PERFORMANCE OF ADAPTIVE BEAMFORMERS FOR EXTRACTING AUDIO SIGNALS

Chaoying Bao

Maritime Division
Defence Science and Technology Group, Department of Defence, Australia
Email: chaoying.bao@dsto.defence.gov.au

Abstract

The need to extract a single audio signal of interest from a multi-source and noisy environment is common across many disciplines. Adaptive beamforming, due to its superior interference rejection and noise suppression, is a preferred processing technique for obtaining high quality audio in noisy environments. In this paper, we compare the performances of two different types of adaptive beamformers for the purpose of extracting audio signals. One of the beamformers is the robust Capon beamformer (RCB) where the beamforming part is carried out in the frequency domain, resulting in low computational complexity but relatively high latency. The other beamformer is the block constrained least mean square beamformer (BCLMSB) where the beamforming part is carried out in the time domain. It has relatively higher computational cost but no latency. The performances of these two adaptive beamformers are evaluated in terms of fidelity of the beamformer output and robustness of the system under various conditions.

1. Introduction

Adaptive beamforming is a preferable technique over its non-adaptive counterpart for extracting a single audio signal of interest from a multi-source and noisy environment due to its far superior interference rejection and noise suppression. In an earlier study [1], we have examined the performance of so called time-frequency domain adaptive beamformers where the beamforming part is carried out in the frequency domain. The main motivation of using time-frequency domain adaptive beamformers is that their computational complexity can be ten to hundreds times smaller than that of direct time domain adaptive beamformers. Two widely used frequency domain adaptive beamformers have been evaluated and compared. One is the so called robust Capon beamformer (RCB) and the other is a standard minimum variance distortionless response (MVDR) beamformer. The RCB has been found to outperform the standard MVDR both in signal fidelity and in robustness at expense of slightly higher computational cost. In a later study [2], we have compared the performance of several time domain adaptive beamformers where the whole process including adaptive beamforming is all carried out in the time domain. The main advantage of time domain adaptive beamformers over time-frequency domain beamformers is that they have little or no latency (which is the delay time between output and input of the beamformer). Therefore, they are widely used in the telecommunication field where latency normally cannot be tolerated. In the study, the tapped delay line (TDL) structure has been considered for time domain adaptive beamforming. Three different adaptive algorithms have been used for obtaining the optimal TDL filters. These are the sample matrix inversion method, the recursive least squares method with sliding window, and the block constrained least mean square method with diagonal loading. The block constrained least mean square beamformer (BCLMSB) has been found to give a good performance in signal fidelity with the least computational complexity. In this paper, we will compare the performances of the winners of the two previous studies, which are the RCB and the BCLMSB, for extracting a moving audio signal of interest from a multi-source and noisy environment in the situation where the signal model is inaccurate. Various conditions that may affect the performance are considered. These include the effect of bearing rate of moving sources, the effect of bearing separation of interference and signal sources, and the effect of steering direction of the beamformers.

2. Beamformers

2.1 Robust Capon beamformer

In a time-frequency domain beamformer, the time series output of each array element is divided into blocks. Each block of data is Fourier transformed, and the results for each array element in the same frequency bin f are combined into an array output (which is also beamforming input) vector $\mathbf{x}(f)$. The beamforming output in the frequency domain at frequency bin f is given by

$$y(f) = \mathbf{w}(f)^H \mathbf{x}(f), \tag{1}$$

where $\mathbf{w}(f)$ is a vector of complex weights given by a beamforming algorithm, and superscript ^{H} denotes the conjugate transpose. After $y(f)$ has been computed at each frequency bin, the time domain signal $y(t)$ is obtained by the inverse Fourier transformation.

