Experimental study of robust beamforming techniques for acoustic applications

Conference Paper · October 2017

DOI: 10.1109/WASPAA.2017.8170000

5 authors, including:

Yingke Zhao
Northwestern Polytechnical University
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Jesper Rindom Jensen
Audio Analysis Lab, Aalborg University
102 PUBLICATIONS **423** CITATIONS
SEE PROFILE

Mads Græsbøll Christensen
Aalborg University
247 PUBLICATIONS **2,222** CITATIONS
SEE PROFILE

Jingdong Chen
Institute of Electrical and Electronics Engineers
307 PUBLICATIONS **4,721** CITATIONS
SEE PROFILE

EXPERIMENTAL STUDY OF ROBUST BEAMFORMING TECHNIQUES FOR ACOUSTIC APPLICATIONS

Yingke Zhao[1,2], Jesper Rindom Jensen[2], Mads Græsbøll Christensen[2], Simon Doclo[3] and Jingdong Chen[1]

[1]CIAIC and School of Marine Science and Technology, Northwestern Polytechnical University
[2] Audio Analysis Lab, AD:MT, Aalborg University, Aalborg, Denmark
[3] University of Oldenburg, Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany

ABSTRACT

In this paper, we investigate robust beamforming techniques for wideband signal processing in noisy and reverberant environments. In such environments, steering vector estimation errors are inevitable, leading to a degradation of the beamformer performance. Here, we study two types of beamformers that are robust against steering vector estimation errors. The first type includes robust Capon beamformers, where the underlying principle is to add a steering vector uncertainty constraint and/or a norm constraint to the optimization problem to improve the beamformer's robustness. The second type is the amplitude and phase estimation method, which utilizes both temporal and spatial smoothing. Experiments are presented to demonstrate the performance of the considered robust beamformers in acoustic environments. The results show that the robust beamformers outperform the non-robust beamformers in terms of predicted speech quality and intelligibility for different steering vector and covariance matrix estimation errors.

Index Terms— Microphone array, Capon beamforming, steering vector error, robust beamforming, APES beamforming.

1. INTRODUCTION

Beamforming [1–5] can be used to solve many acoustic problems and is, therefore, an important topic of research in the field of acoustic signal processing. Traditional beamforming methods, such as the data-independent delay-and-sum (DAS) beamformer and the standard data-dependent Capon beamformer (SCB) [6], introduce speech distortion if there are steering vector estimation errors. SCB, also known as the minimum power distortionless response (MPDR) beamformer, has been shown to be more sensitive to steering vector estimation errors than the minimum variance distortionless response (MVDR) beamformer (using the noise covariance matrix) [7–9]. However, accurately estimating the noise covariance matrix is not trivial. Herein, we focus on the beamformers based on the noisy signal covariance matrix, and seek to increase their robustness against the steering vector errors and covariance matrix estimation errors by generalizing techniques from robust narrowband beamforming to broadband speech scenarios.

Robust beamforming [10–16] has been widely investigated in narrowband signal processing, e.g., in radar and sonar applications. Based on the SCB, in [10–13] several robust beamformers, namely the norm constraint Capon beamformer (NCCB), the robust Capon beamformer (RCB) and the double-constraint robust Capon beamformer (DCRCB) have been derived to estimate the spectrum of the desired signal. By adding steering vector uncertainty and/or norm constraints to the traditional Capon method, these approaches give an accurate power estimate of the desired signal and a better performance than the SCB in terms of suppressing the interferences. Alternatively, another promising method is the amplitude-and-phase-estimation (APES) beamformer [17–21], which uses both the temporal as well as spatial averaging to obtain an estimate of the noise covariance matrix. When processing wideband signals, such as speech, the narrowband robust Capon and APES beamformers can be applied at each frequency bin in the STFT domain to form a wideband beamformer [22–24]. For example, in [22] the RCB was studied for processing speech signals, but only two channels and a simple alphabetical task were considered in the simulations.

Although robust beamforming has been extensively studied in a variety of narrowband applications, its application in wideband acoustic signal processing is not very common. In this paper, we hence study and experimentally compare the performance of several types of wideband robust beamforming algorithms for multichannel speech processing. On the one hand, we investigate the robustness against steering vector estimation errors by imposing uncertainty and norm constraints to the traditional Capon beamformer. On the other hand, we consider the APES method to deal with covariance matrix estimation errors.

