Acoustics 2019
Sound Decisions: Moving forward with Acoustics

Use of robust Capon beamformer for extracting audio signals

Chaoying Bao (1, 2), Ling Jia (1), and Jie Pan (1)

(1) Department of Mechanical Engineering, University of Western Australia, WA, Australia
(2) Defence Science and Technology, Australia

ABSTRACT

The need to extract a single audio signal of interest from a multi-source and noisy environment is common across many disciplines. Use of sensor arrays with adaptive beamforming is a preferred approach for obtaining high quality audio signals in many applications. One of such applications is applying the microphone array technology to extract signals of interest for the purpose of condition monitoring in an industrial site. In that work, we employed a so called Robust Capon Beamformer (RCB) to perform the signal extraction. The adaptive algorithm in RCB was originally developed to deal with the uncertainty existing in the signal model. This uncertainty is also known as signal mismatch. There is an important parameter called the error allowance in the algorithm to regulate the robustness of the beamformer. For an application of signal direction of arrival or localization, which RCB was originally proposed for, the appropriate value of the error allowance is purely defined by anticipated errors in steering vectors. For an application of audio signal extraction, however, the appropriate value of the error allowance has been found to be related also to other factors in the system, such as signal to noise ratio, signal direction of arrival, and the distance between the source and the array. In this paper, we examine the influence of those factors to the appropriate value of the error allowance in the context of machinery condition monitoring through numerical simulation, and provide an empirical formula for choosing the appropriate value of the error allowance in that context.

1 INTRODUCTION

The need to extract a single audio signal of interest from a multi-source and noisy environment is common across many disciplines. Use of sensor arrays with adaptive beamforming is a preferred approach for obtaining high quality audio signals in many applications. One of such applications is applying the microphone array technology to extract signals of interest for the purpose of condition monitoring in an industrial site (Bao et al., 2018). In that work, the beamforming task was divided into two phases. In Phase 1, the whole area of a selected site was scanned to find exact locations of sound sources of interest. In Phase 2, a signal of interest was extracted in time series form by beamforming only to the source location which is obtained in Phase 1.

There are many adaptive beamforming (ABF) algorithms that can be used for audio signal extraction. Depending on what domain where beamforming is carried out, they can be categorized in two classes. One class is time domain beamforming where the whole process including the beamforming part takes place in the time domain. The other is time-frequency domain beamforming (TFDBF) where beamforming is carried out in the frequency domain. In the previous studies (Bao, 2005, Bao, 2014), the performances of several ABF algorithms from both classes were examined and evaluated. According to those studies, TFDBF is a better option for the condition monitoring application where a certain amount of latency can be tolerated.

The beamformer used for TFDBF was a so called Robust Capon Beamformer (RCB) proposed by Li et al. (Li et al., 2003). The adaptive algorithm in RCB was originally developed to deal with the uncertainty existing in the signal model. This uncertainty is also known as signal mismatch. The error allowance, which is a small positive number proportional to signal mismatch, was introduced in the algorithm to regulate the robustness of the beamformer. For an application of signal direction of arrival or localization, which RCB was originally proposed for, the appropriate value for the error allowance is solely determined by anticipated errors in steering vectors. For an application of audio signal extraction, however, apart from dealing with signal mismatch the error allowance also serves an additional purpose of limiting white noise gain (Bao, 2005). Some degree of allowance is always required even in the absence of any signal mismatch. So far we have not found a method to set the appropriate value of the error allowance in the literature for an application of signal extraction. In our earlier study (Bao et al., 2018), the value of the error allowance was chosen by a method of trial-and-error, which is far from ideal. The objective of this study is to identify those factors in the system that affect the value of the error allowance and provide a method to choose the appropriate one. This will be achieved through numerical simulations in the context of condition monitoring in an industrial site.