One method for calculating $\mathbf{w}$ (for simplicity we henceforth drop index f) is the Robust Capon Beamforming (RCB) scheme proposed by Li et al. [3]. Let $\mathbf{x}_n$ denote the array output vector for the n^{th} block, and $\hat{\boldsymbol{R}}$ the sample estimate of the cross spectral matrix,

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n^H . \tag{2}$$

Let $\mathbf{v}$ denote the theoretical signal response vector obtained from the signal model of the array; this is the array output vector when a signal arrives from the look-direction in the absence of noise. The RCB algorithm anticipates possible errors in the signal model. The actual signal response vector, $\mathbf{v}_{\text{a}}$, is estimated by solving the following quadratic problem,

$$\min_{\mathbf{v}_a} \mathbf{v}_a^H \hat{\mathbf{R}}^{-1}\mathbf{v}_a \quad \text{subject to} \left\|\mathbf{v}_a - \mathbf{v}\right\|^2 \le \varepsilon , \tag{3}$$

where ε is a small positive number proportional to the signal mismatch. Once $\mathbf{v}_{\text{a}}$ is obtained (see [3] for details), the RCB solution for $\mathbf{w}$ can be computed using the formula:

$$\mathbf{w}_{RCB} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{v}_a}{\mathbf{v}_a^H\hat{\mathbf{R}}^{-1}\mathbf{v}_a} . \tag{4}$$

As explained in [3], the RCB approach belongs to the class of extended diagonally loaded MVDR beamformers. The degree of loading is determined by value of ε. It should be noted that the original purpose of the RCB method is to make the algorithm robust to errors in the signal model, and the appropriate value ε is determined by the anticipated signal mismatch. For audio applications, however, loading is introduced for an additional purpose [1]. Some degree of loading is always required in audio processing to limit the so called white noise gain (WNG) [4] of the beamformer, even in the absence of any signal mismatch.

2.2 Block constraint least mean square beamformer with diagonal loading

The BCLMSB is based on one of the most commonly used time domain adaptive beamforming structures. It employs banks of adaptive finite impulse response (FIR) filters, or a tapped delay line (TDL) structure, between the delays and summation point of the well-known conventional delay-and-sum beamformer, as illustrated in Figure 1. The coefficients or weights of those FIR filters are adapted to minimise the noise and interference from non-look-directions. Following the notation in Frost [5] and referring to Figure 1, we have the beamforming input vector,

