Experimental study of robust acoustic beamforming for speech acquisition in reverberant and noisy environments

Zhao, Yingke; Jensen, Jesper Rindom; Jensen, Tobias; Chen, Jingdong; Christensen, Mads Græsbøll

Published in:
Applied Acoustics

DOI (link to publication from Publisher):
10.1016/j.apacoust.2020.107531

Publication date:
2020

Document Version
Early version, also known as pre-print

Experimental Study of Robust Acoustic Beamforming for Reverberant and Noisy Environments

Yingke Zhao[a,*], Jesper Rindom Jensen[b], Tobias Lindstrøm Jensen[c], Jingdong Chen[a], Mads Græsbøll Christensen[b]

[a]*Center of Intelligent Acoustics and Immersive Communications and School of Marine Science and Technology, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China.*
[b]*Audio Analysis Lab, CREATE, Aalborg University, Aalborg 9000, Denmark*
[c]*Faculty of Engineering and Science, Aalborg University, Aalborg 9220, Denmark*

Abstract

The performance of adaptive beamformers suffers from significant degradation in the presence of steering vector errors, statistics estimation errors, and reverberation. To address this issue, narrowband robust beamforming methods are extended to the wideband case for processing speech signals in this paper. We study two types of methods. In the first type, the robustness of the beamformer is improved by adding a norm constraint and/or a steering vector uncertainty constraint to the optimization problem. It is worth noticing that the norm constraint also helps to control the sidelobes of the beampattern, which makes the beamformers able to suppress the interferences and the reflections of the desired signal, thereby improving the robustness of the beamformers against reverberation. Another type of methods is developed by using the spatial smoothing technique. The noise covariance matrix is implicitly estimated first by subtracting a delay-and-sum beamforming estimate of the desired signal covariance matrix from the observed signal covariance matrix, which helps improve the robustness of the beamformer. Experiments are performed to investigate the performance of the developed robust beamformers in acoustic environments. The results show that the robust beamformers outperform the non-robust counterparts in terms of: 1) robust performance in reverberation and different noise levels; 2) resilience against steering vector and noisy signal covariance matrix estimation errors; and 3) better predicted speech quality and intelligibility measured using the output SINR, PESQ, and STOI scores under reverberant conditions.

Keywords: Microphone array, Capon beamforming, steering vector error, reverberation, robust beamforming, amplitude-and-phase estimation beamforming

1. Introduction

A beamformer, which is basically an optimal spatial filter, can be applied in many acoustic applications to acquire high-fidelity signals of interest and eliminate interference and noise. It is the core part of an array system for signal processing [1–5] and its performance plays an important role on the overall performance of the signal processing system. However, steering vector estimation errors may cause the desired signal to be distorted, which makes the robustness of the beamformers against steering vector estimation errors an essential consideration. Beamformers including delay and sum beamformer (DSB) and minimum power distortionless response (MPDR) beamformer (also known as the Capon beamformer) [6], are sensitive to steering vector estimation errors, which introduce severe speech distortion. The minimum variance distortionless response (MVDR) beamformer on the other hand, is robust against steering vector estimation errors. However, the MVDR beamformer is difficult to implement in practice, since the estimation of the noise covariance matrix is not trivial [7–9]. Herein, we focus on the beamformers based on the observed signal statistics, and seek to increase their robustness against steering vector estimation errors by generalizing techniques from robust narrowband beamforming to broadband speech scenarios.

In practice, the steering vector mismatch may cause the desired signal to be canceled as if it is an interference. Furthermore, the observed signal covariance matrix estimation errors also challenges the ability of the beamformer to maintain robust performance. It is interesting to note that the difference between the estimated covariance matrix and the theoretical one can be viewed as steering vector estimation errors [10]. The reverberation caused by the sound reflections in the acoustic environment makes accurate DOA estimation hard to obtain which increases the likelihood of the steering errors in practice. Moreover, under reverberant condition, the signal model mismatch appears when the free field steering vector is used to model the signal propagation. Additionally, coherent reflections of the desired signal may lead to cancellation of the desired signal. A pos-

*This work was supported in part by the National Natural Science Foundation of China (NSFC) "Distinguished Young Scientists Fund" under Grant No. 61425005 and the NSFC-ISF (Israel Science Foundation) joint research program under Grant No. 61761146001. The work of Y. Zhao was supported in part by the China Scholarship Council.

*Corresponding author

Email addresses: yingkezhao_nwpu@163.com (Yingke Zhao), jrj@create.aau.dk (Jesper Rindom Jensen), tlj@its.aau.dk (Tobias Lindstrøm Jensen), jingdongchen@ieee.org (Jingdong Chen), mgc@create.aau.dk (Mads Græsbøll Christensen)

sible way to handle the reverberation is to utilize the transfer function when developing beamformers. However, the estimation of room impulse response (RIR) is a difficult task [11].

Robust beamforming [10, 12–18] has been widely used in narrow band signal processing. In [12, 13], in order to control the amplification of spatially white noise of the beamformer, a norm constraint is added to the Capon beamformer which gives the norm constraint Capon beamformer (NCCB). The norm constraint turns out to help improve the robustness of the beamformer against the steering vector estimation error. Another way to deal with the steering vector inaccuracy is to apply a steering vector uncertainty constraint to the optimization problem, which leads to the beamforming method called the robust Capon beamformer (RCB) [14]. By taking into account the norm constraint as well as the steering vector uncertainty constraint, the doubly constrained robust Capon beamformer (DCRCB) can be derived [10]. An alternative way is to apply the spatial smoothing technique to improve the robustness of the beamformer, which leads to a promising method called the amplitude-and-phase-estimation beamformer (APES) [19–23]. The APES beamformer can be interpreted as first estimating the noise covariance matrix by using the delay and sum method within the subarray. The robustness of the beamformer is improved by utilizing the noise statistics in forming the spatial filter. In order to apply these methods to wideband signals such as speech signals, the signal can be transformed into the frequency domain, after which the narrowband robust methods can be applied in each subband independently [24–27]. The broadband filtered signal is then synthesized from the outputs of subband filters [26]. For example, in [24] the RCB was studied for processing speech signals, but only two channels and a simple alphabetical task were considered in the simulations. In [28], the inequality-constrained minimum variance beamformer is developed by adding inequality constraints to the original MVDR optimization problem for acoustic signal processing. However, the noise covariance matrix need to be estimated, which limits the application of this method, since multichannel noise tracking is still an open problem.

In this paper, we apply beamformers using the observed signal statistics to process speech signals in reverberant and noisy environments, and seek to improve the robustness of the beamformer against the steering vector estimation errors. The observed signal statistics estimation errors and reverberation can be regarded as steering vector estimation errors, which makes the beamformers also robust against statistics estimation errors and reverberation. In our previous work [25], we evaluated the performance of the robust beamformers in speech signal processing by taking into account different noise conditions and small amounts of reverberation. The experimental results showed that the robust beamformers are promising in both improving the speech quality and speech intelligibility. Based on the former work in wideband robust beamforming, we will give more details about how to solve the optimization problems of the beamforming methods in this paper. Furthermore, the performance of the beamformers under different reverberation conditions is further studied. Robust beamformers is promising in dealing with reverberation, since the added norm constraint is able to lower the sidelobes of the beampattern which helps reduce the reflections of the signal and thus reduce cancellation of the desired signal.

The rest of this paper is organized as follows. Section 2 depicts the signal model and problem formulation. Section 3 reviews the standard approaches for beamforming. Section 4 defines and studies the robust beamformers including the robust Capon/MPDR type of methods and the APES method. Section 5.1 continues to analyze the approach by introducing some performance measures. Experimental results are then presented in Section 5 before the paper is concluded in Section 6.

2. Signal model and problem formulation

We consider to use a uniform linear array (ULA), which consists of M microphones, to capture an acoustic signal of interest in noisy environments. If the signal of interest impinges on the array from the farfield and there is no reverberation, the observation signal at the mth microphone is written in the time domain as

$$\begin{aligned} y_m(t) &= x_m(t) + v_m(t) \\ &= x(t-\tau_m) + v_m(t), \quad m = 1, 2, \cdots, M, \end{aligned} \tag{1}$$

where $y_m(t)$, $x_m(t)$ and $v_m(t)$ are, respectively, the observation signal, the speech signal, and the additive noise (including both non-stationary noise and stationary noise) from microphone m at time t, $\tau_m = (m-1)\tau_0$ is the time difference of arrival (TDOA), τ_0 is the TDOA between the second and first microphones, i.e., $\tau_0 = (\delta \cos\theta_{\mathrm{d}})/c$, with θ_{d} being the source incidence angle, δ denoting the spacing between neighboring microphones, and c being the speed of sound in air.

Speech signals are nonstationary and we generally process such signals in the short-time-Fourier-transform (STFT) domain. In such a domain, the signal model in (1) rewritten as:

$$\begin{aligned} Y_m(n,\omega) &= X_m(n,\omega) + V_m(n,\omega) \\ &= e^{-j(m-1)\omega\tau_0} X(n,\omega) + V_m(n,\omega), \end{aligned} \tag{2}$$

where n and ω denote, respectively, the frame and frequency indices. $Y_m(n,\omega)$, $X_m(n,\omega)$ and $V_m(n,\omega)$ are the STFT coefficients of $y_m(t)$, $x_m(t)$ and $v_m(t)$, respectively. $X(n,\omega)$ is the noise free speech signal at the reference (first) microphone. In case that there exists reverberation, the signal model in (2) should be modified as

$$Y_m(n,\omega) = D_m(n,\omega)X(n,\omega) + V_m(n,\omega), \tag{3}$$

where $D_m(n,\omega)$ is the relative transfer function between the mth microphone and the reference (first) microphone. The signal model mismatch between (2) and (3) may dramatically degrade the beamforming performance if robustness is not taken into account.