2. SIGNAL MODEL

Let us consider the acoustic scenario with a single source located in the far field and a uniform linear array (ULA) consisting of M omnidirectional microphones to pick up the signal radiated from the source. If we neglect reverberation (this will be considered in the simulations in Section 4), the signal received by the mth microphone ($m = 1, 2, \ldots, M$) can be written as

$$y_m(t) = x_m(t) + v_m(t) = x(t - \tau_m) + v_m(t), \qquad (1)$$

where $y_m(t)$, $x_m(t)$, and $v_m(t)$ are the time domain noisy observation, the clean speech, and the additive noise (containing both stationary background noise and interferences), respectively, $x(t)$ denotes the clean speech received at the first microphone, and $\tau_m = (m-1)\tau_0$ is the time difference of arrival (TDOA) between the mth and the first microphone, $\tau_0 = (\delta \cos\theta_{\mathrm{d}})/c$ with δ being the spacing between neighboring microphones, θ_{d} denoting the DOA of the desired signal and $c = 340$ m/s being the speed of sound in air. We assume that all the signals are real-valued, zero-mean and wideband sequences, and the clean speech and the additive noise are uncorrelated. Without loss of generality, we choose $x(t)$ as the desired signal.

In the short-time-Fourier-transform (STFT) domain, (1) can be

This work was supported in part by the Villum Foundation and the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005. The work of Y. Zhao was supported in part by the China Scholarship Council.

written 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,ω) denote frame index and frequency index, respectively.

In vector form, (2) can be written as

$$\begin{aligned} \mathbf{y}(n,\omega) &= [Y_1(n,\omega),\ Y_2(n,\omega),\ \ldots,\ Y_M(n,\omega)]^T \\ &= \mathbf{a}(n,\omega)X(n,\omega) + \mathbf{v}(n,\omega), \end{aligned} \tag{3}$$

where $\mathbf{a}(n,\omega) \triangleq [1,\ e^{-j\omega\tau_0},\ \cdots,\ e^{-j(M-1)\omega\tau_0}]^T$ is the steering vector of the desired signal, and $[\cdot]^T$ denotes the transpose operator. The covariance matrix of $\mathbf{y}(n,\omega)$ is defined as

$$\mathbf{R}_{\mathbf{y}}(n,\omega) \triangleq E[\mathbf{y}(n,\omega)\mathbf{y}^H(n,\omega)], \tag{4}$$

where $E[\cdot]$ denotes mathematical expectation, and $[\cdot]^H$ denotes the conjugate-transpose operator. In practice, $\mathbf{R}_{\mathbf{y}}(n,\omega)$ is estimated by using a short-time average of the sample vectors, i.e.,

$$\widehat{\mathbf{R}}_{\mathbf{y}}(n,\omega) = \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{y}(n-k,\omega)\mathbf{y}^H(n-k,\omega), \tag{5}$$

where N is the number of recent frames. The mismatch between (4) and (5) may significantly degrade the beamforming performance.

By applying a spatial filter $\mathbf{h}(n,\omega)$ to the noisy observation in (3), we get

$$Z(n,\omega) = \mathbf{h}^H(n,\omega)\mathbf{y}(n,\omega) = X_{\mathrm{fd}}(n,\omega) + V_{\mathrm{rn}}(n,\omega), \tag{6}$$

where $Z(n,\omega)$ is an estimate of the desired signal, 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 vectors $\mathbf{x}(n,\omega)$ and $\mathbf{v}(n,\omega)$ are defined similarly to (3). With the beamforming model given in (6), the objective of beamforming is then to find an optimal filter $\mathbf{h}(n,\omega)$ so that $Z(n,\omega)$ is a good estimate of $X(n,\omega)$, which will be discussed in the next section.

3. ROBUST BEAMFORMERS

In this section, we review the narrowband SCB beamformer and its robust versions as well as the APES beamformer to estimate the desired signal $X(n,\omega)$. To process wideband speech signals, these narrowband beamformers will be applied at each frequency bin.