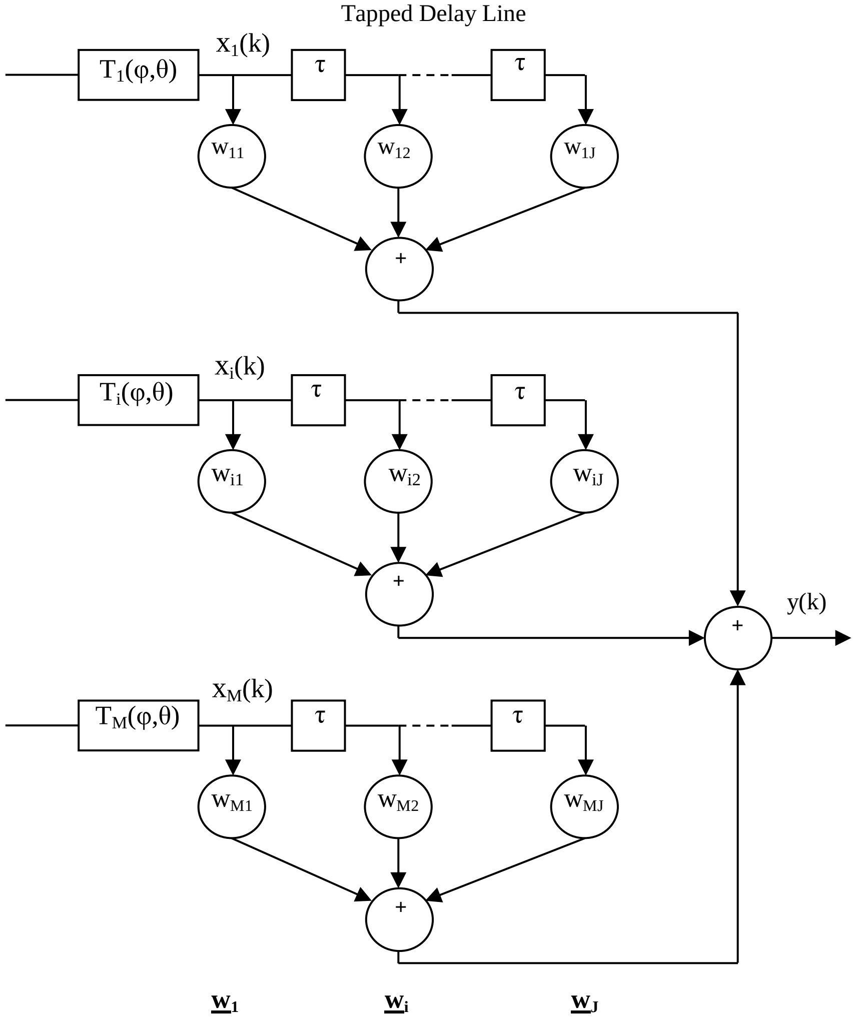

Figure 1. Adaptive beamformer structure using TDL filter

$$\mathbf{X}^T(k) = [x_1(k),\ x_2(k),\ ...,\ x_{MJ}(k)], \tag{5}$$

the weight vector,

$$\mathbf{W}^T(k) = [w_1(k),\ w_2(k),\ ...,\ w_{MJ}(k)], \tag{6}$$

and the covariance matrix,

$$\mathbf{R}_{XX} = E[\mathbf{X}(k)\mathbf{X}^T(k)], \tag{7}$$

where $E[\ ]$ denotes the statistical expectation. The output of the beamformer is

$$y(k) = \mathbf{X}^T(k)\mathbf{W}. \tag{8}$$

The optimal weights are the solution of the following minimization problem:

$$\underset{\mathbf{W}}{\text{minimise}} \ \mathbf{W}^T \mathbf{R}_{XX} \mathbf{W} \tag{9}$$

$$\text{subject to } \mathbf{C}^{\mathrm{T}}\mathbf{W} = \mathbf{F}_0, \tag{10}$$

where **C** is the constraint matrix and $\mathbf{F}_0$ is the vector specifying the frequency response in the look-direction. Thus

$$\mathbf{W}_{opt} = \mathbf{R}_{XX}{}^{-1}\mathbf{C}[\mathbf{C}^T\mathbf{R}_{XX}{}^{-1}\mathbf{C}]^{-1}\mathbf{F}_0. \tag{11}$$

In order to avoid an excessive computational burden of matrix inversions in Eq. (11), Frost [5] developed a simple stochastic gradient-descent algorithm called the constrained LMS (CLMS) algorithm where the weights of the adaptive beamformer can be iteratively updated to converge to $\mathbf{W}_{opt}$ via the following equations:

$$\mathbf{W}(0) = \mathbf{F}, \tag{12}$$
$$\mathbf{W}(k+1) = \mathbf{P}[\mathbf{W}(k) - \mu\, y(k)\mathbf{X}(k)] + \mathbf{F}, \tag{13}$$
$$\mathbf{F} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{F}_0, \tag{14}$$
$$\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T, \tag{15}$$

where μ is the step size for regulating the convergence rate of the algorithm.

Including diagonal loading in CLMS is straight forward. Eq. (13) can also be written as

$$\mathbf{W}(k+1) = \mathbf{P}[\mathbf{W}(k) - \mu\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{W}(k)] + \mathbf{F}. \tag{16}$$

Substituting instantaneous estimation of covariance matrix $\mathbf{X}(k)\mathbf{X}^T(k)$ with its loaded version of $\mathbf{X}(k)\mathbf{X}^T(k)+\sigma\mathbf{I}$ gives

$$\mathbf{W}(k+1) = \mathbf{P}[\mathbf{W}(k) - \mu(\mathbf{X}(k)\mathbf{X}^T(k) + \sigma\mathbf{I})\mathbf{W}(k)] + \mathbf{F}, \tag{17}$$

where σ is the loading coefficient. Eq. (17) can further be simplified as

$$\mathbf{W}(k+1) = \mathbf{P}[(1-\mu\sigma)\mathbf{W}(k) - \mu\, y(k)\mathbf{X}(k)] + \mathbf{F}. \tag{18}$$

Derivation of a block version of Eq. (18) is also straight forward. By replacing $y(k)\mathbf{X}(k)$ with its time average, we finally have the BCLMSB in which the equation for updating $\mathbf{W}(k)$ is

$$\mathbf{W}(k+1) = \mathbf{P}[(1-\mu\sigma)\mathbf{W}(k) - \mu\sum_{i=0}^{N-1} y(kN+i)\mathbf{X}(kN+i)] + \mathbf{F}, \tag{19}$$

where N is the block length. It should be noted that k in Eq. (19) stands for kth block instead of kth sample.