By stacking all the M observation signals in a vector form, we obtain

$$\begin{aligned} \mathbf{y}(n,\omega) &= [Y_1(n,\omega) \quad Y_2(n,\omega) \quad \dots \quad Y_M(n,\omega)]^T \\ &= \mathbf{a}(n,\omega)X(n,\omega) + \mathbf{v}(n,\omega), \end{aligned} \tag{4}$$

where $\mathbf{a}(n,\omega) = \begin{bmatrix} 1 & e^{-j\omega\tau_0} & \cdots & e^{-j(M-1)\omega\tau_0} \end{bmatrix}^T$ is the signal propagation vector (in the same form as the steering vector) of the source signal if there is no reverberation, and $[\cdot]^T$ denotes the transpose operator. In order to do adaptive beamforming in the following sections, the second order statistics of the signals are needed. The covariance matrix of $\mathbf{y}(n,\omega)$ is defined as follows:

$$\mathbf{R}_{\mathbf{y}}(n,\omega) = E[\mathbf{y}(n,\omega)\mathbf{y}^H(n,\omega)], \tag{5}$$

where $E[\cdot]$ denotes mathematical expectation, and $[\cdot]^H$ is the conjugate-transpose operator. In practice, $\mathbf{R}_{\mathbf{y}}(n,\omega)$ is not accessible and often estimated using the short-time average method, i.e.,

$$\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{y}(n-k,\omega)\mathbf{y}^H(n-k,\omega), \tag{6}$$

where N is the total number of most recent frames used in the short-time average. The mismatch between (5) and (6) may significantly degrade the beamforming performance [10].

The objective of beamforming is then to develop an optimal filter to obtain a good estimate of the desired signal. By applying a spatial filter $\mathbf{h}(n,\omega)$ to the noisy observations in (4), we get

$$\begin{aligned} Z(n,\omega) &= \mathbf{h}^H(n,\omega)\mathbf{y}(n,\omega) \\ &= X_{\mathrm{fd}}(n,\omega) + V_{\mathrm{rn}}(n,\omega), \end{aligned} \tag{7}$$

where $Z(n,\omega)$ is an estimate of $X(n,\omega)$, while $X_{\mathrm{fd}}(n,\omega) = \mathbf{h}^H(n,\omega)\mathbf{x}(n,\omega)$ and $V_{\mathrm{rn}}(n,\omega) = \mathbf{h}^H(n,\omega)\mathbf{v}(n,\omega)$ are the filtered desired signal and the residual noise, respectively. The derivation of $\mathbf{h}(n,\omega)$ will be discussed in the following sections.

3. Standard Capon and MVDR beamformers

3.1. Standard Capon beamformer (SCB)

The standard Capon beamformer, also known as the minimum power distortionless response (MPDR) beamformer, is derived by minimizing the output power of the array with the constraint that the signal from the desired look direction is undistorted. Mathematically, the SCB is obtained by solving the following optimization problem

$$\begin{aligned} \mathbf{h}_{\mathrm{SCB}}(n,\omega) = \arg\min_{\mathbf{h}(n,\omega)} \ & \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega) \\ \text{subject to } & \mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) = 1, \end{aligned} \tag{8}$$

where $\bar{\mathbf{a}}(n,\omega)$ is an estimate of the steering vector. The solution of (8) is

$$\mathbf{h}_{\mathrm{SCB}}(n,\omega) = \frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}. \tag{9}$$

It has been shown in [7–9] that the MPDR beamformer is sensitive to steering vector estimation errors.

3.2. Minimum variance distortionless response (MVDR) beamformer

Unlike the SCB which utilizes the noisy statistics, the MVDR beamformer is derived by using the noise covariance matrix, which makes the MVDR beamformer robust against steering vector inaccuracy. The MVDR beamformer is obtained by solving the following problem

$$\begin{aligned} \mathbf{h}_{\mathrm{MVDR}}(n,\omega) = \arg\min_{\mathbf{h}(n,\omega)} \ & \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{v}}(n,\omega)\mathbf{h}(n,\omega) \\ \text{subject to } & \mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) = 1, \end{aligned} \tag{10}$$

where $\bar{\mathbf{R}}_{\mathbf{v}}(n,\omega)$ is an estimate of the noise covariance matrix $\mathbf{R}_{\mathbf{v}}(n,\omega)$, and $\mathbf{R}_{\mathbf{v}}(n,\omega)$ has a similar definition to $\mathbf{R}_{\mathbf{y}}(n,\omega)$ in (5). The solution to (10) is

$$\mathbf{h}_{\mathrm{MVDR}}(n,\omega) = \frac{\bar{\mathbf{R}}_{\mathbf{v}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{v}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}. \tag{11}$$

The implementation of the MVDR filter in (11) requires to estimate the noise covariance matrix, which is not trivial in practice. Note that, theoretically, the beamformers in (9) and (11) are strictly equivalent if there is no errors in statistics estimation.

4. Robust beamformers

In this section, we discuss how to derive optimal beamformers, which are robust against steering vector estimation errors.

4.1. SCB with diagonal loading

The robustness of SCB against steering vector estimation errors can be improved by applying the common diagonal loading technique. The SCB with a fixed amount of diagonal loading is derived as

$$\mathbf{h}_{\mathrm{SCBD}}(n,\omega) = \frac{\left[\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \alpha\mathbf{I}\right]^{-1}\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)\left[\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \alpha\mathbf{I}\right]^{-1}\bar{\mathbf{a}}(n,\omega)}, \tag{12}$$

where the parameter $\alpha \geq 0$ controls the amount of diagonal loading, and $\mathbf{I}$ is an identity matrix of size $M \times M$. However, the proper selection of α is not obvious. In the next subsection, the norm constrained Capon beamformer (NCCB) is introduced to adaptively adjust the diagonal loading amount.

4.2. Norm constrained Capon beamformer

By adding a norm constraint to the problem in (8), the NCCB method is obtained [12]. The norm constraint helps limit the amplification of the spatial white noise as well as improving the robustness of the beamformer against steering vector estimation error [12]. The optimization problem for NCCB is stated as

$$\begin{aligned} \mathbf{h}_{\mathrm{NCCB}}(n,\omega) = \arg\min_{\mathbf{h}(n,\omega)} \ & \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega) \\ \text{subject to } & \mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) = 1 \\ & \|\mathbf{h}(n,\omega)\|^2 \leq \zeta, \end{aligned} \tag{13}$$

where ζ is a user-chosen parameter. The value of ζ plays an important role on the performance and robustness of NCCB, which will be discussed in Section 5. The Lagrangian of (13) is

$$\begin{aligned} &L(\mathbf{h}(n,\omega),\mu,\nu) \\ &= \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega) + 2\mu[\mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) - 1] \\ &\quad + \nu(\|\mathbf{h}(n,\omega)\|^2 - \zeta), \end{aligned} \tag{14}$$

where μ and ν are Lagrangian multipliers $\nu \geq 0$.

We first consider the case with $\nu = 0$, which means the norm constraint is not active. The problem then degenerates to (8), and the solution is given by (9).

The second scenario is when $\nu > 0$. By taking the derivative of the Lagrangian with respect to $\mathbf{h}(n,\omega)$ and equating the result to 0, we obtain

$$\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega) + \mu\bar{\mathbf{a}}(n,\omega) + \nu\mathbf{h}(n,\omega) = 0, \tag{15}$$

from which we readily derive

$$\mathbf{h}_{\mathrm{NCCB}}(n,\omega) = -\mu(\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega). \tag{16}$$

The beamformer in (16) can be interpreted as diagonal loading type of method, which is widely used in beamforming. The adaptive parameter ν controls the amount of diagonal loading and makes it possible to obtain a proper amount of diagonal loading. With the distortionless constraint, we also have

$$\mu = -\frac{1}{\bar{\mathbf{a}}^H(n,\omega)\left[\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I}\right]^{-1}\bar{\mathbf{a}}(n,\omega)}. \tag{17}$$

Substituting (17) into (16), we get

$$\mathbf{h}_{\mathrm{NCCB}}(n,\omega) = \frac{(\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)\left[\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I}\right]^{-1}\bar{\mathbf{a}}(n,\omega)}. \tag{18}$$

Let us consider the equality constraint of ζ first, we then have that

$$\left\| \frac{(\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)(\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \nu\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)} \right\|^2 = \zeta. \tag{19}$$

Applying eigenvalue decomposition to the matrix $\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)$ gives

$$\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) = \mathbf{U}(n,\omega)\mathbf{\Gamma}(n,\omega)\mathbf{U}^H(n,\omega), \tag{20}$$

where $\mathbf{U}(n,\omega) = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_M]$ contains the all the eigenvectors and $\mathbf{\Gamma}(n,\omega)$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues, i.e., $\mathbf{\Gamma}_{mm}(n,\omega) = \gamma_m$. We assume that the eigenvalues are sorted in a descending order, i.e., $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_M \geq 0$. With (19) and according to [10], we get the bounds on ν as

$$0 \leq \nu \leq \frac{\gamma_1 - \sqrt{\zeta M}\gamma_M}{\sqrt{\zeta M} - 1}. \tag{21}$$

There is a unique solution for ν in the proposed range with a given ζ [10].

With an estimate of the parameter ν from (19) and by using the eigenvalue decomposition result in (20), (18) can be rewritten as

$$\mathbf{h}_{\mathrm{NCCB}}(n,\omega) = \frac{\mathbf{U}(\mathbf{\Gamma} + \nu\mathbf{I})^{-1}\mathbf{U}^H\bar{\mathbf{a}}}{\bar{\mathbf{a}}^H\mathbf{U}(\mathbf{\Gamma} + \nu\mathbf{I})^{-1}\mathbf{U}^H\bar{\mathbf{a}}}, \tag{22}$$

where the inverse of the diagonal matrix $\mathbf{\Gamma} + \nu\mathbf{I}$ can be easily computed. We omit the frame index and frequency index for clarity. Compared with the filter in (12), the NCCB method adaptively chooses the diagonal loading amount.