The rest of the paper is organized as follows. In Section 2, we briefly review some important formulas in the development of the RCB algorithm and show how the error allowance is defined. In the context of audio signal extraction for machinery condition monitoring, we show that the appropriate value of the error allowance is not solely determined by the signal mismatch. In Section 3, we first define the environment and parameters of the numerical simulation. We then present the simulation results and thereby identify those factors that affect the value of the error allowance and obtain an empirical formula for choosing the appropriate value. We conclude the paper in Section 4.

2 RCB FOR AUDIO SIGNAL EXTRACTION

In TFDBF, 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 the beamformer input vector $\mathbf{x}(f)$. The frequency domain beamformer output $y(f)$ is then 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.

In RCB, $\mathbf{w}$ (for simplicity we henceforth drop index *f*) is obtained as follows. Let $\mathbf{x}_n$ denote the beamformer input vector for the n^{th} block, and $\widehat{\mathbf{R}}$ the sample estimate of the cross spectral matrix,

$$\widehat{\boldsymbol{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_{\boldsymbol{n}} \mathbf{x}_{\boldsymbol{n}}^H. \tag{2}$$

Let $\mathbf{v}$ denote the assumed steering vector obtained from the signal model of the system. The RCB algorithm anticipates possible errors in the signal model. The actual steering vector, $\mathbf{v}_a$, is then estimated by solving the following quadratic problem,

$$\min_{\mathbf{v}_a} \mathbf{v}_a^H \widehat{\mathbf{R}}^{-1} \mathbf{v}_a \qquad \text{subject to } \|\mathbf{v}_a - \mathbf{v}\|^2 \le \varepsilon, \tag{3}$$

where ε is a small positive number proportional to the signal mismatch, and is termed the error allowance in this paper. Once $\mathbf{v}_a$ is obtained (see Li et al., 2003 for details), the RCB solution for $\mathbf{w}$ is given by

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

As explained in our previous study (Bao, 2005), for an application of signal direction of arrival or localization, which RCB was originally proposed for, the appropriate value for the error allowance is solely determined by anticipated errors in steering vectors. For an application of audio signal extraction, however, apart from dealing with signal mismatch, the error allowance serves to limit the so-called white noise gain (WNG) (Cox et al, 1987) of a beamformer. The WNG must not be too high because the weights of the beamformer $\mathbf{w}$ are obtained through averaging *N* input blocks whereas a time domain audio signal is calculated using individual input blocks (i.e., no averaging). Within each input block the noise may be stronger in different bearings. With a high WNG, this noise might be amplified and interfere with the signal of interest. Therefore some degree of allowance is always required even in the absence of any signal mismatch. This is illustrated by an example in Figure 1 which shows the performance of the signal extraction by a RCB beamformer against the different values of the error allowance. In the example, which is a numerical simulation of the signal extraction of three sound sources, we assume there is no errors in $\mathbf{v}_a$.

Therefore the value for ε in Equation (3) should be able to be set arbitrarily small. This is true for an application of source localisation. However, for an application of signal extraction, the performance of a RCB beamformer will significantly deteriorate if the value of ε is set too small, as shown in Figure 1.

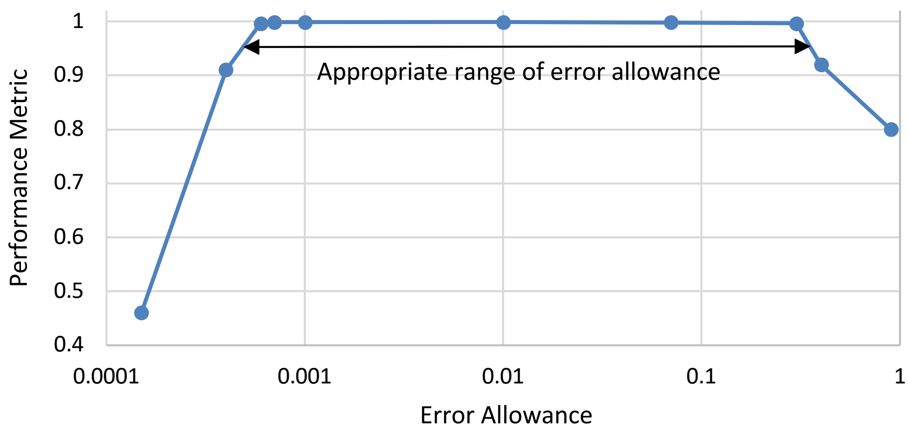


Figure 1: Performance of signal extraction by a RCB beamformer against the error allowance: higher value better performance.

