Performance of time domain and time-frequency domain adaptive beamformers with moving sound sources

Chaoying Bao

Defence Science and Technology Organization, Australia

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 a previous study, we have examined the performance of two classes of adaptive beamforming, namely, time domain and time-frequency domain adaptive beamforming, under the conditions that the sound sources were stationary and the signal model was accurate. In this paper, we extend the study to the situations where the sources are moving and certain amount of signal mismatch is allowed. Four different types of adaptive beamformers are considered. The robust Capon beamformer is used for time-frequency domain beamforming. The tapped delay line (TDL) structure is adopted for time domain adaptive beamforming. Three different adaptive algorithms are 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. In the paper, the performances of those four adaptive beamformers are evaluated in terms of fidelity of the beamformer output, robustness of the system, and the computational complexity of the algorithm. It has been found that the robust Capon beamformer provides better performance than the time domain beamformers.

Keywords: Beamforming, Audio I-INCE Classification of Subjects Number: 74.3

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 a previous study [1], we have examined the performance of two classes of adaptive beamforming, namely, time domain and time-frequency domain adaptive beamforming, under the conditions that the sound sources were stationary and the signal model was accurate. In this paper, we extend the study to the situations where the sources are moving and certain amount of signal mismatch is allowed. Four different types of adaptive beamformers are considered. The robust Capon beamformer is used for time-frequency domain beamforming. The tapped delay line (TDL) structure is adopted for time domain adaptive beamforming. Three different adaptive algorithms are 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. In the paper, the performances of those four adaptive beamformers are evaluated in terms of fidelity of the beamformer output, robustness of the system, and the computational complexity of the algorithm.

2. Beamformers

2.1 Time-frequency domain adaptive beamformer (TFDABF)

The beamformer in the time-frequency domain implementation begins with time series array data. The array data are Fourier transformed to the frequency domain and processed by a frequency domain beamformer. The output of that process is then inverse-Fourier transformed back to time domain for an audio output. Figure 1 shows the schematic presentation of the time-frequency domain beamformer.

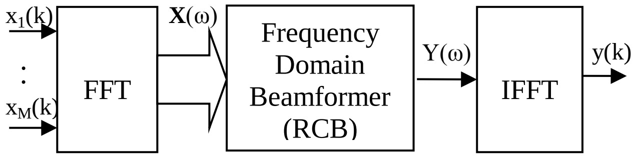

Figure 1 – Schematic presentation of the time-frequency domain beamformer

The beamformer used for the frequency domain processing is the Robust Capon Beamformer (RCB) presented in [2], a type of diagonally loaded Minimum Variance Distortionless Response (MVDR) beamformer. 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 for the loading is determined by the anticipated signal mismatch. For the audio application, however, loading is introduced for another purpose [3]. Some degree of loading is 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. The WNG must not be too high because the weights of a classic frequency domain beamformer are obtained through averaging L blocks of data whereas for audio signals they need to be calculated using individual blocks of data (i.e. no averaging). Within each block of data the noise may be stronger in different bearings than those of an average. With a high WNG, this noise might be amplified and interfere with the signal of interest [3].

2.2 Time domain adaptive beamformers (TDABF)

The beamformers used for the time domain implementation are based on one of the most commonly used time domain adaptive beamforming structure that employs banks of adaptive finite impulse response (FIR) filters, or tapped delay line (TDL) structure, between the delays and summation point of the well-known conventional delay-and-sum beamformer, as illustrated in Figure 2. The coefficients or weights of those FIR filters are adapted accordingly to the characteristics of noise and interference presented such that the noise and interference from non-look-directions are minimized. Three adaptive algorithms are considered for obtaining those optimal FIR 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.

2.2.1 Sample matrix inversion method (SMI)

Following the notation in Frost [5] and referring to Figure 2, we have the input vector,