4.3. Robust Capon beamformer (RCB)

Another robust solution for beamforming is the robust Capon beamformer[14]. The robustness of the filter against steering vector estimation error is obtained by adding an uncertainty constraint to the steering vector. RCB is obtained by solving the following optimization problem:

$$\begin{aligned} \mathbf{a}_{\mathrm{RCB}}(n,\omega) = \arg\min_{\mathbf{a}(n,\omega)} \ & \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) \\ \text{subject to } & \|\mathbf{a}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|^2 \leq \epsilon, \end{aligned} \tag{23}$$

where $\epsilon \geq 0$ is a user-chosen parameter. The problem is convex and can be solved by using the method of Lagrange multipliers. The Lagrangian of (23) is

$$\begin{aligned} L(\mathbf{a}(n,\omega),\xi) &= \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) \\ &\quad + \xi(\|\mathbf{a}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|^2 - \epsilon), \end{aligned} \tag{24}$$

where $\xi \geq 0$ is the Lagrangian multiplier. We then have

$$\begin{aligned} \mathbf{a}_{\mathrm{RCB}}(n,\omega) &= \left(\frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)}{\xi} + \mathbf{I}\right)^{-1}\bar{\mathbf{a}}(n,\omega) \\ &= \left(\mathbf{I} - \left(\xi\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \mathbf{I}\right)^{-1}\right)\bar{\mathbf{a}}(n,\omega). \end{aligned} \tag{25}$$

With the uncertainty constraint, the parameter ξ can be obtained by finding the root of the following function:

$$\left\| \left(\xi\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega) + \mathbf{I}\right)^{-1}\bar{\mathbf{a}}(n,\omega)\right\|^2 = \epsilon. \tag{26}$$

This is again a nonlinear equation. The function on the left-hand side is a monotonically decreasing function of ξ, so there is a unique solution to ξ for a given ϵ [14]. We can get an upper bound and lower bound for ξ as

$$\frac{\|\bar{\mathbf{a}}(n,\omega)\| - \sqrt{\epsilon}}{\gamma_1\sqrt{\epsilon}} \leq \xi \leq \frac{\|\bar{\mathbf{a}}(n,\omega)\| - \sqrt{\epsilon}}{\gamma_M\sqrt{\epsilon}}. \tag{27}$$

With the solution to ξ from (26), we have a revised estimate of the steering vector according to (25). The RCB has a similar form to (9), and can be obtained as

$$\begin{aligned} \mathbf{h}_{\mathrm{RCB}}(n,\omega) &= \frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{RCB}}(n,\omega)}{\mathbf{a}_{\mathrm{RCB}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{RCB}}(n,\omega)} \\ &= \frac{\left(\bar{\mathbf{R}}_{\mathbf{y}} + \frac{1}{\xi}\mathbf{I}\right)^{-1}\bar{\mathbf{a}}}{\bar{\mathbf{a}}^H\left(\bar{\mathbf{R}}_{\mathbf{y}} + \frac{1}{\xi}\mathbf{I}\right)^{-1}\bar{\mathbf{R}}_{\mathbf{y}}\left(\bar{\mathbf{R}}_{\mathbf{y}} + \frac{1}{\xi}\mathbf{I}\right)^{-1}\bar{\mathbf{a}}} \\ &= \frac{\mathbf{U}\left(\mathbf{\Gamma} + \frac{1}{\xi}\mathbf{I}\right)^{-1}\mathbf{U}^H\bar{\mathbf{a}}}{\bar{\mathbf{a}}^H\mathbf{U}\left(\frac{1}{\xi^2}\mathbf{\Gamma}^{-1} + \mathbf{\Gamma} + \frac{2}{\xi}\mathbf{I}\right)^{-1}\mathbf{U}^H\bar{\mathbf{a}}}. \end{aligned} \tag{28}$$

Due to computing the eigen-decomposition of matrix $\bar{\mathbf{R}}_{\mathbf{y}}$, the complexity of RCB is $\mathcal{O}(M^3)$ flops which is the same as SCB.

4.4. Double-constraint robust Capon beamformer (DCRCB)

In this section, we recall the derivation of the double-constraint robust Capon beamformer presented in [10]. By taking into account the uncertainty constraint as well as the norm constraint of the steering vector, the optimization problem can be formulated as

$$\begin{aligned}\mathbf{a}_{\text{DCRCB}}(n,\omega) = \arg\min_{\mathbf{a}(n,\omega)} \ & \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) \\ \text{subject to } & \|\mathbf{a}(n,\omega)-\bar{\mathbf{a}}(n,\omega)\|^2 \leq \epsilon \\ & \|\mathbf{a}(n,\omega)\|^2 = M. \end{aligned} \tag{29}$$

Due to the equality constraint, the problem in (29) is not convex contrary to the previous beamformer design problems. Since the estimated steering vector satisfies $\|\bar{\mathbf{a}}(n,\omega)\|^2 = M$, we have

$$\|\mathbf{a}(n,\omega)-\bar{\mathbf{a}}(n,\omega)\|^2 = 2M - 2\Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)], \tag{30}$$

where $\Re[\cdot]$ denotes taking the real part of a complex number. An equivalent formulation of the problem can be obtained as

$$\begin{aligned}\mathbf{a}_{\text{DCRCB}}(n,\omega) = \arg\min_{\mathbf{a}(n,\omega)} \ & \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) \\ \text{subject to } & \Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] \geq \delta \\ & \|\mathbf{a}(n,\omega)\|^2 = M, \end{aligned} \tag{31}$$

where $\delta = M - \frac{\epsilon}{2}$. The Lagrangian of (31) is

$$\begin{aligned}L(\mathbf{a}(n,\omega),\lambda,\varrho) = & \ \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) \\ & + \lambda\left(\delta - \Re\left[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)\right]\right) + \varrho(M - \|\mathbf{a}(n,\omega)\|^2). \end{aligned} \tag{32}$$

The KKT conditions of the problem are

$$\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) - \lambda\bar{\mathbf{a}}(n,\omega) - \varrho\mathbf{a}(n,\omega) = 0 \tag{33}$$

$$\Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] \geq \delta \tag{34}$$

$$\|\mathbf{a}(n,\omega)\|^2 = M \tag{35}$$

$$\lambda(\delta - \Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)]) = 0, \ \lambda \geq 0. \tag{36}$$

The optimization problem will only be solved by an optimal point if some constraint qualification (or regularity conditions) are satisfied [29]. Note that the KKT conditions are only necessary (but not sufficient) conditions since the problem is non-convex. As constraint qualification, we consider the linear independence constraint qualification (LICQ), i.e., the gradients for active inequality constraint and equality constraint are linear independent at solution $\mathbf{a}^\star(n,\omega)$. For the problem in (31), the gradients are $\nabla_{\mathbf{a}(n,\omega)}\Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] = 2\bar{\mathbf{a}}(n,\omega)$ and $\nabla_{\mathbf{a}(n,\omega)}\|\mathbf{a}(n,\omega)\|_2^2 = 2\mathbf{a}(n,\omega)$. Consider that there is linear dependence and the LICQ is not satisfied, then

$$\mathbf{a}^\star(n,\omega) = \alpha\bar{\mathbf{a}}(n,\omega), \ \alpha \in \mathbf{C}\,. \tag{37}$$

Due to the equality $M = \|\mathbf{a}^\star(n,\omega)\|_2^2 = |\alpha|^2\|\bar{\mathbf{a}}(n,\omega)\|_2^2$, we have that $|\alpha| = 1$. The objective in (31) in this case becomes

$$\begin{aligned} f\left(\mathbf{a}^\star(n,\omega)\right) &= (\alpha\bar{\mathbf{a}}(n,\omega))^H\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)(\alpha\bar{\mathbf{a}}(n,\omega)) \\ &= |\alpha|^2\bar{\mathbf{a}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega). \end{aligned} \tag{38}$$

With the equality constraint, which implies that $|\alpha| = 1$, the objective is constant. We can now consider that either *1)* the LICQ is satisfied and compute all $\mathbf{a}(n,\omega)$ that solves the KKT or *2)* all $\alpha\bar{\mathbf{a}}(n,\omega)$, with $|\alpha| = 1$ and feasible $\|\alpha\bar{\mathbf{a}}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|_2^2 = |1-\alpha|^2\|\bar{\mathbf{a}}(n,\omega)\|_2^2 = |1-\alpha|^2 M \leq \epsilon$ are possible solutions. Since all the latter solutions have the same objective, and we are simply seaking *a solution*, we can pick $\bar{\mathbf{a}}(n,\omega)$ ($\alpha = 1$) and all possible $\mathbf{a}(n,\omega)$ that satisfy the KKT conditions as candidates and select the one with the smallest objective as the solution.

Assume now the LICQ holds. Furthermore, there is a unique (λ,ν) that solves the KKT system if and only if the LICQ is satisfied [29]. We first consider the case $\lambda = 0$, then the KKT conditions become

$$\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) = \varrho\mathbf{a}(n,\omega) \tag{39}$$

$$\|\mathbf{a}(n,\omega)\|_2^2 = M, \ \Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] \geq \delta\,. \tag{40}$$

From this we can obtain a solution for the case $\lambda = 0$:

- Compute the eigen-decomposition of $\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)$ and evaluate the following starting from the eigenvector associated with the largest eigenvalue of $\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)$, i.e., γ_1.
  1. For each eigenvector $\mathbf{u}_m$, compute a candidate solution $\check{\mathbf{a}}(n,\omega)$ by scaling the eigenvector to have $\|\check{\mathbf{a}}(n,\omega)\|_2^2 = M$ and phase rotating to maximize $\Re[\check{\mathbf{a}}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)]$, i.e., $\check{\mathbf{a}}(n,\omega) = \sqrt{M}e^{j\angle\mathbf{u}_m^H\bar{\mathbf{a}}(n,\omega)}\mathbf{u}_m$.
  2. Test feasibility, i.e., if $\|\check{\mathbf{a}}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|_2^2 \leq \epsilon$ or $\Re[\check{\mathbf{a}}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] \geq \delta$.
  3. If feasible and $f(\check{\mathbf{a}}(n,\omega)) < f(\bar{\mathbf{a}}(n,\omega))$, return $\check{\mathbf{a}}(n,\omega)$ as a solution for $\mathbf{a}_{\text{DCRCB}}(n,\omega)$.