To set an appropriate value of the error allowance for a specified WNG is a complex problem and still an ongoing research. The aim of this study is however to identify those factors that affect the value of the error allowance through numerical simulations, and thereby provide an empirical formula for choosing the value of the error allowance in the context of condition monitoring in an industrial site.

3 NUMERICAL SIMULATION

In order to focus on the important factors that affect the selection of the error allowance, we would like to keep the complexity of the environment and parameters of the simulation minimal.

The array used in the simulation is a uniform linear array consisting of 8 omnidirectional microphones with an inter-element spacing of 0.2 meters. The sound speed is 340 m/s. Thus, the array covers a frequency range from 100 to 850 Hz. The sampling frequency is 8000 Hz.

Three incoherent monopole sound sources are assumed in a free field and located in an area of 60 meters x 8 meters. One of them (normally the middle one) is taken as the signal of interest and the other two as the interferences. Each source emits a different narrow band signal consisting of two frequencies, one around 100 Hz and the other around 800 Hz, covering both ends of the frequency range. Independent and identically distributed (IID) random noise is also added at each sensor to simulate all other noises in the field.

The performance of signal extraction is evaluated in terms of a metric of coherence. 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 signal extraction output *y(t)*,

$$\gamma = \max_{\tau} \frac{|\,\mathrm{COV}\{s(t)y(t-\tau)\}\,|}{\sqrt{\mathrm{COV}\{s^2(t)\}\mathrm{COV}\{y^2(t-\tau)\}}}. \tag{5}$$

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 the output of signal extraction, and value 1 indicating that the two waveforms are proportionally identical.

For a given set of parameter settings in the simulation, we can obtain a coherence number of the signal extraction using Equation (5) for a value of the error allowance. By plotting coherence numbers against values of the error allowance, we get a performance curve such as the one shown in Figure 1. There are some typical features of performance curves that are well illustrated by the plot in Figure 1. Firstly, there is a high coherence plateau for a certain range of the error allowance. The range is termed the appropriate range of the error allowance. Outside the left hand side of that range, the coherence drops rapidly to a very low level as the value of the error allowance decreases. When the error allowance is too low the RCB algorithm designs the beam pattern more aggressively by having a narrower main beam at the expense of increasing the WNG elsewhere. A high WNG makes the performance of the signal extraction deteriorate (Bao, 2005). Beyond the right hand side of the appropriate range, the coherence decreases gradually towards the level of that of the conventional delay-and-sum beamformer (DASB) as the value of the error allowance increases towards its maximum (which is unity). This is due to the fact that a RCB behaves more and more like a DASB when the error allowance is high (Li et al., 2003). Naturally, the value of the error allowance for the signal extraction should be set inside that appropriate range.

In this section, we shall identify those factors that might affect the size and position of the appropriate range. The factors considered are listed as follows:

- Signal distance, the distance from the signal source to the center of the array;
- Signal to Noise Ratio (SNR), the value listed is a mean value over all elements of the array;
- Signal to Interference Ratio (SIR), the value listed is a mean value over all elements of the array;
- Interference angle, the angle between interference and signal sources respect to the center of the array;
- Signal angle, the incident angle from the signal source to the center of the array.

In the simulation, the influences of those factors are investigated by a control variable method. We first define a default setting as listed in Table 1, which represents the scenario most likely encountered in the context considered in this paper. We then examine the effect of a given factor by varying this factor while keeping all other factors constant at a default setting.

Table 1: Default Settings

Factors	Signal Distance	SNR	SIR	Interference Angle	Signal Angle
Default Setting	2m and 8m	3dB	-3dB	±45°	0°

All results presented in this section are obtained from the average of 200 independent runs. The phase relationships among the signals of each source and also among the frequencies within a signal are randomly selected in each independent run. The IID noise added is also randomly generated in each run.