$$X^T(k) = [x_1(k),\ x_2(k),\ ...,\ x_{MJ}(k)], \tag{1}$$

the weight vector,

$$W^T(k) = [w_1(k),\ w_2(k),\ ...,\ w_{MJ}(k)], \tag{2}$$

and the covariance matrix,

$$R_{XX} = E[X(k)X^T(k)], \tag{3}$$

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

$$y(k) = X^T(k)W. \tag{4}$$

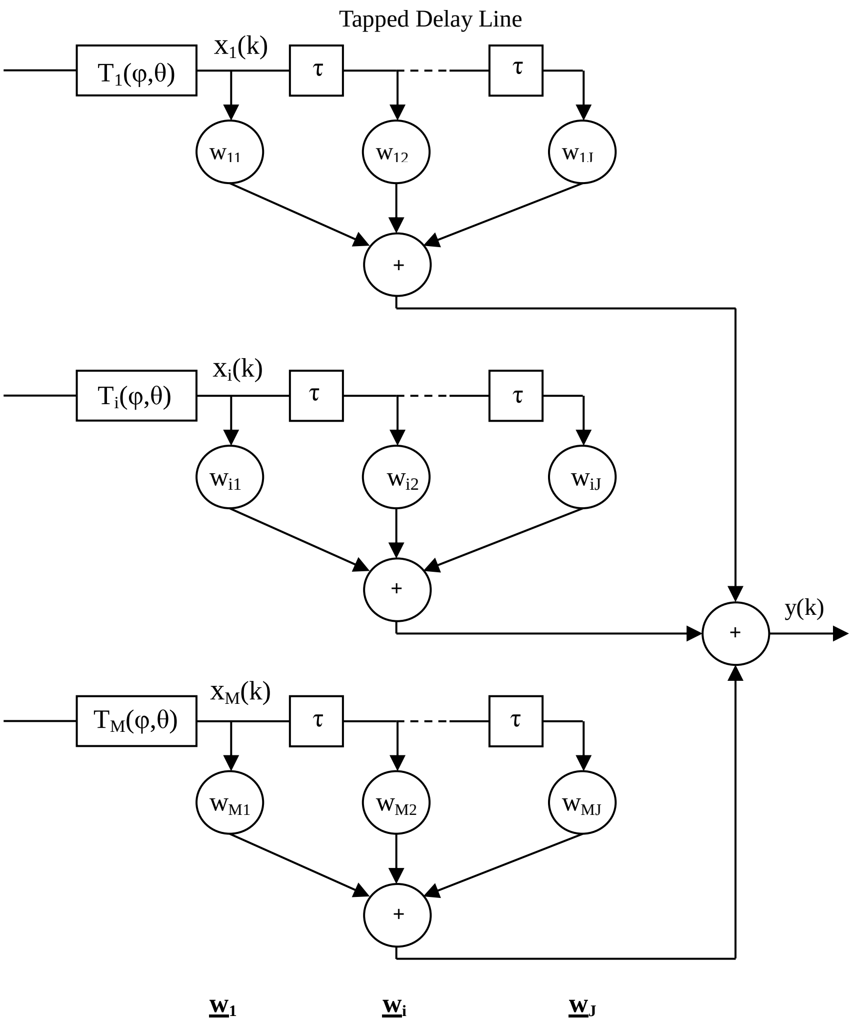

Figure 2 – Adaptive beamformer structure using TDL filter

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

$$\underset{w}{\text{minimise}} \; W^T R_{XX} W \tag{5}$$

$$\text{subject to } \mathrm{C}^{\mathrm{T}} W = F_0, \tag{6}$$

where C is the constraint matrix and F_0 is the vector specifying the frequency response in the look direction. Thus

$$W_{opt} = R_{XX}{}^{-1} C [C^T R_{XX}{}^{-1} C]^{-1} F_0 . \tag{7}$$

If we use the sample covariance matrix

$$\hat{R}_{XX}(k) = \frac{1}{L} \sum_{i=0}^{L-1} X(kL+i) X^T(kL+i) \tag{8}$$

in place of R_{XX} in Eq. (7), we have a SMI beamformer. It is a block data processor with L being the block length, and k in Eq. (8) stands for kth block instead of kth sample.

2.2.2 Recursive least squares method with sliding window (RLSSW)

The essence of RLS algorithm is to estimate the inverse of covariance matrix in Eq. (7) recursively so that the expensive operation of matrix inversion can be avoided [6]. In RLSSW, the statistical expectation in Eq. (3) is replaced by the sample averaging in time with a rectangular sliding window,

$$R_{XX}(k) = \sum_{i=k-L+1}^{k} X(i)X^T(i), \tag{9}$$

where L is the sliding window length. Eq. (9) can also be written as

$$R_{XX}(k) = R_{XX}(k-1) + X(k)X^T(k) - X(k-L)X^T(k-L). \tag{10}$$

Using Matrix Inversion Lemma twice, we obtain the equations for updating the inverse of $R_{XX}(k)$:

$$R_{XXL}^{-1}(k) = R_{XX}^{-1}(k-1) - \frac{R_{XX}^{-1}(k-1)X(k-L)X^T(k-L)R_{XX}^{-1}(k-1)}{-1 + X^T(k-L)R_{XX}^{-1}(k-1)X(k-L)}, \tag{11}$$

$$R_{XX}^{-1}(k) = R_{XXL}^{-1}(k-1) - \frac{R_{XXL}^{-1}(k-1)X(k)X^T(k)R_{XXL}^{-1}(k-1)}{1 + X^T(k)R_{XXL}^{-1}(k-1)X(k)}. \tag{12}$$