3. Performance evaluation

In this section, the performances of the two beamformers are evaluated in the presence of signal mismatch and with moving sound sources. The evaluation is based on the coherence between the beamformer output and the signal of interest, as well as the robustness of the beamformers. The coherence γ is based on the correlation coefficient, defined in a way that accounts for a time delay between the signal *s(t)* and the beamformer output *y(t)*,

$$\gamma = \max_{\tau} \frac{|\mathrm{COV}\{s(t)y(t-\tau)\}|}{\sqrt{\mathrm{COV}\{s^2(t)\}\mathrm{COV}\{y^2(t-\tau)\}}}. \qquad (20)$$

Here COV{} denotes the covariance. Note that γ has a value between 0 and 1, with value 0 indicating no correlation between the signal and beamformer output, and value 1 indicating that the two waveforms are proportionally identical.

The array considered is a uniform linear array (ULA) consisting of 32 sensors with an inter-element spacing of 0.75m. The speed of sound is taken as 1500 m/s. As a result, the design frequency of the array is 1000 Hz. The sampling frequency is chosen as 4000 Hz.

The frequency resolution (binwidth) in the RCB is 8 Hz. If the frequency resolution is too low, adequate beamforming in the frequency domain is not achieved, resulting in a lower coherence. On the other hand, a finer frequency resolution demands longer integration time for obtaining the cross spectral matrix, which in the case of moving sound sources results in a poor estimate of that matrix and hence again a lower coherence. The choice of 8 Hz is found to be a good compromise in the simulations. The latency resulted from using this frequency resolution is about 3 seconds. The bearing resolution of the RCB is 1°, which limits the maximum of the look direction mismatch (LDM) of the beamformer to 0.5°.

The FIR filters in the BCLMSB have 33 taps. A larger number of taps will always result in a higher coherence but is accompanied by a greater computational cost. Therefore, a trade-off between these two is made, with 33 taps achieving a high coherence for the given sound sources while maintaining reasonable computational complexity. In order to increase the bearing resolution of the beamformer, the interpolation technique described in [6] is used. The interpolation factor is chosen as 30, which gives bearing resolutions ranging from about 1° to 10° and the corresponding maximum of LDM between about 0.5° and 5°, depending of the steering direction.

One signal and two interference sound sources are in far field to the array. They all emit sound waves consisting of four tonal components. The tonal frequencies of the signal are chosen to be close to 200, 400, 600 and 800 Hz. These frequencies are selected to simulate a broad range of low, mid and high frequencies, relative to the design frequency of the array. The tonal frequencies of the interferences are shifted 20 Hz away from those of the signal, one to the lower end and the other to the higher end. As there is a performance difference in the RCB depending on whether those frequencies of sound sources are on or off the FFT frequencies [1], the tonal frequencies are randomly chosen within one binwidth around 200, 400, 600 and 800 Hz in each simulation run in order to have a fair comparison between the performances of the RCB and the BCLMSB. The phases of the tonal components in the signal and interferences are also randomly selected at the start of each simulation run. A uniformly distributed random noise of 0 dB relative to the signal is added at each sensor output to simulate all the other noises in the system. The signal source moves in a manner that results in a constant bearing rate to the array. The powers of the interferences are 10 dB higher than that of the signal. The signal model error, or signal mismatch, is realised by applying certain percentage of random deviations to the values of each sensor output. In this study, 30% of signal mismatch is applied.

Table 1. Latency and computational complexity of the two beamformers

	RCB	BCLMSB
Latency (sec)	3.008	0
Complexity ratio	1	39

The results shown in the following subsections are all obtained by averaging 200 simulation runs. All the adjustable parameters of the two beamformers, such as the user parameter in the RCB, and step size μ, loading coefficient σ, and block length N in the BCLMSB, are optimally tuned. Table 1 compares the latency and computational complexity of the two beamformers under those optimally tuned parameters.