The reason we can stop is that we are evaluating the eigenvectors in order, and if a scaled eigenvector is feasible associated with the unique ϱ, then this must be the solution with the smallest objective.

We can continue and consider $\lambda > 0$ which leads to

$$\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega) - \lambda\bar{\mathbf{a}}(n,\omega) - \varrho\mathbf{a}(n,\omega) = 0 \tag{41}$$

$$\|\mathbf{a}(n,\omega)\|_2^2 = M, \ \Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)] = \delta. \tag{42}$$

From the first equation, consider the case with a selected ϱ such that $\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega) - \varrho\mathbf{I}$ is positive definite. The revised steering vector can be derived as

$$\mathbf{a}(n,\omega) = \lambda(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega) - \varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega). \tag{43}$$

From the equality constraints we then have

$$\lambda^2\|(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega) - \varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)\|_2^2 = M. \tag{44}$$

Since the inequality constraint is active, we have

$$\begin{aligned}&\Re[\mathbf{a}^H(n,\omega)\bar{\mathbf{a}}(n,\omega)]\\&\quad=\Re[\lambda\bar{\mathbf{a}}^H(n,\omega)(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)]\\&\quad=\sqrt{M}\frac{\bar{\mathbf{a}}^H(n,\omega)(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)}{\|(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)\|_2}\\&\quad=\delta.\end{aligned}\tag{45}$$

This is a non-linear equation in the real scalar $\varrho\in\mathbf{R}$. By squaring, (45) can be rewritten as [10, (51)–(52)]:

$$\begin{aligned}\rho&=\frac{M}{\delta^2}\\&=\frac{\bar{\mathbf{a}}^H(n,\omega)(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-2}\bar{\mathbf{a}}(n,\omega)}{(\bar{\mathbf{a}}^H(n,\omega)(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega))^2}.\end{aligned}\tag{46}$$

There is a unique solution to (46) for $\varrho\in(-\infty,1/\gamma_1)$ [10].

Furthermore a lower bound on ϱ can be obtained as follows [10]:

$$\varrho\geq\frac{\frac{1}{\gamma_1}\sqrt{M\rho}-\frac{1}{\gamma_M}}{\sqrt{M\rho}-1}.\tag{47}$$

The algorithm is summarized as follows:

- Compute the unique solution to (46) for $\varrho\in\left[\frac{\frac{1}{\gamma_1}\sqrt{M\rho}-\frac{1}{\gamma_M}}{\sqrt{M\rho}-1},1/\gamma_1\right)$.
- Compute $\lambda=\frac{\sqrt{M}}{\|(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega)\|_2}$.
- A candidate solution is obtained as

$$\mathbf{a}(n,\omega)=\lambda(\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)-\varrho\mathbf{I})^{-1}\bar{\mathbf{a}}(n,\omega).\tag{48}$$

- Select the solution with the smallest objective

$$\mathbf{a}_{\mathrm{DCRCB}}(n,\omega)=\begin{cases}\mathbf{a}(n,\omega) & \text{if } f(\mathbf{a}(n,\omega))\leq f(\bar{\mathbf{a}}(n,\omega))\\ \bar{\mathbf{a}}(n,\omega) & \text{otherwise .}\end{cases}\tag{49}$$

After we get a revised estimate of the steering vector, the DCRCB is formed in a same way as the RCB in (28), i.e.,

$$\mathbf{h}_{\mathrm{DCRCB}}(n,\omega)=\frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{DCRCB}}(n,\omega)}{\mathbf{a}_{\mathrm{DCRCB}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{DCRCB}}(n,\omega)}.\tag{50}$$

4.5. Amplitude-and-phase-estimation (APES) beamformer

The APES beamformer uses the spatial smoothing technique to improve the robustness against steering vector and covariance matrix estimation errors.

Let $\bar{M}<M$ denote the number of microphones in the subarray. The lth subarray contains the microphones from number l to $l+\bar{M}-1$. The subarray observation signal vector is written as

$$\bar{\mathbf{y}}_l(n,\omega)=\begin{bmatrix}Y_l(n,\omega) & Y_{l+1}(n,\omega) & \cdots & Y_{l+\bar{M}-1}(n,\omega)\end{bmatrix}^T,\tag{51}$$

and the corresponding subarray steering vector of size $\bar{M}\times 1$ is

$$\bar{\mathbf{a}}_l(n,\omega)=\begin{bmatrix}e^{-jl\omega\tau_0} & e^{-j(l+1)\omega\tau_0} & \cdots & e^{-j(l+\bar{M}-1)\omega\tau_0}\end{bmatrix}^T,\tag{52}$$

where $l=0,\ldots,L-1$, and $L=M-\bar{M}+1$. With the ULA assumption, the subarray steering vectors are related as

$$\bar{\mathbf{a}}_l(n,\omega)=e^{-jl\omega\tau_0}\bar{\mathbf{a}}_0(n,\omega).\tag{53}$$

Using (53), one can rewrite $\bar{\mathbf{y}}_l(n,\omega)$ as

$$\bar{\mathbf{y}}_l(n,\omega)=e^{-jl\omega\tau_0}\bar{\mathbf{a}}_0(n,\omega)X(n,\omega)+\bar{\mathbf{v}}_l(n,\omega).\tag{54}$$

The principle of APES is to minimize the least-squares error between the beamformer's output and the desired signal for each subarray with the constraint that the signal from the desired look direction is undistorted[22]. If we combine (54) with the APES principle, an estimate of $X(n,\omega)$ and the filter $\bar{\mathbf{h}}(n,\omega)$ of length $\bar{M}$ can be obtained by solving the following optimization problem

$$\min_{\bar{\mathbf{h}}(n,\omega),X(n,\omega)}\ J(n,\omega)\ \text{ subject to }\ \bar{\mathbf{h}}^H(n,\omega)\bar{\mathbf{a}}_0(n,\omega)=1,\tag{55}$$

where

$$\begin{aligned}&J(n,\omega)\\&=\sum_{k=0}^{N-1}\sum_{l=0}^{L-1}\left|\bar{\mathbf{h}}^H(n,\omega)\bar{\mathbf{y}}_l(n-k,\omega)e^{jl\omega\tau_0}-X(n-k,\omega)\right|^2\end{aligned}\tag{56}$$

Let us reformulate the optimization problem in (55) over only $\bar{\mathbf{h}}(n,\omega)$. To this end, let

$$\mathbf{g}(n-k,\omega)=\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{y}}_l(n-k,\omega)e^{jl\omega\tau_0}.\tag{57}$$

Note that

$$\begin{aligned}&\frac{1}{L}\sum_{l=0}^{L-1}\left|\bar{\mathbf{h}}^H(n,\omega)\bar{\mathbf{y}}_l(n-k,\omega)e^{jl\omega\tau_0}-X(n-k,\omega)\right|^2\\&\quad=\bar{\mathbf{h}}^H(n,\omega)\left[\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{y}}_l(n-k,\omega)\bar{\mathbf{y}}_l^H(n-k,\omega)\right]\bar{\mathbf{h}}(n,\omega)\\&\quad-\bar{\mathbf{h}}^H(n,\omega)\mathbf{g}(n-k,\omega)\mathbf{g}^H(n-k,\omega)\bar{\mathbf{h}}(n,\omega)\\&\quad+|X(n-k,\omega)-\bar{\mathbf{h}}^H(n,\omega)\mathbf{g}(n-k,\omega)|^2.\end{aligned}\tag{58}$$

Minimizing (58) with respect to $X(n-k,\omega)$, we obtain an estimate of $X(n-k,\omega)$ as

$$X(n-k,\omega)=\bar{\mathbf{h}}^H(n,\omega)\mathbf{g}(n-k,\omega).\tag{59}$$

Substituting (59) into (55) gives

$$\begin{aligned}\bar{\mathbf{h}}_{\mathrm{APES}}(n,\omega)=\arg\min_{\bar{\mathbf{h}}(n,\omega)}\ &\bar{\mathbf{h}}^H(n,\omega)\hat{\mathbf{Q}}(n,\omega)\bar{\mathbf{h}}(n,\omega)\\&\text{subject to }\ \bar{\mathbf{h}}^H(n,\omega)\bar{\mathbf{a}}_0(n,\omega)=1,\end{aligned}\tag{60}$$

where

$$\hat{\mathbf{Q}}(n,\omega) = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{y}}_l(n-k,\omega)\bar{\mathbf{y}}_l^H(n-k,\omega) - \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{g}(n-k,\omega)\mathbf{g}^H(n-k,\omega). \quad (61)$$

This expression can be interpreted as an estimate of the noise covariance matrix. Note that we should have $NL \geq M$ to ensure that $\hat{\mathbf{Q}}(n,\omega)$ is positive-definite. The solution to (60) is then given by

$$\bar{\mathbf{h}}_{\mathrm{APES}}(n,\omega) = \frac{\hat{\mathbf{Q}}^{-1}(n,\omega)\bar{\mathbf{a}}_0(n,\omega)}{\bar{\mathbf{a}}_0^H(n,\omega)\hat{\mathbf{Q}}^{-1}(n,\omega)\bar{\mathbf{a}}_0(n,\omega)}. \quad (62)$$

With (57), (59) and (62), the estimate of $X(n,\omega)$ is

$$Z_{\mathrm{APES}}(n,\omega) = \bar{\mathbf{h}}^H(n,\omega)\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{y}}_l(n,\omega)e^{jl\omega\tau_0}. \quad (63)$$

As shown in (63), the filtering procedure of APES beamformer is different from the other methods, since it applies a spatial smoothing technique.

We summarize SCB and all the robust beamformers in Table 1.