The reason that we use a rectangular sliding window instead of a commonly adopted infinite exponential window to formulate the recursive algorithm is that we want to use the diagonal loading technique to improve the robustness of the adaptive algorithm. As discussed in Ma and Wu [7], including diagonal loading in RLS with an infinite exponential window is not trivial. Any amount of loading added in the initial condition will reduce rapidly towards zero after a certain number of iterations. Whereas for RLSSW, as there is no forgetting factor in the updating equations, any initial loading added will remain unchanged. To further reduce the computational complexity of the adaptive algorithm, W_{opt} in Eq. (7) will be updated in a block manner.

2.2.3 Block constrained least mean square method with diagonal loading (BCLMSDL)

In order to further reduce the computational complexity of the RLS algorithm, Frost [5] developed a simple stochastic gradient-descent algorithm called the constrained LMS (CLMS) algorithm where the weights of the adaptive beamformers can be iteratively updated to converge to W_{opt} via the following equations:

$$W(0) = F, \tag{13}$$

$$W(k+1) = P[W(k) - \mu\, y(k)X(k)] + F, \tag{14}$$

$$F = C(C^TC)^{-1}F_0, \tag{15}$$

$$P = I - C(C^TC)^{-1}C^T, \tag{16}$$

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

Including diagonal loading in CLMS is straight forward. Eq. (14) can also be written as

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

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

$$W(k+1) = P[W(k) - \mu(X(k)X^T(k) + \sigma I)W(k)] + F, \tag{18}$$

where σ is the loading coefficient for regulating the mount of loading added. Eq. (18) can further be simplified as

$$W(k+1) = P[(1-\mu\sigma)W(k) - \mu\, y(k)X(k)] + F. \tag{19}$$

Derivation of a block version of Eq. (19) is also straight forward. By replacing $y(k)X(k)$ with its time average, we obtain the block version of equation for updating $W(k)$:

$$W(k+1) = P[(1-\mu\sigma)W(k) - \mu\sum_{i=0}^{L-1} y(kL+i)X(kL+i)] + F, \tag{20}$$

where L is the block length. It should be noted that k in Eq. (20) stands for kth block instead of kth sample.

3. Performance evaluation

In this section, the performances of the four beamformers are evaluated in the presence of signal mismatch and with a moving sound source. The evaluation is based on the coherence between the beamformer output and the signal of interest, as well as the robustness of the beamformers and the computational complexity. 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)*,

$$\gamma = \max_{\tau} \frac{|\,\mathrm{COV}\{s(t)y(t-\tau)\}\,|}{\sqrt{\mathrm{COV}\{s^2(t)\}\mathrm{COV}\{y^2(t-\tau)\}}}, \tag{21}$$

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 consisting of 10 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 TFDABF is 8 Hz. This frequency resolution is experimentally found capable of adequately representing frequency characteristics of the signal, without excessive long integration time. If the frequency resolution is too low, adequate beamforming in 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 bearing resolution of the TFDABF is 1°, which limits the maximum of the look direction mismatch (LDM) of the beamformer to 0.5°.

The FIR filters in the TDABF have 32 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 32 taps achieving high coherence for the given sound sources while maintaining a reasonable computational cost. In order to increase the bearing resolution of the TDABF, the interpolation technique described in [8] is used. The interpolation factor is chosen as 30, which gives the bearing resolutions between about 1° and 10° and the corresponding maximum of LDM between about 0.5° and 5°, dependent of steering directions.

One signal and two interference sound sources are in far field to the array. They all emit sound waves consisting of two tonal components. The tonal frequencies of the signal are chosen as around 700 and 900 Hz. 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 performance difference in TFDABF depending on whether those frequencies of sound sources are on or off the FFT frequencies [1], the tonal frequencies in TFDABF are randomly chosen within one binwidth around 700 and 900 Hz in each simulation run in order to have a fair comparison to the time domain beamformers. 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 two interference sources are stationary and separated from the signal source in bearing of about 30 and -30 degrees apart from the signal source, respectively. The powers of both interferences are 10 dB higher than that of the signal. The error on the signal model or signal mismatch is realised by applying certain percentage of random deviations to the normal values of each sensor output. In this study, 30% of signal mismatch is applied.