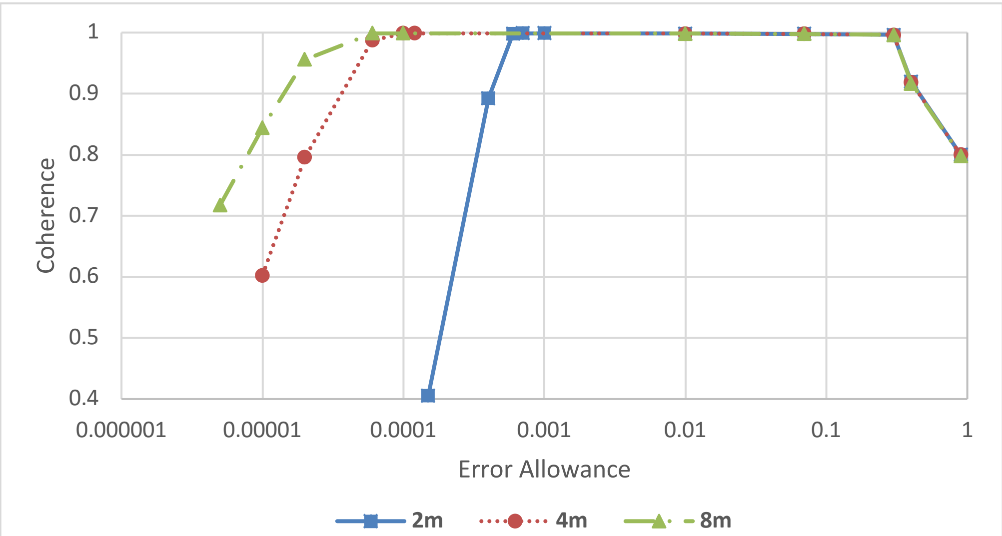


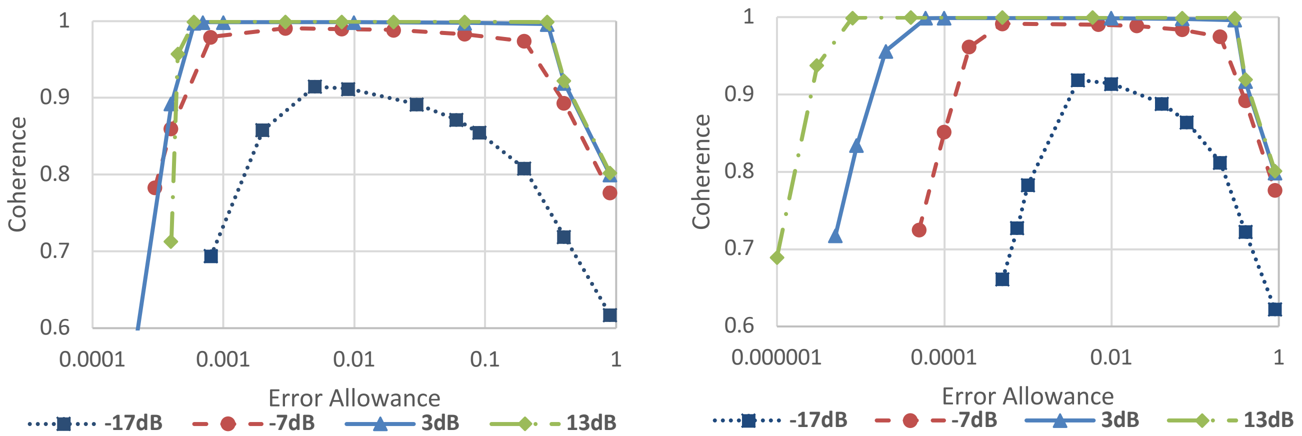
Figure 2: Effect of signal distance

3.1 Effect of signal distance

Figure 2 shows the performance curves with three distance settings, 2m, 4m and 8m, respectively. It can be seen that the appropriate range of the error allowance is affected by the signal distance; the further the distance of the signal source from the array, the wider the appropriate range at the lower end. For the default setting, it appears that the signal distance does not affect the appropriate range at the higher end. In order to capture the effect of the signal distance at the both ends, we will have two default settings for the signal distance: 2m and 8m.