3.1 Effect of bearing rate of the moving sources

As the adaptive algorithms of both RCB and BCLMSB are developed on the assumption of stationary sources, it is important to examine the robustness of both beamformers to the moving sources with different bearing rates.

In this subsection, the signal source moves around the broad side of the array with a bearing rate ranging from 0.01 to 0.8 degree/second. The two inference sources move around at the bearings of 30° and -30°, respectively, with a constant bearing rate of 0.05 degree/second.

Figure 2 shows the time history of the coherences achieved by the two beamformers in a scenario where the signal source moves from -0.9° to 0.9° with a bearing rate of 0.2 degree/second. It can be seen that the trends of the coherence of the two beamformers are similar. They both roughly follow the change of LDM. For example, the coherence reaches to its peak at about 4.5 second when LDM becomes zero, and the coherence goes lower as LDM peaks at around 2 and 7 seconds. This indicates that both beamformers are robust to the moving source at this bearing rate. The time averaged coherences of the two beamformers in this scenario are very similar.

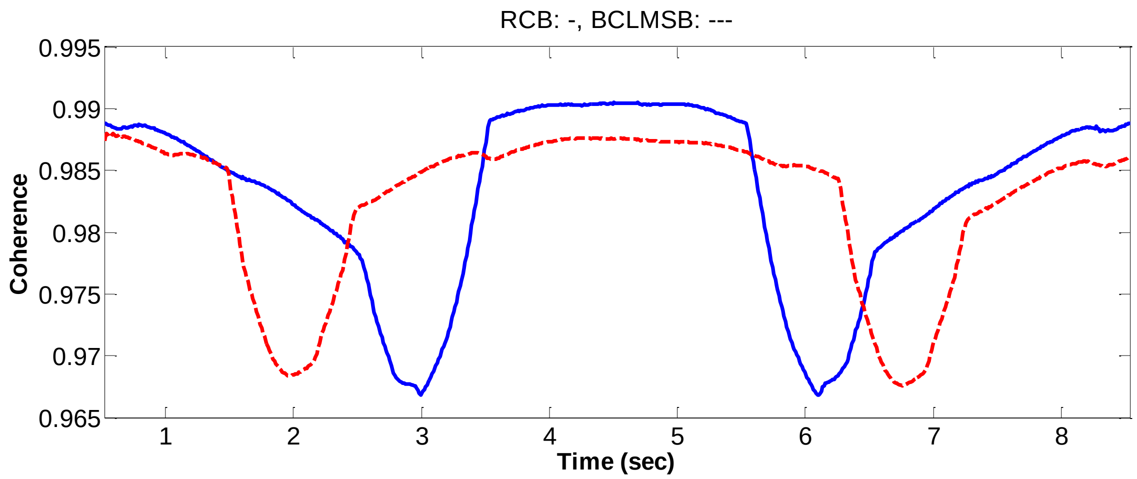

Figure 2. Time history of coherence of RCB and BCLMSB, with the signal source moving from -0.9° to 0.9° at a bearing rate of 0.2 degree/second. The time averaged coherences of RCB and BCLMSB are 0.9832 and 0.9831, respectively.

Figure 3 plots the time history of the coherences of the two beamformers in a scenario where the signal source moves from -1.2° to 1.2° with a bearing rate of 0.4 degree/second. The trends of the coherence of the two beamformers are different in this scenario. The coherence of the BCLMSB still roughly follows the change of LDM. The coherence peaks when there is no LDM and goes down as LDM increases. The coherence of the RCB, however, does not follow the change of LDM anymore. For example, the coherence drops to its lowest level at about 3 seconds when there is no LDM. This indicates that while the BCLMSB is still robust to the moving source in this scenario, the RCB is not. The coherence level of the BCLMSB is very similar to that of the scenario in Figure 2. But for the RCB, there is a period of time (between 2.5 and 3.5 seconds) where the coherence level drops below 0.96 and down to 0.91(comparing to Figure 2 where the coherence level is above 0.965 at all times).