Table 1: Beamformers

SCB	$\min_{\mathbf{h}(n,\omega)} \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega)$ s. t. $\mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) = 1$
filter	$\mathbf{h}_{\mathrm{SCB}}(n,\omega) = \frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}{\bar{\mathbf{a}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\bar{\mathbf{a}}(n,\omega)}$
NCCB	$\min_{\mathbf{h}(n,\omega)} \mathbf{h}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}(n,\omega)\mathbf{h}(n,\omega)$ s. t. $\mathbf{h}^H(n,\omega)\bar{\mathbf{a}}(n,\omega) = 1$, $\|\mathbf{h}(n,\omega)\|^2 \leq \zeta$
filter	$\mathbf{h}_{\mathrm{NCCB}}(n,\omega) = \frac{\mathbf{U}(\mathbf{\Gamma}+\nu\mathbf{I})^{-1}\mathbf{U}^H\bar{\mathbf{a}}}{\bar{\mathbf{a}}^H\mathbf{U}(\mathbf{\Gamma}+\nu\mathbf{I})^{-1}\mathbf{U}^H\bar{\mathbf{a}}}$
RCB	$\min_{\mathbf{a}(n,\omega)} \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega)$ s. t. $\|\mathbf{a}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|^2 \leq \epsilon$
filter	$\mathbf{h}_{\mathrm{RCB}}(n,\omega) = \frac{\mathbf{U}\left(\mathbf{\Gamma}+\frac{1}{\xi}\mathbf{I}\right)^{-1}\mathbf{U}^H\bar{\mathbf{a}}}{\bar{\mathbf{a}}^H\mathbf{U}\left(\frac{1}{\xi^2}\mathbf{\Gamma}^{-1}+\mathbf{\Gamma}+\frac{2}{\xi}\mathbf{I}\right)^{-1}\mathbf{U}^H\bar{\mathbf{a}}}$
DCRCB	$\min_{\mathbf{a}(n,\omega)} \mathbf{a}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}(n,\omega)$ s. t. $\|\mathbf{a}(n,\omega) - \bar{\mathbf{a}}(n,\omega)\|^2 \leq \epsilon$, $\|\mathbf{a}(n,\omega)\|^2 = M$
filter	$\mathbf{h}_{\mathrm{DCRCB}}(n,\omega) = \frac{\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{DCRCB}}(n,\omega)}{\mathbf{a}_{\mathrm{DCRCB}}^H(n,\omega)\bar{\mathbf{R}}_{\mathbf{y}}^{-1}(n,\omega)\mathbf{a}_{\mathrm{DCRCB}}(n,\omega)}$
APES	$\min_{\bar{\mathbf{h}}(n,\omega)} \bar{\mathbf{h}}^H(n,\omega)\hat{\mathbf{Q}}(n,\omega)\bar{\mathbf{h}}(n,\omega)$ s. t. $\bar{\mathbf{h}}^H(n,\omega)\bar{\mathbf{a}}_0(n,\omega) = 1$
filter	$\bar{\mathbf{h}}_{\mathrm{APES}}(n,\omega) = \frac{\hat{\mathbf{Q}}^{-1}(n,\omega)\bar{\mathbf{a}}_0(n,\omega)}{\bar{\mathbf{a}}_0^H(n,\omega)\hat{\mathbf{Q}}^{-1}(n,\omega)\bar{\mathbf{a}}_0(n,\omega)}$

5. Simulations

We will study in this section the performance of the robust beamformers in reverberant and noisy environments.

5.1. Performance measures

In this subsection, we present some performance measures to evaluate the aforementioned robust beamformers. Without loss of generality, we consider the signal of interest received at the first microphone as the desired signal. The input signal-to-interference-and-noise-ratio (SINR) is defined as

$$\mathrm{iSINR} = \frac{E[|X(n,\omega)|^2]}{E[|V_1(n,\omega)|^2]}. \quad (64)$$

The output SINR, according to the model given in (7), is

$$\mathrm{oSINR} = \frac{E[|X_{\mathrm{fd}}(n,\omega)|^2]}{E[|V_{\mathrm{rn}}(n,\omega)|^2]}. \quad (65)$$

Note that due to the use of the spatial smoothing, the definition of SINR with the APES method in (63) is different from that of other methods. First, we rewrite (63) as

$$Z_{\mathrm{APES}}(n,\omega) = X_{\mathrm{fd,APES}}(n,\omega) + V_{\mathrm{rn,APES}}(n,\omega), \quad (66)$$

where

$$X_{\mathrm{fd,APES}}(n,\omega) = \bar{\mathbf{h}}^H(n,\omega)\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{x}}_l(n,\omega)e^{jl\omega\tau_0} \quad (67)$$

and

$$V_{\mathrm{rn,APES}}(n,\omega) = \bar{\mathbf{h}}^H(n,\omega)\frac{1}{L}\sum_{l=0}^{L-1}\bar{\mathbf{v}}_l(n,\omega)e^{jl\omega\tau_0} \quad (68)$$

are filtered desired signal and residual noise, respectively. Similar to (65), the output SINR for the APES beamformer is then defined as

$$\mathrm{oSINR}[\bar{\mathbf{h}}(n,\omega)] = \frac{E[|X_{\mathrm{fd,APES}}(n,\omega)|^2]}{E[|V_{\mathrm{rn,APES}}(n,\omega)|^2]}. \quad (69)$$

We also use the perceptual evaluation of speech quality (PESQ) [30] and the short-time objective intelligibility measure (STOI) [31] to evaluate the speech quality and intelligibility of the enhanced speech signal.

5.2. Experimental setup

In this subsection, we evaluate the robust beamformers with numerical experiments under reverberant and noisy environment. Both simulated signal and real environment recorded signal are considered. The speech quality and speech intelligibility improvement are shown in the following experiments. We consider a simulated room of size $5\mathrm{m} \times 5\mathrm{m} \times 3\mathrm{m}$. The desired signal is located at $(3.25, 1.2, 1.5)$ with DOA being $\theta_{\mathrm{d}} = 30°$. In order to simulate babble noise, we consider multiple interferences here. There are six interferences located at $(1.3, 1.5, 1.5)$, $(1.0, 2.2, 1.5)$, $(1.5, 3.7, 1.5)$, $(2.75, 4, 1.5)$, $(3.8, 3.5, 1.5)$ and

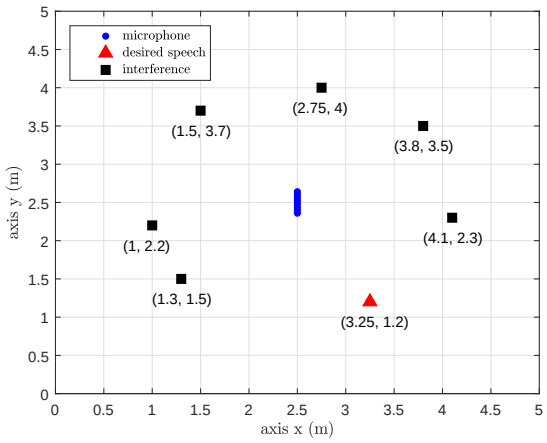
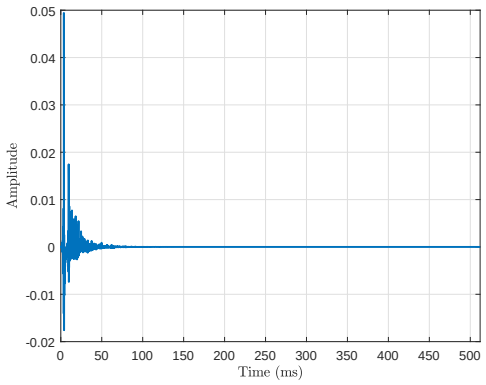


Figure 1: The room setup.

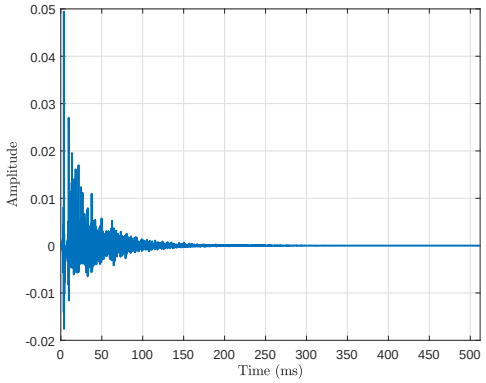
$(4.1, 2.3, 1.5)$, respectively. An ULA containing 16 microphones is located at the center of the room with $\delta = 0.04$m. The first microphone is located at $(2.5, 2.2, 1.5)$. The microphone and speaker setup is shown in Fig. 1. The room impulse response is generated by using the image model based method [32]. The room impulse responses with different reverberation times are shown in Fig. 2. The speech signal is taken from the TIMIT database[33] and down sampled to 8k Hz for our use. Each experiment is repeated with the speech signal from 10 different speakers, and 100 seconds of speech are used totally. The interferences are first scaled to have the same power before convolved with the corresponding room impulse responses. The convolved interferences at the first microphone are then added together as the interference signal. The interference signal and white Gaussian noise are scaled and added to the desired signal with a certain input SIR and SNR. The interference signal at the other microphones is proper scaled based on the input SIR at the first microphone. Unless clarified, the input SIR mentioned in the following section is at the first microphone. The time domain noisy signal is then transformed to the frequency domain by applying the STFT, the overlap rate is set to be 75% with a frame length of 128. The sound speed is set as $c = 340$ m/s.