3.2 Effect of SNR

Figure 3 shows the performance curves with four SNR settings, 10 dB apart. It can be seen that the level of SNR affects the appropriate range of the error allowance especially at the lower end for the signal distance at 8m. It also affects the coherence value achieved; the lower the level of SNR, the lower the value of the coherence. In order to achieve satisfactory performance, SNR needs to be above a certain level. In the following simulations, the level of SNR is fixed at 3 dB.

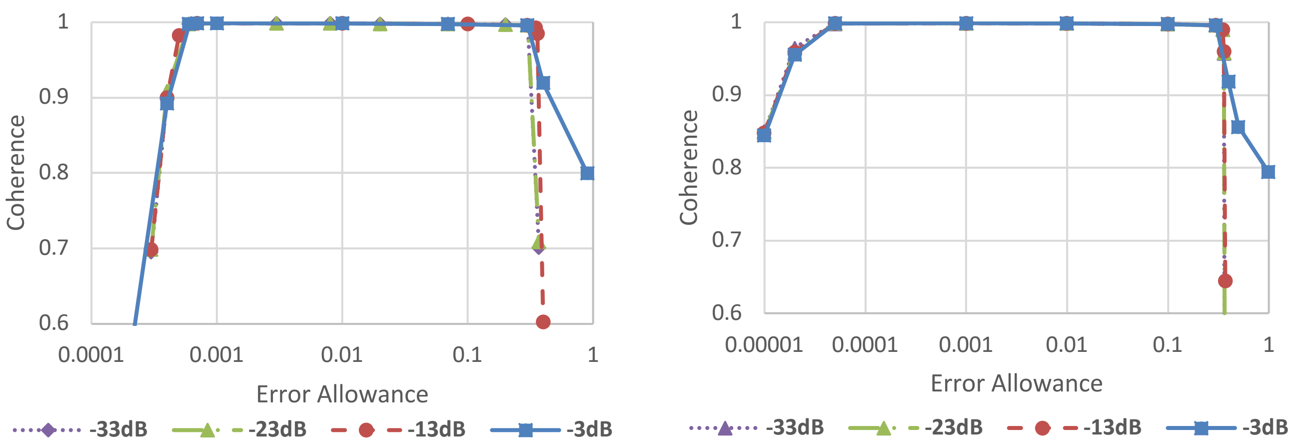


(a) (b)

Figure 3: Effect of SNR with the signal distance at (a) 2m and (b) 8m.

3.3 Effect of SIR

Figure 4 shows the performance curves with four SIR settings, 10 dB apart. It can be seen that the level of SIR does not affect the appropriate range of the error allowance for the large range of SIR covered. Therefore in the following simulations, the level of SIR is fixed at -3 dB.

