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 709–716, 2003.

[4] S. Leukimmiatis, D. Dimitriadis, and P. Maragos, "An optimum microphone array post-filter for speech applications," in *Ninth International Conference on Spoken Language Processing*, 2006.

[5] N. Q. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1830–1840, 2010.

[6] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, and Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 6, pp. 1240–1250, 2013.

[7] Y. A. Huang, A. Luebs, J. Skoglund, and W. B. Kleijn, "Globally optimized least-squares post-filtering for microphone array speech enhancement," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 380–384.

[8] W. Jin, M. J. Taghizadeh, K. Chen, and W. Xiao, "Multi-channel noise reduction for hands-free voice communication on mobile phones," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 506–510.

[9] P. Pertilä and J. Nikunen, "Microphone array post-filtering using supervised machine learning for speech enhancement," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[10] Y. Xu, J. Du, Z. Huang, L.-R. Dai, and C.-H. Lee, "Multi-objective learning and mask-based post-processing for deep neural network based speech enhancement," *arXiv preprint arXiv:1703.07172*, 2017.

[11] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702–1726, 2018.

[12] M. Taseska and E. A. Habets, "Minimum bayes risk signal detection for speech enhancement based on a narrowband doa model," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 539–543.

[13] ——, "Doa-informed source extraction in the presence of competing talkers and background noise," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, p. 60, 2017.

[14] J. Flanagan, J. Johnston, R. Zahn, and G. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *The Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508–1518, 1985.

[15] P. D. Teal, T. D. Abhayapala, and R. A. Kennedy, "Spatial correlation for general distributions of scatterers," *IEEE signal processing letters*, vol. 9, no. 10, pp. 305–308, 2002.

[16] C. A. Anderson, P. D. Teal, and M. A. Poletti, "Spatially robust far-field beamforming using the von mises (-fisher) distribution," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2189–2197, 2015.

[17] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. T. Jr, "Measurement of correlation coefficients in reverberant sound fields," *Jasa*, vol. 52, no. 6, pp. 1072–1077, 1955.

[18] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, 1993.

[19] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on*. IEEE, 2014, pp. 313–317.

[20] E. A. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *The Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3464–3470, 2007.

[21] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, 2007.

[22] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2008.