3.1. Standard Capon beamformer (SCB)

The SCB is obtained by solving the following constrained optimization problem [6]:

$$\min_{\mathbf{h}(\omega)}\ \mathbf{h}^H(\omega)\mathbf{R}_{\mathbf{y}}(\omega)\mathbf{h}(\omega)\ \text{ s.t. }\ \mathbf{h}^H(\omega)\overline{\mathbf{a}}(\omega) = 1, \tag{7}$$

where $\overline{\mathbf{a}}(\omega)$ is an estimate of the steering vector $\mathbf{a}(\omega)$. Note that the time frame index in the above formulation is omitted for ease of presentation. The solution of (7) is given by

$$\mathbf{h}(\omega) = \frac{\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\overline{\mathbf{a}}(\omega)}{\overline{\mathbf{a}}(\omega)^H\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\overline{\mathbf{a}}(\omega)}. \tag{8}$$

3.2. Norm constraint Capon beamformer (NCCB)

To limit the amplification of spatially white noise, i.e., the white noise gain (WNG), it has been proposed in [25] to add a norm constraint to (7), which also improves the robustness of the SCB against steering vector estimation errors. The corresponding constrained optimization problem hence becomes

$$\begin{aligned} \min_{\mathbf{h}(\omega)}\ \mathbf{h}^H(\omega)\mathbf{R}_{\mathbf{y}}(\omega)\mathbf{h}(\omega)\ \text{ s.t. }\ \mathbf{h}^H(\omega)\overline{\mathbf{a}}(\omega) = 1 \\ \|\mathbf{h}(\omega)\|^2 \le \zeta, \end{aligned} \tag{9}$$

where ζ is a parameter, which will be discussed in Section 4. The solution of (9) is given by:

$$\mathbf{h}(\omega) = \frac{[\mathbf{R}_{\mathbf{y}}(\omega) + \lambda\mathbf{I}]^{-1}\overline{\mathbf{a}}(\omega)}{\overline{\mathbf{a}}(\omega)^H[\mathbf{R}_{\mathbf{y}}(\omega) + \lambda\mathbf{I}]^{-1}\overline{\mathbf{a}}(\omega)}, \tag{10}$$

where λ is the so-called diagonal loading parameter, which is related to ζ (see also [12, 25]), and $\mathbf{I}$ denotes the identity matrix of size $M \times M$.

3.3. Robust Capon beamformer (RCB) and double-constraint robust Capon beamformer (DCRCB)

Instead of using a norm constraint, another way to deal with the steering vector error is to add an uncertainty constraint to (7). Assuming that the steering vector belongs to the sphere $\|\mathbf{a}(\omega) - \overline{\mathbf{a}}(\omega)\|^2 \le \epsilon$, where ϵ is a control parameter. The RCB problem is given by [11, 13]

$$\min_{\mathbf{a}(\omega)}\ \mathbf{a}^H(\omega)\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\mathbf{a}(\omega)\ \text{ s.t. }\ \|\mathbf{a}(\omega) - \overline{\mathbf{a}}(\omega)\|^2 \le \epsilon. \tag{11}$$

The robustness of the RCB can be further increased by introducing an additional norm constraint to the estimated steering vector $\mathbf{a}(\omega)$ in (11). This is also known as the double-constraint RCB (DCRCB), which is obtained by solving:

$$\begin{aligned} \min_{\mathbf{a}(\omega)}\ \mathbf{a}^H(\omega)\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\mathbf{a}(\omega)\ \text{ s.t. }\ \|\mathbf{a}(\omega) - \overline{\mathbf{a}}(\omega)\|^2 \le \epsilon \\ \|\mathbf{a}(\omega)\|^2 = M. \end{aligned} \tag{12}$$

The constrained optimization problems in (9), (11) and (12) are known as quadratically constrained quadratic programming problems (QCQP), which can be solved using the Lagrange multiplier method [10, 12, 13]. The solutions of the RCB and the DCRCB can be obtained in a similar way. By solving the optimization problems in (11) and (12), we first obtain an estimate of the steering vector, i.e., $\hat{\mathbf{a}}(\omega)$ [12]. By substituting this estimate into (8), the robust beamformers are obtained as

$$\mathbf{h}(\omega) = \frac{\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\hat{\mathbf{a}}(\omega)}{\hat{\mathbf{a}}(\omega)^H\mathbf{R}_{\mathbf{y}}^{-1}(\omega)\hat{\mathbf{a}}(\omega)}. \tag{13}$$

More details can be found in [12].