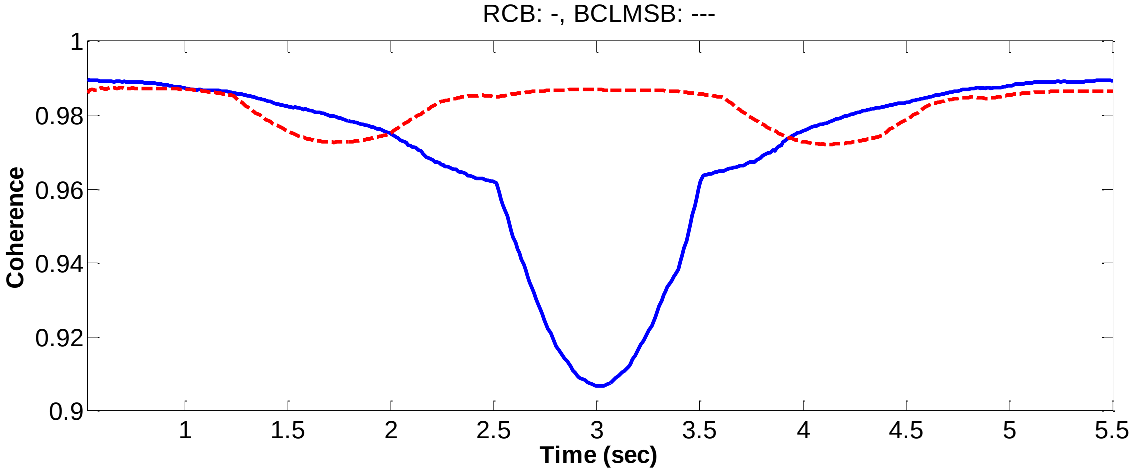

Figure 3. Time history of coherence of RCB and BCLMSB, with the signal source moving from -1.2° to 1.2° at a bearing rate of 0.4 degree/second. The time averaged coherences of RCB and BCLMSB are 0.9703 and 0.9825, respectively.

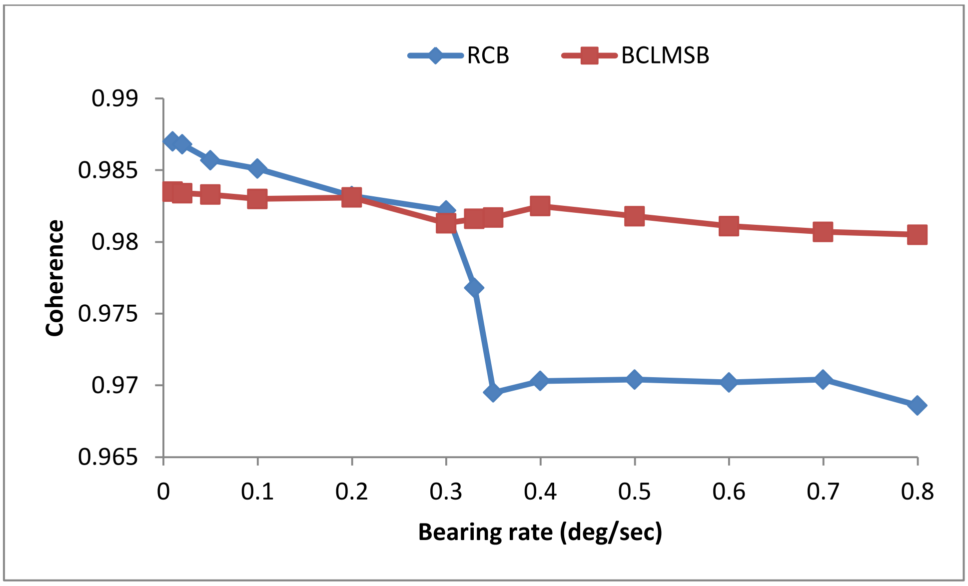

Figure 4. The effect of bearing rate of the moving signal source on the coherence achieved by RCB and BCLMSB.