5.3. Simulation results

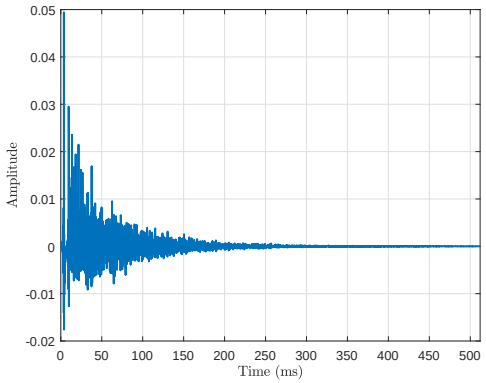
In the first experiment, we evaluate the influence of the DOA estimation error ($\Delta\theta$) on the performance of the beamformer. For comparison, the performance of DAS, MVDR and SCB are also studied. As discussed in Section 4.1, diagonal loading is commonly used in practice when applying the SCB method. We already studied in our former work [25] that the SCB method without diagonal loading even degrades the speech quality and speech intelligibility. In the experiment, we also add a fixed amount of diagonal loading to the SCB method. The results are shown in Fig. 3. We can notice that the DOA estimation error influences the performance of the beamformer dramatically. The performance of all the methods are decreasing with increasing amount of DOA estimation error. The MVDR method has the best performance in all cases. The adaptive methods outperform the fixed beamforming (DAS). We can notice that



(a) Room impulse response with $T_{60} \approx 150$ ms.

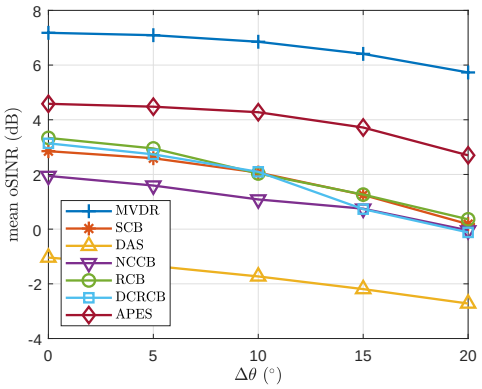


(b) Room impulse response with $T_{60} \approx 300$ ms.

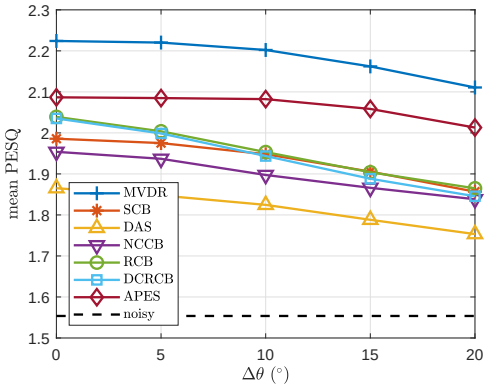


(c) Room impulse response with $T_{60} \approx 450$ ms.

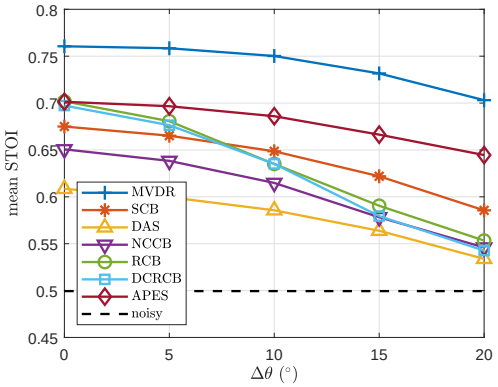
Figure 2: The room impulse response from the desired signal to the first microphone.



(a) The mean output SINR.

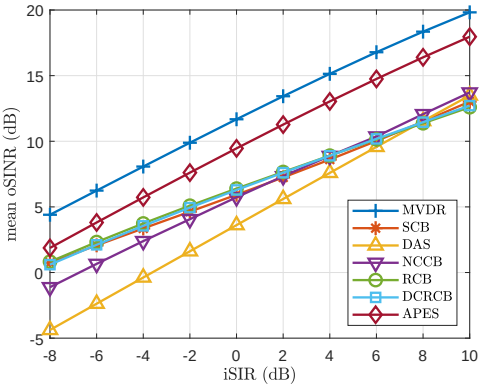


(b) The mean PESQ score.

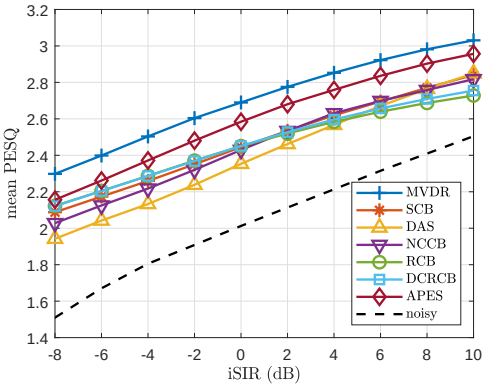


(c) The mean STOI score.

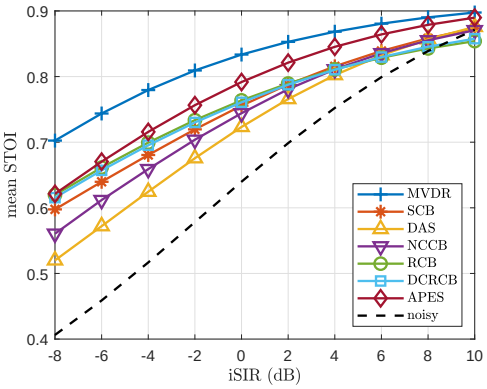
Figure 3: The beamformers performance with different amount of DOA estimation errors. We set $M = 8$, $N = 100$, $\theta_{\mathrm{d}} = 30°$, iSNR $= 20$ dB, iSIR $= -5$ dB and $T_{60} \approx 150$ ms.



(a) The mean output SINR.

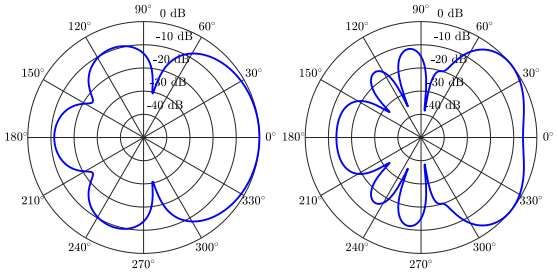


(b) The mean PESQ score.



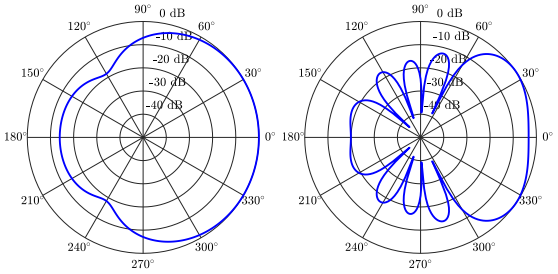
(c) The mean STOI score.

Figure 4: The beamformers performance with different input SIRs. We set $M = 8$, $N = 100$, $\theta_{\mathrm{d}} = 30°$, $\Delta\theta = 5°$, iSNR $= 20$ dB and $T_{60} \approx 150$ ms.



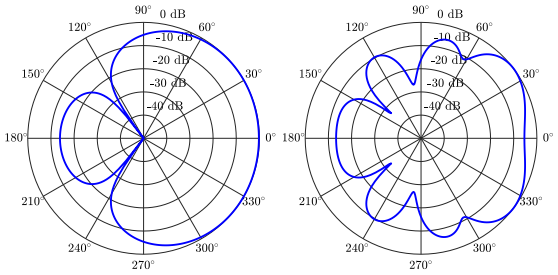
(a) MVDR, $f = 750$ Hz

(b) MVDR, $f = 2750$ Hz



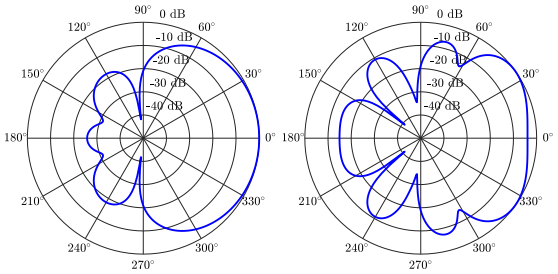
(c) SCB, $f = 750$ Hz

(d) SCB, $f = 2750$ Hz



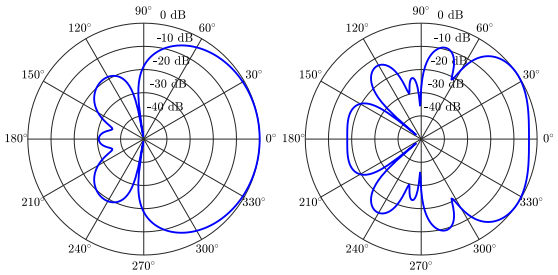
(e) NCCB, $f = 750$ Hz

(f) NCCB, $f = 2750$ Hz



(g) RCB, $f = 750$ Hz

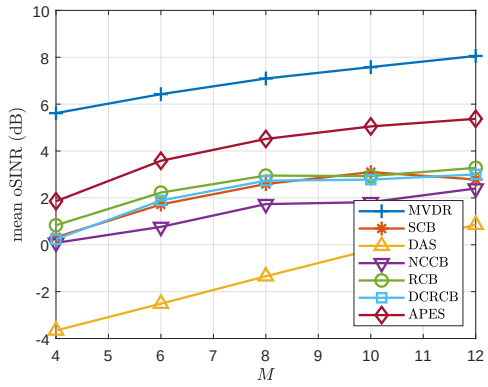
(h) RCB, $f = 2750$ Hz



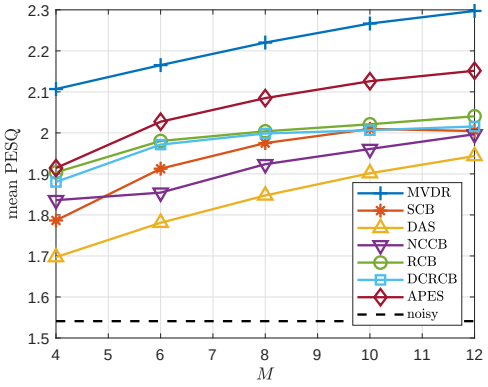
(i) DCRCB, $f = 750$ Hz

(j) DCRCB, $f = 2750$ Hz

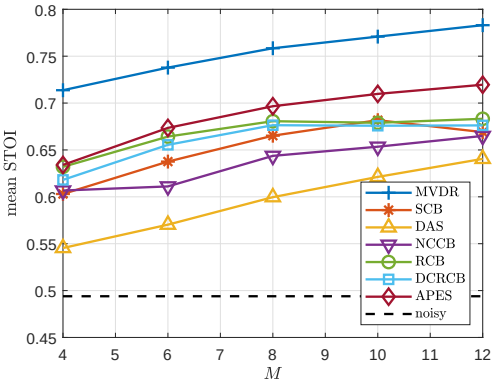
Figure 5: The beampatterns of different beamformers with different frequencies. We set $M = 8$, $N = 100$, $\theta_{\mathrm{d}} = 30°$, $\Delta\theta = 5°$, iSNR $= 20$ dB, iSIR $= -5$ dB and $T_{60} \approx 150$ ms.