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

As an alternative approach to robust Capon beamformers, in this subsection we present the APES bemformer, where the microphone array is divided into several subarrays, and a spatial smoothing technique is then utilized to improve the robustness against steering vector and covariance matrix estimation errors.

Let $\overline{M} < M$ be an integer, which denotes the number of microphones in the subarray. For the lth subarray, the $\overline{M} \times 1$ vectors $\overline{\mathbf{a}}_l(n,\omega) = [e^{-jl\omega\tau_0},\ e^{-j(l+1)\omega\tau_0},\ \cdots,\ e^{-j(l+\overline{M}-1)\omega\tau_0}]^T$ and

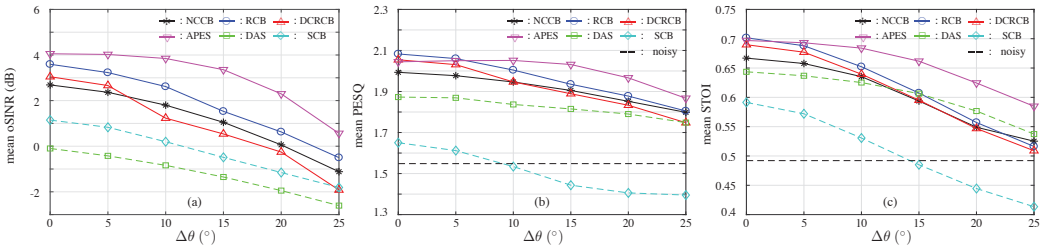

Figure 1: Performance with different DOA estimation errors, $M = 8$, $T_{60} \approx 150$ ms, iSNR $= 10$ dB and iSIR $= -5$ dB.

$\overline{\mathbf{y}}_l(n,\omega) = [Y_l(n,\omega),\ Y_{l+1}(n,\omega),\ \cdots,\ Y_{l+\overline{M}-1}(n,\omega)]^T$ denote the subarray steering vector and subarray observed signal vector, where $l = 0,\ldots,L-1$, and $L = M - \overline{M} + 1$. With the ULA assumption, the subarray steering vectors are related as

$$\overline{\mathbf{a}}_l(n,\omega) = e^{-jl\omega\tau_0}\overline{\mathbf{a}}_0(n,\omega). \tag{14}$$

Hence, using (14), $\overline{\mathbf{y}}_l(n,\omega)$ can be written as

$$\overline{\mathbf{y}}_l(n,\omega) = e^{-jl\omega\tau_0}\overline{\mathbf{a}}_0(n,\omega)X(n,\omega) + \overline{\mathbf{v}}_l(n,\omega). \tag{15}$$

Combining (15) with the APES principle, which aims to minimize the least-squares error between the beamformer output and the desired signal for each subarray [20], an estimate of $X(n,\omega)$ and the filter $\overline{\mathbf{h}}$ of length $\overline{M}$ can be obtained by solving the following problem

$$\min_{\overline{\mathbf{h}},X(n,\omega)} \sum_{k=0}^{N-1}\sum_{l=0}^{L-1} |\overline{\mathbf{h}}^H\overline{\mathbf{y}}_l(n-k,\omega)e^{jl\omega\tau_0} - X(n-k,\omega)|^2$$
$$\text{s.t. } \overline{\mathbf{h}}^H\overline{\mathbf{a}}_0(n,\omega) = 1. \tag{16}$$

Let $\mathbf{g}(n,\omega) = (1/L)\sum_{l=0}^{L-1}\overline{\mathbf{y}}_l(n,\omega)e^{jl\omega\tau_0}$, and notice that