The results shown in the following are all obtained by averaging 200 simulation runs. All the adjustable parameters of the four beamformers, such as the integration time and the user parameter of RCB, step size μ, loading coefficient σ, block length L, sliding window length L, etc., are optimally tuned.

Scenario 1: Signal source at broadside with lower bearing rate.

In this scenario, the signal source moves from -1° to 1° with a bearing rate of 0.02°/second. The bearing resolution of the time domain beamformers in this bearing range is about 1°, similar to that of

TFDABF. The main beam width of the beamformers in this direction is also the narrowest. The bearings of the two stationary inference sources are -30° and 30°, respectively.

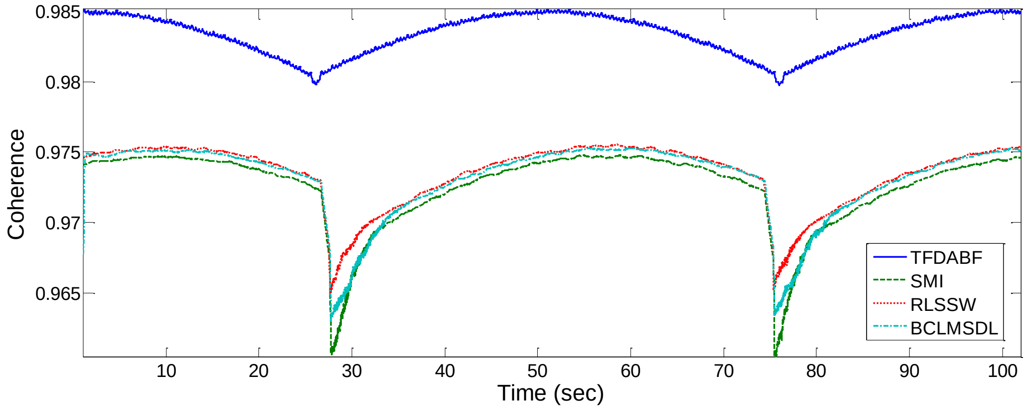

Figure 3 – Time history of the coherences of the four beamformers in Scenario 1

Figure 3 plots the time history of the coherences achieved by the four beamformers. Following observations can be made.

- The coherences of the four beamformers all undergo up and down courses. These ups and downs are coincident with the change of LDM. The coherences peak when there is no LDM and drop to the bottom when LDM is at its maximum.
- The TFDABF achieves higher coherence than those of the TDABFs. The gaps between the highs and lows of the coherence of the former are also smaller than those of the latter (about 0.004 against 0.010). Considering the fact that the coherence achieved by TDABF with stationary sources and without model errors is very similar to that of TFDABF [1], this result indicates that the TFDABF with RCB provides more robust performance.
- As for the performances of the three TDABFs, they are quite similar.

Scenario 2: Signal source at broadside with higher bearing rate

In this scenario, the signal source moves from -5.2° to 5.2° with a bearing rate of 0.2°/second, 10 times higher than that in Scenario 1. The bearing resolution of the time domain beamformers in this bearing range remains about 1°. The main beam width of the beamformers in this direction is also narrow. The bearings of the two stationary inference sources are -30° and 30°, respectively.