(a) The mean output SINR.

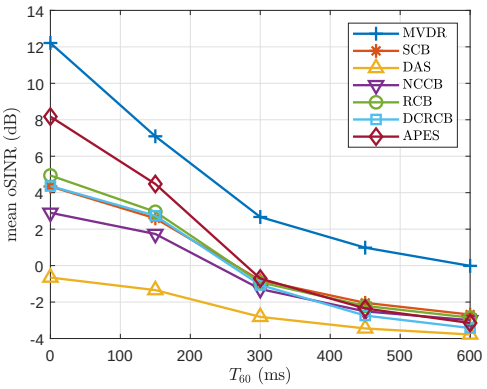


(b) The mean PESQ score.

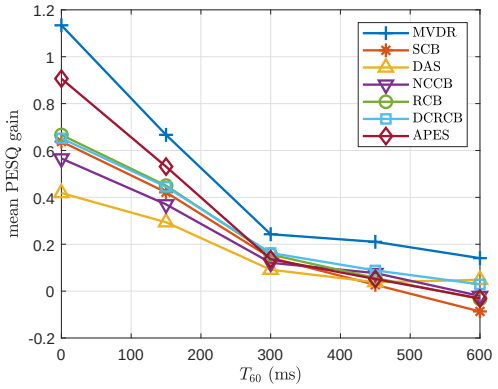


(c) The mean STOI score.

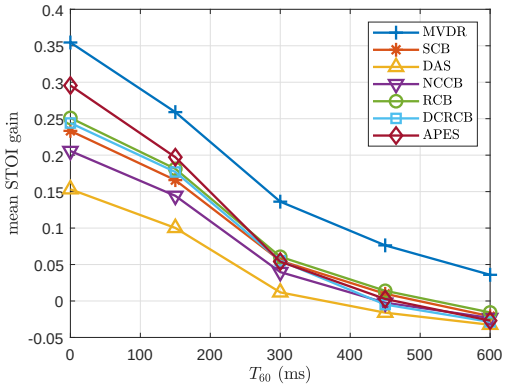
Figure 6: The beamformers performance with different number of microphones. We set $\theta_{\mathrm{d}} = 30°$, $\Delta\theta = 5°$, iSNR $= 20$ dB, iSIR $= -5$ dB and $T_{60} \approx 150$ ms.



(a) The mean output SINR.



(b) The mean PESQ gain.



(c) The mean STOI gain.

Figure 7: The beamformers performance with different amount of reverberation. We set $M = 8$, $N = 100$, $\theta_{\mathrm{d}} = 30°$, $\Delta\theta = 5°$, iSNR $= 20$ dB and iSIR $= -5$ dB.

the RCB and DCRCB methods outperform the SCB when the DOA error is small. It should be noticed that even though the APES beamformer shows better performance in improving the oSINR, the definition of oSINR for APES is different from that for the other methods as seen in (65) and (69). For RCB and DCRCB, the choice of ϵ should be frequency dependent ideally. We use a same ϵ for all the frequency bins in practice for simplicity. Here the proper values of these parameters are chosen based on experiments, which are listed in Table 2 for different DOA estimation errors. For NCCB, we have $\zeta = \beta/M$, and for APES, the subarray microphone number is chosen as $\bar{M} = M - 1$. We choose $\alpha = 10^{-4}$ for the SCB method.

Table 2: Parameters for NCCB, RCB and DCRCB with different DOA errors

$\Delta\theta$	$0°$	$5°$	$10°$	$15°$	$20°$
NCCB (β)	1.5	1.75	1.75	1.75	1.75
RCB (ϵ)	0.1	0.1	0.25	0.25	0.25
DCRCB (ϵ)	0.1	0.1	0.1	0.25	0.25

The second experiment is to study the performance of the beamformers under different iSIRs. Figure 4 shows that the robust beamformers outperform DAS method both in improving the speech quality and speech intelligibility. The APES method gives better performance than the other robust methods under all the conditions, while RCB and DCRCB behave similarly under most iSIR conditions. The results illustrate that SCB has less speech quality and speech intelligibility improvement than RCB, DCRCB and APES methods under low iSIR conditions. This is simply because it suffers from steering vector and covariance matrix estimation errors, which forces the mainlobe to point to a wrong direction and causes higher level of sidelobes as well. The DAS beamformer suffers from the high white noise amplification in low frequency which limits its performance when compares to the robust methods.

In order to analyse the behavior of different beamformers, we also illustrate the beampatterns for different methods in Fig. 5. The beampattern is defined as

$$\mathcal{B}[\mathbf{h}(n,\omega),\theta] = \mathbf{a}^H(n,\omega)\mathbf{h}(n,\omega), \tag{70}$$

which is used to describe the sensitivity of the beamformer to the plane wave impinging on the array from different directions. One can notice from Fig. 5 that the beampattern varies significantly with different frequency bins. At low frequency, the DOA estimation error does not affect the mainlobe direction much because the mainlobe is already quite wide. But the difference between the sidelobes is significant. More specifically, RCB, DCRCB have lower sidelobes compared to the other methods, which helps to better suppress the interferences and eliminate the white Gaussian noise. There is not too much of difference between the beampatterns of RCB and DCRCB at low frequency. At high frequency, It can be noticed that the steering directions of RCB and DCRCB are corrected because of the uncertainty constraint. The NCCB method gives high sidelobe at low frequency and shallow valley at unwanted directions at high frequency, which makes its performance not

as good as RCB and DCRCB as shown in Fig. 3 and Fig. 4. Additionally, from Fig. 5 (f), we can notice that the pointing direction is not well corrected by using NCCB.

The next experiment studies the performance versus different number of microphones. To estimate the observed signal covariance matrix, the proper short-time average length for different M is set according to experiments, which are listed in Table 3. The choice of parameters for different methods are also listed in Table 3. As shown in Fig. 6, the performance of robust beamformers and DAS improves with the increasing of microphone number. But the performance increasing of RCB and DCRCB are slowing down after M reaches 8. The performance of RCB with a fixed amount of diagonal loading starts to decrease after M reaching 10. This is due to the estimation error of the covariance matrix, which increases with the number of microphones. Interestingly, the performance of the APES method keeps increasing with the increasing of M, which makes it promising in further improving the speech quality and speech intelligibility with large number of microphones.

Table 3: Parameters for NCCB, RCB, DCRCB and SCB with different number of microphones.

M	4	6	8	10	12
N	40	80	100	120	160
NCCB (β)	2.75	2	1.75	1.75	1.75
RCB (ϵ)	0.1	0.1	0.1	0.5	1
DCRCB (ϵ)	0.1	0.1	0.1	0.25	0.5
SCB (α)	10^{-5}	10^{-5}	10^{-4}	10^{-4}	10^{-4}

The next experiment is to study the performance of the beamformers under different reverberation conditions. The parameters are selected by experiments which are shown in Table. 4. We select $\alpha = 10^{-4}$ for SCB method. As illustrated in Fig. 7, an increasing reverberation time dramatically decreases the speech quality and speech intelligibility. The RCB, DCRCB and APES still outperform the standard methods in most cases. However, when T_{60} reaches 450 ms, the robust methods have almost the same performance with the DAS beamformer.

Table 4: Parameters for NCCB, RCB and DCRCB with different amount of reverberations.

T_{60}	0 ms	150 ms	300 ms	450 ms	600 ms
NCCB (β)	1.75	1.75	1.5	1.25	1.25
RCB (ϵ)	0.1	0.1	0.25	0.25	0.25
DCRCB (ϵ)	0.1	0.1	0.25	0.75	1

We further study the performance of the robust filters under non-stationary noise. The number of microphones is set to 8. We consider one interference which is located at $(3.8, 3.5, 1.5)$. The interference signal is then scaled to get the iSIR at the first microphone as -5 dB. We set $T_{60} = 150$ ms and $N = 100$. The noise data in this experiment is taken from the AURORA database [35]. The multichannel non-stationary noise is generated by using the method proposed in [36]. The noise signal is scaled to have iSNR $= 20$ dB. We set the parameter $\alpha = 10^{-4}$ for SCB, $\beta = 1.75$ for NCCB and $\epsilon = 0.1$ for both RCB and DCRCB. The results are illustrated in Table 5, Table 6, and Table 7. We can notice that the RCB, DCRCB, and APES beamformer outperform the SCB and DAS method under different noise conditions.

Table 5: Performance of beamformers under restaurant noise condition. The PESQ and STOI of the noisy signal are 1.4623 and 0.6291, respectively.

Restaurant	MVDR	SCB	DAS	NCCB	RCB	DCRCB	APES
mean oSINR	20.0866	5.5134	-3.1152	3.5572	6.5247	6.4153	6.8056
mean PESQ	3.1465	2.3345	1.9228	2.2377	2.3920	2.3963	2.4600
mean STOI	0.9244	0.7684	0.6867	0.7312	0.7783	0.7789	0.8280

Table 6: Performance of beamformers under street noise condition. The PESQ and STOI of the noisy signal are 1.4574 and 0.6283, respectively.

Street	MVDR	SCB	DAS	NCCB	RCB	DCRCB	APES
mean oSINR	20.0089	5.5126	-3.1323	3.5427	6.5273	6.4168	6.9064
mean PESQ	3.1791	2.3233	1.9173	2.2282	2.3752	2.3785	2.4524
mean STOI	0.9284	0.7672	0.6857	0.7298	0.7778	0.7784	0.8266

Table 7: Performance of beamformers under station noise condition. The PESQ and STOI of the noisy signal are 1.4610 and 0.6299, respectively.