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left|\overline{\mathbf{h}}^H\overline{\mathbf{y}}_l(n-k,\omega)e^{jl\omega\tau_0} - X(n-k,\omega)\right|^2 \\
&= \overline{\mathbf{h}}^H\left[\frac{1}{L}\sum_{l=0}^{L-1}\overline{\mathbf{y}}_l(n-k,\omega)\overline{\mathbf{y}}_l^H(n-k,\omega)\right]\overline{\mathbf{h}} \\
&- \overline{\mathbf{h}}^H\mathbf{g}(n-k,\omega)\mathbf{g}^H(n-k,\omega)\overline{\mathbf{h}} \\
&+ |X(n-k,\omega) - \overline{\mathbf{h}}^H\mathbf{g}(n-k,\omega)|^2.
\end{aligned} \tag{17}$$

Minimizing (17), we obtain an estimate of $X(n,\omega)$ as

$$Z(n,\omega) = \overline{\mathbf{h}}^H\mathbf{g}(n,\omega). \tag{18}$$

Substituting (18) into (16), the problem reduces to

$$\min_{\overline{\mathbf{h}}} \overline{\mathbf{h}}^H\hat{\mathbf{Q}}\overline{\mathbf{h}} \text{ s.t. } \overline{\mathbf{h}}^H\overline{\mathbf{a}}_0(n,\omega) = 1, \tag{19}$$

where

$$\begin{aligned}
\hat{\mathbf{Q}} =& \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{L}\sum_{l=0}^{L-1}\overline{\mathbf{y}}_l(n-k,\omega)\overline{\mathbf{y}}_l^H(n-k,\omega) \\
&- \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{g}(n-k,\omega)\mathbf{g}^H(n-k,\omega),
\end{aligned} \tag{20}$$

can be interpreted as an estimate of the noise covariance matrix. Note that $NL \geq M$ to ensure that $\hat{\mathbf{Q}}$ is positive-definite. The solution to (19) is then given by

$$\overline{\mathbf{h}} = \frac{\hat{\mathbf{Q}}^{-1}\overline{\mathbf{a}}_0(n,\omega)}{\overline{\mathbf{a}}_0^H(n,\omega)\hat{\mathbf{Q}}^{-1}\overline{\mathbf{a}}_0(n,\omega)}. \tag{21}$$

4. SIMULATIONS

Now, we study the performance of different robust beamformers with simulated room acoustics. Reverberation and different noise levels are considered. The signal model given in Section 2 neglects the effect of reverberation. Generally, the received signal in reverberant environment is approximately written as:

$$Y_m(n,\omega) = D_m(n,\omega)S(n,\omega) + V_m(n,\omega), \tag{22}$$

where $D_m(n,\omega)$ is the STFT of the room impulse response from the source $S(n,\omega)$ to the mth microphone. The signal model mismatch between (2) and (22) may dramatically degrade the beamforming performance if robustness is not taken into account in beamforming.

4.1. Experimental setup

The room impulse response is generated by using the image model of [26], with a room of size $5\text{ m} \times 5\text{ m} \times 3\text{ m}$. We consider a uniform linear array with 8 microphones located at the center of the room with δ being 0.04 m, and the desired signal which is 1 meter away from the array center propagates from the direction $\theta_\text{d} = 30°$. To simulate an acoustic multi-interference scene, we set 5 competing speakers at 82°, 127°, 230°, 281° and 325° respectively. Moreover, we also consider the DOA estimation error here, assuming the only information we know about the desired signal is an imprecise DOA: $\theta = \theta_\text{d} + \Delta\theta$, where $\Delta\theta$ models the DOA estimation error. The clean speech signal is taken from the TIMIT database [27] and downsampled to 8000 Hz in our experiments. The speech signals from ten different speakers are used. The microphone signal is generated by convolving the clean speech with the corresponding room impulse response. After that, both white Gaussian noise and convolved interferences are added to the desired speech. The signal is then transformed into the STFT domain with a 128-point FFT and the overlap between neighboring frames is 75%. The $\mathbf{R}_\mathbf{y}(n,\omega)$ matrix is estimated by using (5) with the most recent $N = 20$ frames for $M = 8$.