Figure 4 plots the time averaged coherence achieved by the two beamformers against the bearing rate of the moving signal source. It can be seen that the performance of the BCLMSB is quite robust to the bearing rate of the moving source. Its coherence drops very little as the bearing rate increases. The performance of the RCB, on the other hand, is sensitive to the bearing rate of the moving source. It appears that there exists a turning point for the RCB. When the bearing rate is slower than 0.3 degree/second, its coherence is slightly higher than that of the BCLMSB. But when the bearing rate is higher than 0.3 degree/second, its coherence is noticeably lower than that of the BCLMSB. The reason for this different robustness to the bearing rate between the two beamformers can be explained as follows. In the RCB, a certain time (known as integration time, reflected by N in Equation (2)) is required to have a workable estimate of the cross spectral matrix $\mathbf{R}$. For the system set up in this study, the integration time needed is about 3 seconds. If the bearing rate is greater than 0.34 degree/second, the signal source will have moved more than 1° (which is the bearing resolution of the RCB) during one integration time. This severely violates the stationary assumption and degrades the beamforming performance. As for the BCLMSB, the integration or block time (reflected by N in Equation (19)) is much shorter (0.001 second for the system setup in this study). Consequently, the BCLMSB has a much better robustness to the moving sources with high bearing rate. In conclusion, the rule of thumb of keeping high level of coherence in the situation of moving sources is that the product of the bearing rate of the signal source times the integration time of the beamformer should be smaller than the bearing resolution of the beamformer.

3.2 Effect of bearing separation between interference and signal sources

In this subsection, we examine the effect of bearing separation between interference and signal sources on the performances of the two beamformers. In simulations, the signal source moves from -0.75° to 0.75° with a bearing rate of 0.1 degree/second. The two inference sources are symmetrically at different side of the signal sources, moving at a bearing rate of 0.1 degree/second. Different bearing separations between interference and signal sources are used, ranging from 5° to 60°.

Figure 5 shows the time averaged coherence achieved by the two beamformers against the bearing separation. It can be seen that both beamformers achieve high coherence if the bearing separation is greater than 15°. The coherence of the BCLMSB drops away much faster than that of the RCB when the bearing separation is smaller than 15°.

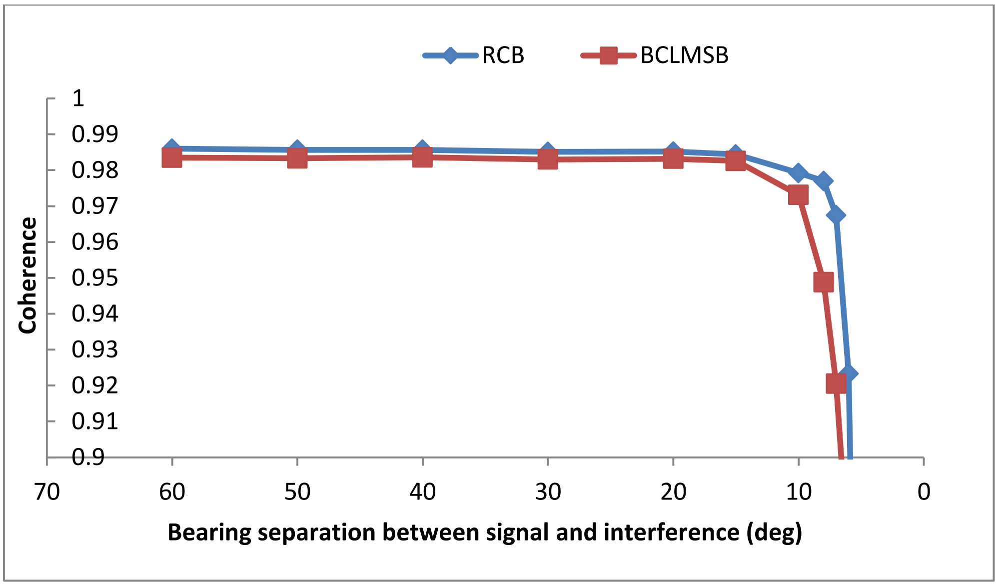

Figure 5. The effect of bearing separation on the coherence achieved by RCB and BCLMSB.