Station	MVDR	SCB	DAS	NCCB	RCB	DCRCB	APES
mean oSINR	20.2860	5.5227	-3.1185	3.5605	6.5388	6.4287	6.8535
mean PESQ	3.1447	2.3320	1.9236	2.2371	2.3876	2.3917	2.4570
mean STOI	0.9255	0.7684	0.6867	0.7311	0.7784	0.7789	0.8272

In the last experiment, we use the real environment recorded signal to test the performance of the robust beamformers. The signals are taken from the single- and multichannel audio recordings database (SMARD) [34]. We use the first ULA array in the database. The desired signal is the male speech (50_male_speech_english) from the configuration 0010 at the direction $\theta_{\mathrm{d}} = 79.8°$. An interference female speech signal (CA02_03) is located at $53.2°$ which is taken from the data in configuration 0110. More details about the room setup and the positions of the array and sources can be found in [34]. The iSIR is set to be 0dB, and we set $N = 100$. The parameters are set by experiments, i.e., $\alpha = 10^{-4}$ for SCB, $\beta = 1.75$ for NCCB and $\epsilon = 0.1$ for both RCB and DCRCB. The results of different methods are listed in Table 8. The performance of the MVDR is the best in both improving the speech quality and speech intelligibility. However, the requirement of estimating the covariance matrix of the noise signal makes MVDR beamforming difficult to implement in practice. We can notice from Table 8 that the NCCB, RCB, DCRCB and APES outperform the SCB method. Moreover, the NCCB, RCB and DCRCB show better performance with real recorded data. The performance of APES is not consistent with the performance

with simulated data, this is because of that the relationship in (53) is not satisfied any more, which degrades the beamforming performance. We remark that from listening to the output of the different methods, we have observed that the filtered signal quality is consistent with the experiment results both with simulated data and real environment recorded data.

Table 8: Performance of beamformers with real environment recorded signal. The PESQ and STOI of the noisy signal are 2.7950 and 0.8334, respectively.

	MVDR	SCB	DAS	NCCB	RCB	DCRCB	APES
oSIR	34.5816	7.8358	1.3095	4.5175	7.3247	7.0856	2.5589
PESQ	2.9986	2.4328	2.7544	2.8451	2.9105	2.9096	2.7822
STOI	0.9108	0.8192	0.8331	0.8530	0.8612	0.8601	0.8350

The simulation results indicate that the robust beamformers perform better than the traditional methods in noisy and reverberant environments. Among the studied robust methods, RCB shows slightly better performance than DCRCB. With real recorded data, the NCCB, RCB, and DCRCB show potential in improving the speech quality and speech intelligibility. With simulated data, APES beamformer is able to further improve the speech quality and speech intelligibility with large number of microphones, however, the performance with real recorded data is not that good. In summary, with the application of robust methods in acoustic signal processing, the robustness of the beamformer against the steering vector, covariance matrix and signal model errors are dramatically improved.

6. Conclusion

This paper considered different robust adaptive beamformers for wideband acoustic signals. Several different robust methods based on the Capon beamformer have been considered, as well as one based on the APES method. The NCCB method belongs to the diagonal loading type of method, but instead of setting a fixed diagonal loading parameter, NCCB adjusts the diagonal loading amount adaptively according to the current signal statistics and the steering vector. Moreover, the RCB and DCRCB intend to adaptively revise the inaccurate steering vector to improve the robustness of the filter against the steering vector estimation errors. Meanwhile, the APES method develops an estimate of the noise covariance matrix before forming the filter, which leads to the improvement of robustness. Experiments were performed in reverberant environments with multiple interference sources. We considered different input SINRs, different numbers of microphones, different amounts of reverberations, and different types of non-stationary noise. The performance of the robust methods with real environment recorded data was also studied. The results showed that the robust methods are able to improve the robustness of the beamformer against the estimation errors of both the steering vector and the covariance matrix. More importantly, these robust adaptive beamformers maintain robustness against the signal model mismatch in reverberant environments. It is worth noticing that the APES beamformer has shown the potential to improve the speech quality and speech intelligibility with large number of microphones. This is because the use of a noise covariance matrix estimate in forming the beamformer, which leads to a low level of signal distortion even when there are estimation errors in the steering vector and the signal covariance matrix.

References

[1] Van Veen BD, Buckley KM. Beamforming: A versatile approach to spatial filtering. IEEE ASSP Mag. 1988;5(2):4-24.

[2] Doclo S, Moonen M. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. IEEE Trans. Signal Process. 2003;51(10):2511-26.

[3] Doclo S, Gannot S, Moonen M, Spriet A. Acoustic beamforming for hearing aid applications. S. Haykin and K. Ray Liu, Ed. New York, NY, USA: Wiley; 2008.

[4] Pan C, Chen J, Benesty J. Performance study of the MVDR beamformer as a function of the source incidence angle. IEEE Trans. Audio, Speech, and Lang. Process. 2014;22(1):67-79.

[5] Hadad E, Doclo S, Gannot S. The binaural LCMV beamformer and its performance analysis. IEEE Trans. Audio, Speech, and Lang. Process. 2016;24(3):543-58.

[6] Capon J. High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE 1969;57(8):1408-18.

[7] Cox H. Resolving power and sensitivity to mismatch of optimum array processors. J. Acoust. Soc. Amer. 1973;54(3):771-85.

[8] Van Trees HL. Detection, estimation, and modulation theory. New York, Wiley; 2004.

[9] Ehrenberg L, Gannot S, Leshem A, Zehavi E. Sensitivity analysis of MVDR and MPDR beamformers. in Proc. 26th Conv. Electr. Electron. Eng. in Israel (IEEE). 2010;416-20.

[10] Li J, Stoica P, Wang Z. Doubly constrained robust Capon beamformer. IEEE Trans. Signal Process. 2004;52(9):2407-23.

[11] Talmon R, Cohen I, Gannot S. Relative transfer function identification using convolutive transfer function approximation. IEEE Trans. Audio, Speech, and Lang. Process. 2009;17(4):546-55.

[12] Cox H, Zeskind RM, Owen MH. Robust adaptive beamforming. IEEE Trans. Acoust. Speech Signal Process. 1987;35(10):1365-76.

[13] Li J, Stoica P. Robust Adaptive Beamforming. New York, Wiley; 2005.

[14] Li J, Stoica P, Wang Z. On robust Capon beamforming and diagonal loading. IEEE Trans. Signal Process. 2003;51(7):1702-15.

[15] Stoica P, Wang Z, Li J. Robust Capon beamforming. IEEE Signal Process. Lett. 2003;10(6):172-5.

[16] Vorobyov SA, Gershman AB, Luo Z, Ma N. Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity. IEEE Signal Process. Lett. 2004;11(2):108-11.

[17] Lorenz RG, Boyd SP. Robust minimum variance beamforming. IEEE Trans. Signal Process. 2005;53(5):1684-96.

[18] Mukherjee A, Swindlehurst AL. Robust beamforming for security in MIMO wiretap channels with imperfect CSI. IEEE Trans. Signal Process. 2011;59(1):351-61.

[19] Li J, Stoica P. An adaptive filtering approach to spectral estimation and SAR imaging. IEEE Trans. Signal Process. 1996;44(6):1469-84.

[20] Liu ZS, Li H, Li J. Efficient implementation of Capon and APES for spectral estimation. IEEE Trans. Aerosp. Electron. Syst. 1998;34(4):1314-9.

[21] Stoica P, Li H, Li J. A new derivation of the APES filter. IEEE Signal Process. Lett. 1999;6(8):205-6.

[22] Stoica P, Moses RL. Spectral Analysis of Signals. Upper Saddle River, NJ: Prentice-Hall; 2005.

[23] Christensen MG, Jakobsson A. Optimal filter designs for separating and enhancing periodic signals. IEEE Trans. Signal Process. 2010;58(12):5969-83.

[24] Ngo K. Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids. Ph.D. dissertation, ESAT, Katholieke Universiteit Leuven, Belgium; 2011.

[25] Zhao Y, Jensen JR, Christensen MG, Doclo S, Chen J. Experimental study of robust beamforming techniques for acoustic applications. in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). 2017;86-90.

[26] Benesty J, Chen J, Huang Y, Microphone Array Signal Processing. Berlin, Germany: Springer-Verlag; 2008.
[27] Zou Q, Yu ZL, Lin Z. A robust algorithm for linearly constrained adaptive beamforming. IEEE Signal Process. Lett. 2004;11(1):26-9.
[28] Pu W, Xiao J, Zhang T, Luo Z. A penalized inequality-constrained minimum variance beamformer with applications in hearing aids. in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). 2017;175-9.
[29] Wachsmuth G. On LICQ and the uniqueness of Lagrange multipliers. Operations Research Letters. 2013;41(1):78-80.
[30] Hu Y, Loizou PC. Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio, Speech, and Lang. Process. 2008;16(1):229-38.
[31] Taal CH, Hendriks RC, Heusdens R, Jensen J. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. IEEE Trans. Audio, Speech, and Lang. Process. 2011;19(7):2125-36.
[32] Allen JB, Berkley DA. Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. 1979;65(4):943-50.
[33] Garofolo J. DARPA TIMIT acoustic-phonetic continuous speech corpus. Gaithersburg, MD, USA: Nat. Inst. of Standards Technol. 1993.
[34] Nielsen JK, Jensen JR, Jensen SH, Christensen MG. The Single- and Multichannel Audio Recordings Database (SMARD). in Proc. 14th International Workshop on Acoustic Signal Enhancement (IWAENC). 2014;40-4.
[35] Hirsch HG, Pearce D. The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions. in Proc. ISCA ITRWASR. 2000;181-188.
[36] Habets EAP, Cohen I, Gannot S. Generating nonstationary multisensor signals under a spatial coherence constraint. J. Acoust. Soc. Am. 2008;124(5);2911-2917.