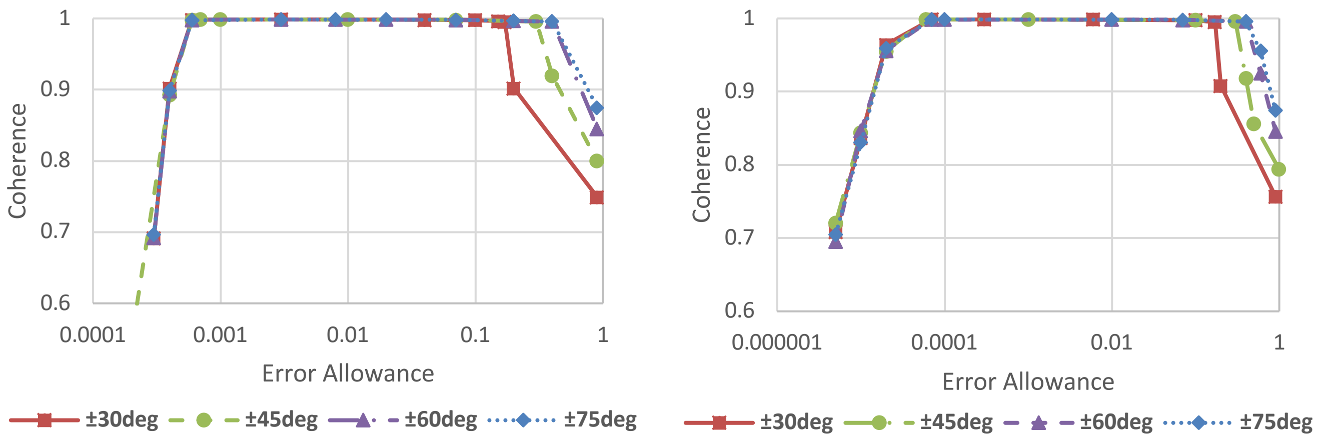


(a) (b)

Figure 4: Effect of SIR with the signal distance at (a) 2m and (b) 8m.

3.4 Effect of interference angle

Figure 5 shows the performance curves with four settings of the interference angle, ±30°, ±45°, ±60° and ±75°, respectively. It can be seen that the appropriate range of the error allowance is slightly affected by the interference angles at the higher end; the closer the interference sources to the signal source, the narrower the appropriate range at the higher end. As such, in the following simulations, the default setting for the interference angle is ±30°.

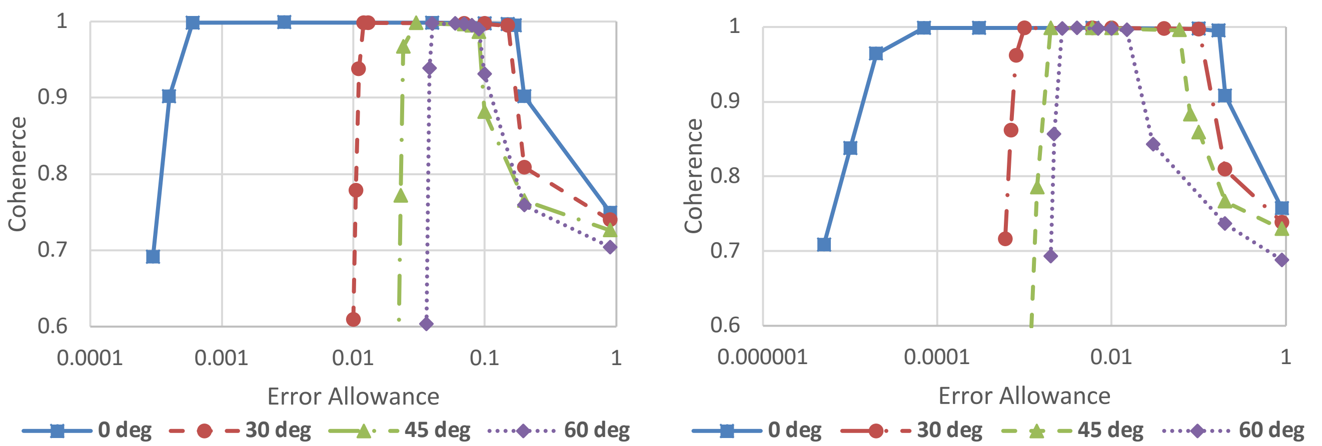


(a) (b)

Figure 5: Effect of interference angle with the signal distance at (a) 2m and (b) 8m.

3.5 Effect of signal angle

Figure 6 shows the performance curves with four settings of the signal angle, 0°, 30°, 45°, and 60°, respectively. It can be seen that the appropriate range of the error allowance is affected by the signal angle. The appropriate range for the signal angle of 0° is much wider than those of all others, especially at the lower end. The appropriate range for the signal angle of 60° is the narrowest among all settings. Therefore, in practice an array should be placed in such a way so that the incident angle of the signal source to the array is as close as possible to 0°, and by all means avoid the signal angle greater than 60°.