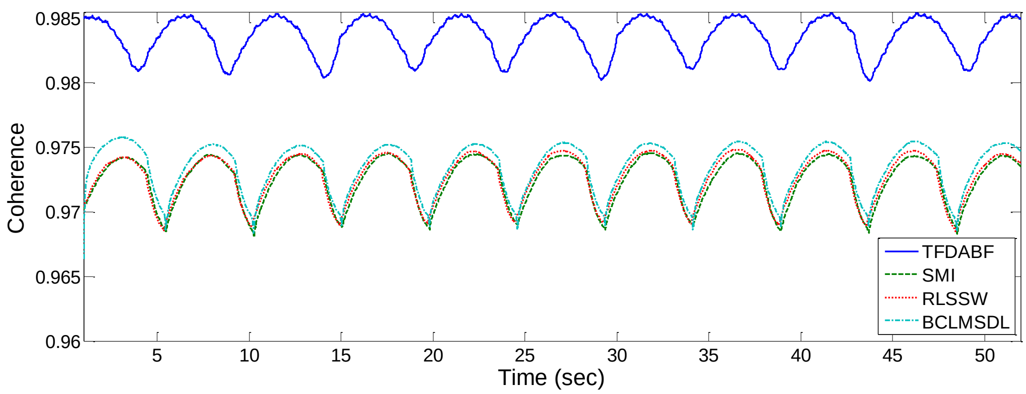

Figure 4 – Time history of the coherences of the four beamformers in Scenario 2

Figure 4 plots the time history of the coherences achieved by the four beamformers. From the figure, the coherences of all four beamformers still go up and down with the change of LDM. The coherence achieved by the TFDABF is similar to that in Scenario 1 and higher than those of the TDABFs, showing the robustness of the TFDABF to fast moving sources. As for the performances of the three TDABFs, they are still similar, with BCLMSDL noticeably better than RLSSW and SMI. This indicates that BCLMSDL is more robust to fast moving sources than the other two TDABFs.

Scenario 3: signal source away from broadside with higher bearing rate

In this scenario, the signal source moves from 50° to 60° with a bearing rate of 0.2°/second. The bearing resolution of the time domain beamformers in this bearing range is about 1.6°, greater than that of TFDABF. The beam widths of the main lobe and side lobes of the beamformers in this direction become broader. The bearings of the two stationary inference sources are 30° and 90°, respectively.

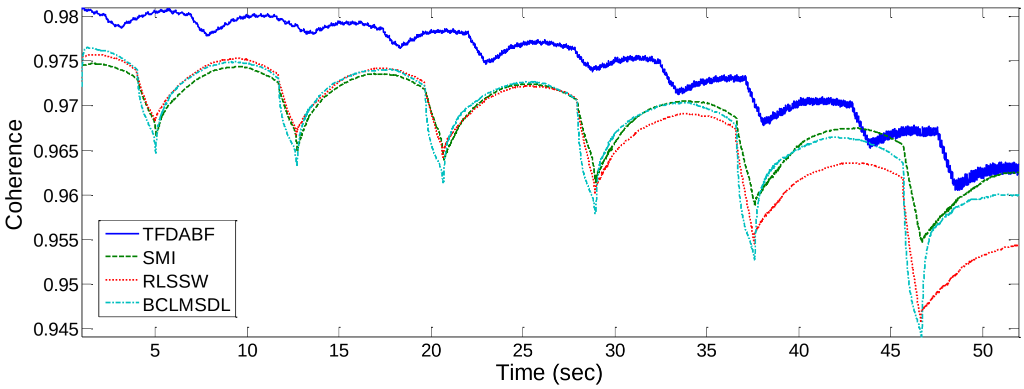

Figure 5 – Time history of the coherences of the four beamformers in Scenario 3

Figure 5 plots the time history of the coherence achieved by the four beamformers.

It can be seen that the coherences of all four beamformers drop as their steering directions move towards endfire direction (ie, from 50° to 60° as time increases). This is because broadening of the beam widths of the beamformers towards endfire allows more noise to get into the beamformer’s output thereby reducing the effectiveness in noise and inference suppression.

Up to 58° (40 seconds in time), the coherence of the TFDABF is still higher than those of the TDABFs. As the bearing resolution of the TDABFs is poorer than that of the TFDABF, the lows of the coherences (where LDM is at its maximum) of the former are much lower than that of the latter. Therefore, overall the TFDABF gives better performance than the TDABFs in terms coherence and robustness in this scenario.

As for the three TDABFs, the coherence of RLSSW drops a lot more than those of the other two as the signal source moves towards endfire direction. This indicates that RLSSW is less robust to a fast moving source towards endfire direction than the other two TDABFs.