4.2. Performance measures

We will now present some performance measures to evaluate the aforementioned beamformers in Section 3. The output SINR, according to the model given in (6), is

$$\text{oSINR} = \frac{E[|X_\text{fd}(n,\omega)|^2]}{E[|V_\text{rn}(n,\omega)|^2]}, \tag{23}$$

Due to the spatial smoothing of APES method in (18), the definition of SINR is different from that of other methods. First, we rewrite (18) as $Z_\text{APES}(n,\omega) = X_\text{fd,APES}(n,\omega) + V_\text{rn,APES}(n,\omega)$, where $X_\text{fd,APES}(n,\omega) = \overline{\mathbf{h}}^H(1/L)\sum_{l=0}^{L-1}\overline{\mathbf{x}}_l(n,\omega)e^{jl\omega\tau_0}$ and

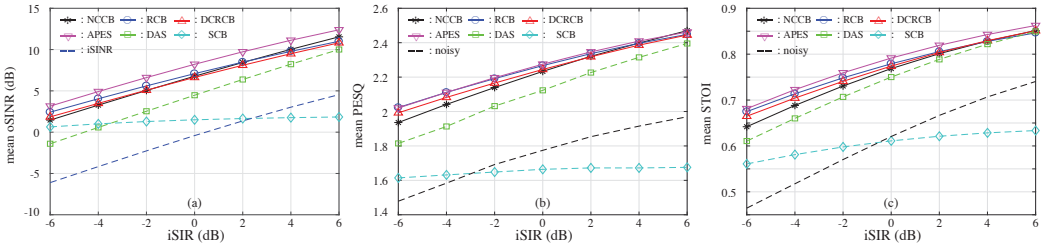

Figure 2: Performance with different input SIR, $M = 8$, $T_{60} \approx 150$ ms, iSNR $= 10$ dB and $\Delta\theta = 5^\circ$.

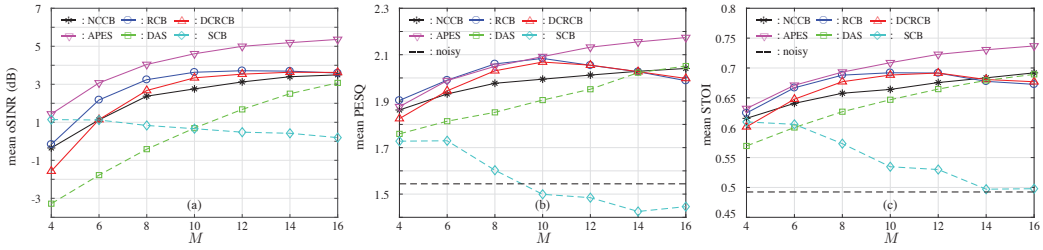

Figure 3: Performance with different M, $T_{60} \approx 150$ ms, iSNR $= 10$ dB, iSIR $= -5$ dB and $\Delta\theta = 5^\circ$.

$V_{\mathrm{rn,APES}}(n,\omega) = \overline{\mathbf{h}}^H(1/L)\sum_{l=0}^{L-1}\overline{\mathbf{v}}_l(n,\omega)e^{jl\omega\tau_0}$. The output SINR for the APES beamformer is then defined as

$$\mathrm{oSINR}[\overline{\mathbf{h}}] = \frac{E[|X_{\mathrm{fd,APES}}(n,\omega)|^2]}{E[|V_{\mathrm{rn,APES}}(n,\omega)|^2]}. \quad (24)$$

Besides evaluating the output SINR, we also use PESQ [28] and STOI [29] as performance measures. The speech distortion measures are more complicated, which will be discussed in a separate study.

4.3. Simulation results

In the first experiment, the influence of the DOA estimation error on the performance is studied. It is worth noting that for different amounts of estimation errors, the smallest possible uncertainty parameter ϵ for RCB and DCRCB should be selected carefully. Here the proper values of those parameters are chosen based on experiments, which are listed in Table 1. For NCCB, we set $\beta_o = 1.5/M$, and for APES, the subarray microphone number is chosen as $M_o = M - 1$. As shown in Fig. 1, with the increasing of the DOA estimation error, the performance decreases for all the beamforming methods. But the robust beamformers outperforms the traditional method in most conditions. Furthermore, the APES beamformer yields good performance with small amount of DOA estimation errors. While the SCB dose not perform well even without DOA errors. This is caused by the mismatch between $\mathbf{R}_{\mathbf{y}}(n,\omega)$ and $\hat{\mathbf{R}}_{\mathbf{y}}(n,\omega)$. Note that even though the APES beamformer always shows better performance in terms of improving the oSINR, the definition of oSINR for APES is different from that for the other methods as seen in (23) and (24).