(a) (b)

Figure 6: Effect of signal angle with the signal distance at (a) 2m and (b) 8m.

3.6 Empirical formula

From the above investigations, among the five factors considered only two factors have a significant impact on the appropriate range of the error allowance. They are the signal distance d and the signal angle θ. The main effect of d is to shift the range towards the lower end, and the amount of the shift diminishes as d increases. This behaviour can be roughly described by $C_1+C_2e^{-d}$. The main effect of θ is to change the size of the range. The size reduces as

θ increases. This behaviour can be roughly realised by times cosθ for the high bound and 1/cosθ for the low bound. With some curve fitting using the data from Sections 3.2 and 3.5, the empirical formula for the low bound of the appropriate value of the error allowance is (0.0015+0.18e^{-d}) / cosθ and for the high bound is (0.03+0.6e^{-d}) × cosθ, with $\theta \leq 60°$ and SNR ≥ 0 dB. It should be noted that the fitting does not need to be precise, as long as the following three rules are not violated for any simulation settings.

- For the low bound, the values obtained from the formula must not be smaller than those from the simulation.
- For the high bound, the values obtained from the formula must not be greater than those from the simulation.
- The values obtained from the formula for the low bound must not be greater than those for the high bound.

We have verified the formulas against the aforementioned three rules. Table 2 shows some of the results from the verification.

Table 2: Comparison of the values from the empirical formulas to those from the simulation

			Signal angle			
			0°	**30°**	**45°**	**60°**
2m	**Low bound**	**Simulation**	0.006	0.012	0.03	0.04
		Formula	0.0259	0.0299	0.0366	0.0517
	High bound	**Simulation**	0.17	0.15	0.09	0.08
		Formula	0.11	0.097	0.079	0.056
4m	**Low bound**	**Simulation**	0.0001	0.0035	0.0066	0.0095
		Formula	0.0048	0.0055	0.0068	0.0096
	High bound	**Simulation**	0.17	0.14	0.07	0.032
		Formula	0.04	0.0355	0.029	0.0205
8m	**Low bound**	**Simulation**	0.00007	0.001	0.002	0.0027
		Formula	0.0016	0.0018	0.0022	0.0031
	High bound	**Simulation**	0.17	0.1	0.06	0.015
		Formula	0.03	0.027	0.022	0.015

4 CONCLUSIONS

In this paper, we have investigated the problem of choosing an appropriate value of the error allowance in the RCB algorithm for signal extraction in the context of machinery condition monitoring in an industrial site. We have found that the distance and the incident angle of the signal source to the array are the two important factors that have a significant impact on the appropriate range of the error allowance. We have obtained empirical formulas for setting the appropriate value in the aforementioned context through numerical simulations.

In the simulation, we have assumed that there is no errors in the steering vector. This is a useful assumption that allows us to demonstrate there is a low bound for the error allowance even in the absence of any signal mismatch. However, in a real life application some degree of signal mismatch is inevitable. Thus, there is a need for further investigation into whether the error allowance set by the signal mismatch overrides or compounds the appropriate value for the signal extraction.

REFERENCES

Bao, C., L. Jia, B. Coral, D. Mathews, H. Sun and J. Pan. 2018. ‘Sound source localization and signal extraction with multiple microphone arrays’. In *Proceedings of Acoustics 2018*, Adelaide, Australia.

Bao, C. 2005. ‘Adaptive beamforming for sonar audio’. In *Proceedings of Acoustics 2005*, 483-485. Busselton, Australia.

Bao, C. 2014. ‘Performance of time domain and time-frequency domain adaptive beamformers with moving sound sources’. In *Proceedings of Inter-noise 2014*, Melbourne, Australia.

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

Li, J., Stoica, P., and Wang, Z. 2003. ‘On robust capon beamforming and diagonal loading’. *IEEE Trans. Signal Processing*, 51 (7): 1702-1715.