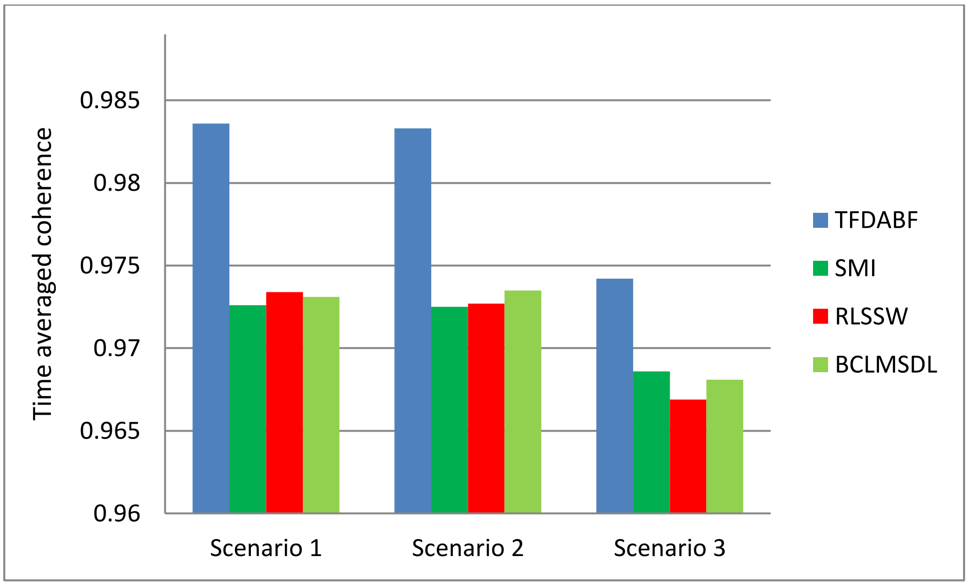

Figure 6 – Time averaged coherences of four beamformers in three scenarios

Table 1 – Computational complexity ratio of the four beamformers

	TFDABF	SMI	RLSSW	BCLMSDL
Complexity ratio	1	27	200	11

Figure 6 summarises the performances of the four beamformers in all three scenarios by comparing their time averaged coherences. Table 1 compares the computational complexity of the four beamformers by the complexity ratio (taking that of TFDABF as 1). It should be noted that the computational complexity of SMI is many times lower than that of RLSSW despite of the matrix inversion in SMI. This is because the matrix inversion in SMI is only carried out once for many hundreds of samples while recursively updating R_{XX}^{-1} in RLSSW has to be performed every sample.

From the figure and the table, the TFDABF scores the best in terms of coherence and the computational complexity. Its coherence is consistently higher than those of the TDABFs in all three scenarios, and its computational complexity is 11 times lower than that of BCLMSDL (the best of TDABFs) and 200 times lower than that of RLSSW. It is also very robust. One of the features of RCB in TFDABF is that the equivalent diagonal loading added in the system is not fixed and uniform as those of the three TDABFs but adapted according to the beamformer’s input [2]. This could be the reason that the TFDABF provides a better performance in terms of coherence and robustness.

One drawback of the TFDABF is its latency. The delay time from input to output of the TFDABF in the system considered in this study is about 1.2 seconds. For those applications where this latency cannot be tolerated, TDABF has to be used. Among the three TDABFs, BCLMSDL appears to be a good choice. It provides a good performance in coherence with the least computational complexity.

4. CONCLUSIONS

We have examined the performance of four adaptive beamformers, one TFDABF and three TDABFs, 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 performance of the beamformers is evaluated in terms of signal fidelity, system robustness, and the computational complexity. The TFDABF has been found to provide a better performance in all three criterions, and therefore is recommended for those applications where a certain amount of latency can be tolerated. Among the three TDABFs, BCLMSDL has been found to give a good performance in signal fidelity with the least computational complexity. It is also more robust to a fast moving source.

REFERENCES

1. Bao C, Farag G, Pan J. Comparison of the performance of time domain and time-frequency domain adaptive beamforming. Proc Oceans’10 IEEE Conference, May 2010; Sydney, Australia 2010.
2. Li J, Stoica P, Wang Z. On robust Capon beamforming and diagonal loading. IEEE Transactions on Signal Processing, vol. 51 (7), July 2003.
3. Bao C. Adaptive beamforming for sonar audio. Proc Acoustics 2005, November 2005; Busselton, Australia 2005. p. 483-5.
4. Cox H, Zeskind R, Owen M. Robust adaptive beamforming. IEEE Trans. ASSP. 1987; 35(10):1365-76.
5. Frost O. An Algorithm for Linearly Constrained Adaptive Array Processing. Proceedings of the IEEE. 1972; vol. 60:926-35.
6. Haykin SS. Adaptive Filter Theory, 4th ed. Upper Saddle River: Prentice Hall; 2002.
7. Ma L, Wu X. Robust Adaptive Beamforming Based on Sliding-window RLS Algorithm. 2nd International Conference on Signal Processing Systems; 2010. V3:793-7.
8. Nielsen RO. Sonar Signal Processing. Artech House; 1991.