Table 1: Parameters for RCB and DCRCB

$\Delta\theta$	0°	5°	10°	15°	20°	25°	30°
RCB (ϵ_o)	0.5	0.5	0.5	1	1	1	1.5
DCRCB (ϵ_o)	0.5	0.5	1	1	1	1.5	1.5

In the second experiment, the performance with different input SIRs are studied. Fig. 2 shows that the robust beamformers outperform DAS and SCB methods both in improving the speech quality and speech intelligibility. The APES method gives better performance under most input SIR conditions, while the other robust methods behave similarly. However, the results illustrate that SCB even degrades the signal quality sometimes. 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 higher level of sidelobes.

The last experiment studies the performance versus different number of microphones. The proper short-time average length for different M is set according experiments. As shown in Fig. 3, the performance of RCB and DCRCB first improves with the increase of microphone number and then begins to decrease after M reaches 10. This is due to the estimation error of the covariance matrix, which increases with the number of microphones. Additionally, this also explains why the performance of SCB deteriorates with increasing number of microphones.

The simulation results indicate that the robust beamformers perform better than the traditional methods in reverberant environments. Among the studied robust methods, APES beamformer has the potential to further improve the speech quality and speech intelligibility with large number of microphones. Moreover, RCB shows slightly better performance than NCCB and DCRCB. 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 improved.

5. CONCLUSION

In this paper, we studied different robust adaptive beamformers for wideband acoustic signal processing. Experiments were performed in reverberant environments with multiple interference sources. The results illustrated that these methods are able to improve the robustness of the beamformer against the estimation errors of steering vector and covariance matrix. Furthermore, these robust adaptive beamformers maintain robustness against the signal model mismatch in reverberant environments. Interestingly, the APES beamformer shows better performance than the other studied methods in terms of signal distortion, which maintain a low level of signal distortion even when there is estimation errors in the steering vector and the signal covariance matrix.

6. REFERENCES

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, 1988.

[2] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2511–2526, 2003.

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

[4] C. Pan, J. Chen, and J. Benesty, "Performance study of the MVDR beamformer as a function of the source incidence angle," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 22, no. 1, pp. 67–79, 2014.

[5] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 24, no. 3, pp. 543–558, 2016.

[6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[7] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc. Amer.*, vol. 54, no. 3, pp. 771-785, 1973.

[8] H. L. Van Trees, *Detection, estimation, and modulation theory*. New York, Wiley, 2004.

[9] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of MVDR and MPDR beamformers," *26th Conv. Electr. Electron. Eng. in Israel (IEEEI)*, pp. 416-420, 2010

[10] J. Li and P. Stoica, *Robust Adaptive Beamforming*. New York, Wiley, 2005.

[11] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, 2003.

[12] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407–2423, 2004.

[13] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, 2003.

[14] S. A. Vorobyov, A. B. Gershman, Z. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 108–111, 2004.

[15] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, 2005.

[16] A. Mukherjee and A. L. Swindlehurst, "Robust beamforming for security in MIMO wiretap channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 351–361, 2011.

[17] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, 1996.

[18] Z.-S. Liu, H. Li, and J. Li, "Efficient implementation of Capon and APES for spectral estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1314–1319, 1998.

[19] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," *IEEE Signal Process. Lett.*, vol. 6, no. 8, pp. 205–206, 1999.

[20] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.

[21] M. G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5969–5983, 2010.

[22] K. Ngo, "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.

[23] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[24] Q. Zou, Z. L. Yu, and Z. Lin, "A robust algorithm for linearly constrained adaptive beamforming," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 26–29, 2004.

[25] H. Cox, R. M. Zeskind, and M. H. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, 1987.

[26] E. A. P. Habets, "Room impulse response generator," Tech. Rep., Technische Universiteit Eindhoven, 2006.

[27] J. Garofolo, "DARPA TIMIT acoustic-phonetic continuous speech corpus," Gaithersburg, MD, USA: Nat. Inst. of Standards Technol. 1993.

[28] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 16, no. 1, pp. 229–238, 2008.

[29] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.