3.3 Effect of steering direction

It is well known that ULA performance will drop somewhat if the beamformer is steered away from broadside towards endfire. In this subsection, we examine the effect of steering direction on the performances of the two beamformers. In the simulation, the signal source moves from 0° to 60° with a bearing rate of 0.1 degree/second. The two inference sources are +15° and -15° from the signal source, respectively, and also move at a bearing rate of 0.1 degree/second.

Figure 6 shows the time averaged coherence achieved by the two beamformers against the steering direction. It can be seen that the performances of both beamformers are very similar for the steering directions from 0° to 55°. For the steering directions from 0° to 30°, both beamformers achieve high coherence above 0.98. For the steering directions greater than 40°, the coherences of both beamformers start to drop.

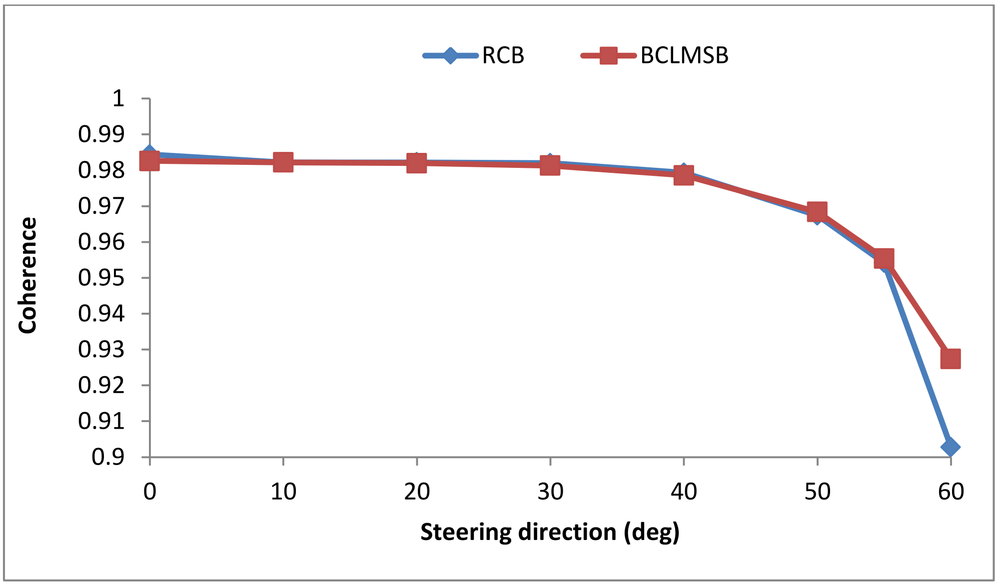

Figure 6. The effect of steering direction on the coherence achieved by RCB and BCLMSB.

4. Conclusions

We have examined the performances of the robust Capon beamformer (RCB) and the block constrained least mean square beamformer (BCLMSB) for extracting a moving audio signal of interest from a multi-source and noisy environment in the situation where the signal model is inaccurate. The performances of the beamformers are evaluated in terms of signal fidelity and system robustness. The RCB has been found to provide a better performance if the bearing rate of the moving signal source is smaller than a certain value (determined by the beamformer's integration time). It also has very significant advantage in computational complexity. Its main drawback is its latency. The BCLMSB, on the other hand, is more robust to a moving signal source with high bearing rate. It also does not have the latency problem. Its main weakness is its significantly higher computational complexity.

References

[1] Bao, C. "Adaptive beamforming for sonar audio", *Proceedings of Acoustics 2005*, Busselton, Australia, November 2005, pp. 483-485.

[2] Bao, C. "Performance of time domain and time-frequency domain adaptive beamformers with moving sound sources", *Proceedings of Inter-noise 2014*, Melbourne, Australia, November 2014.

[3] Li, J., Stoica, P., and Wang, Z. "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Processing*, **51**, (7) pp.1702-1715, 2003.

[4] Cox, H., Zeskind, R., and Owen, M. "Robust adaptive beamforming," *IEEE Trans. ASSP*, **35**, (10) pp.1365-1376, 1987.

[5] Frost, O. "An Algorithm for Linearly Constrained Adaptive Array Processing", *Proceedings of the IEEE*, vol. 60, pp. 926-935, 1972

[6] Nielsen, R.O. *Sonar Signal Processing*, Artech House, 1991.