ROBUST DOA ESTIMATORS FOR USE IN COLOURED NOISE ENVIRONMENTS

P.R. White, R.G. Porges

University of Southampton, Highfield, Southampton, Hants., SO17 1BJ

1. INTRODUCTION

There are a range of optimal, maximum likelihood, Direction Of Arrival (DOA) estimation algorithms [*e.g.* 1,2,3]. The objective of each algorithm is the same, namely to find the signal parameters which maximise the maximum likelihood function. The methods vary in their approach to locating this optimal point. Normally the likelihood function chosen requires the assumption that the background noise is spatial homogeneous (white) and Gaussian. In most oceanic applications the assumption of a Gaussian background noise is not overly restrictive; however the assumption of whiteness is generally not applicable.

In cases where the background noise is Gaussian, but coloured, then the conventional approach to restoring optimality is to estimate the background noise statistics and use these to pre-whiten the data prior to applying the conventional algorithms. This approach requires the opportunity to estimate the background noise statistics in the absence of the signal. In many applications such opportunities are available, whilst in others one cannot guarantee that there will be opportunities to observe "signal free" data. It is to these latter scenarios that we address ourselves. Specifically we assume that the array receives data in an (unknown) coloured Gaussian noise environment and we seek the optimal, maximum likelihood, DOA estimators.

2. HIGH RESOLUTION DOA ESTIMATION

Over many years there has been an abundance of algorithms for estimating DOAs. In this section we review signal models which underlie the vast majority of these algorithms. The models we adopt are as follows: each sensor receives a signal $x_k(t)$, consisting of a signal component $s_k(t)$ and an additive, zero mean, Gaussian noise component $w_k(t)$. Hence for a single sensor we write

$$x_k(t) = s_k(t) + w_k(t)$$

The assumption that the noises are correlated implies that $E[\, w_k(t)\, w_j(t)\,] \neq 0$ for some values of $k \neq j$. Specifically if we arrange these data into column vectors such that

$$\underline{x}(t) = [x_1(t)\;\; x_2(t)\;\; ....\;\; x_L(t)]^\dagger,\;\; \underline{s}(t) = [s_1(t)\;\; s_2(t)\;\; ....\;\; s_L(t)]^\dagger \text{ and } \underline{w}(t) = [w_1(t)\;\; w_2(t)\;\; ....\;\; w_L(t)]^\dagger$$

where L is the total number of sensors and † denotes the Hermitian (conjugate) transpose, the noise correlation matrix R_w is defined by

$$R_w = E[\,\underline{w}(t)\,\underline{w}(t)^\dagger\,]$$

In the case of white noise this matrix is a scaled version of the identity matrix, *i.e.* $R_w = \sigma^2 I$. For a coloured noise scenario no such assumption can be made.

It is usually assumed that the signals are narrow band and in general complex. This assumption, not only covers the case where the signal of interest is close to tonal, but also the more common situation in SONAR systems, where the data is Fourier transformed and each bin of the transform is processed individually. In such cases, assuming only one signal (with centre frequency f_o) is impinging on the array, we can write

$$s_k(t) = A(t-\tau_k)\, e^{2\pi i f_o (t-\tau_k)}$$

so that the signal on the k[th] sensor is a delayed version of $A(t)\, e^{2\pi i f_o t}$, which implies "perfect" propagation and a source in the far field. The delay τ_k is a function of the source and receiver locations. Assuming that the amplitude, $A(t)$, is sufficiently slowly varying as to be constant across the aperture of the array then the signal vector is given by

$$\underline{s}(t) \approx A(t)\, e^{2\pi i f_o t}\, [\, e^{2\pi i f_o \tau_1}\ \ e^{2\pi i f_o \tau_2} \ldots\ e^{2\pi i f_o \tau_L}\,]' = A(t)\, e^{2\pi i f_o t}\, \underline{h}(\theta) \qquad (1)$$

where $'$ denotes straightforward transposition. If the sensor positions are known and the source location only depends upon the DOA (θ) then the delays are a function of θ only, *i.e.* $\tau_k = \tau_k(\theta)$. It is this that motivates our choice of notation in (1) for the vector $\underline{h}(\theta)$.

If we assume that p signals arrive at the array simultaneously, from unique angles $\theta_1 \ldots \theta_p$ then, using superposition, the signal vector becomes

$$\underline{s}(t) = e^{2\pi i f_o t} \sum_{k=1}^{p} A_k(t)\, \underline{h}(\theta_k)$$

The signal correlation matrix, R_s, is given by

$$R_s = E[\,\underline{s}(t)\underline{s}(t)^\dagger\,] = H\, R_A\, H^\dagger$$

where R_A is the amplitude correlation matrix, and H is the $(L\times p)$ matrix formed by concatenating the vectors $\underline{h}(\theta_k)$ so that $H = [\,\underline{h}(\theta_1)\ \underline{h}(\theta_2)\ \ldots\ \underline{h}(\theta_p)\,]$. If the p signals are all uncorrelated then the amplitude correlation matrix R_A is diagonal, with the k[th] term being $E[\,|A_k(t)|^2\,]$.

Further assuming that the signal and noise are uncorrelated allows one to write

$$R_x = E[\ \underline{x}(t)\, \underline{x}(t)^{\dagger}\ ] = H\, R_A\, H^{\dagger} + R_w \tag{2}$$

In this model we have assumed that the amplitudes $A(t)$ are stochastic; hence the appearance of the amplitude correlation matrix R_A. An alternative approach is to assume that the amplitudes are unknown deterministic parameters. In this case it is more conventional to consider the problem in a slightly different form, namely we consider the data matrix X which is formed as

$$X = [\ \underline{x}(t_1)\ \ \underline{x}(t_2)\ \ \underline{x}(t_3)\ \ \ldots.\ \underline{x}(t_N)\ ]^{t}$$

in which $\underline{x}(t_k)$ represents the vector of measured sensor outputs at time instant k and N is the total number of available snapshots (time instants). Further we write

$$X = A\, H^{\dagger} + W \tag{3}$$

where W is the $(N{\times}L)$ matrix of noise terms, similar in form to X, and A is the $(N{\times}p)$ matrix of amplitudes such that $A = \{A_{n,k}\}$ and $A_{n,k}$ is the amplitude of the k^{th} signal at the n^{th} time instant.

In which case we write the signal correlation matrix as follows

$$R_x = E[\ \underline{x}(t)\, \underline{x}(t)^{\dagger}\ ] = H\, A^{\dagger}\, A\, H^{\dagger} + R_w \tag{4}$$

At this stage the difference between (2) and (3) may appear academic, since in practice the matrix R_A would have to be estimated and the "natural" estimator for it would be $A^{\dagger} A / N$ assuming the amplitudes were known.

The two models described by (2) and (4) are dubbed the stochastic and deterministic signal models respectively.

3. HIGH RESOLUTION DOA ESTIMATION IN WHITE NOISE

In practice most high resolution algorithms estimate the array correlation matrix R_x using

$$\hat{R}_x = \frac{1}{N} X^{\dagger} X = \frac{1}{N} \sum_{k=1}^{N} \underline{x}(t_k)\, \underline{x}(t_k)^{\dagger}$$

then by using a signal model attempt to match the measured correlation matrix with that modelled using either (2) or (4). The variety of possibilities of forming this comparison of the modelled and measured matrices gives rise to the wide selection of DOA algorithms.

For example MUSIC [4], on which many of the eigenvector methods are based, searches for values of θ which render $\underline{h}(\theta)$ most close to the subspace spanned by the p largest eigenvectors of $\hat{R}_x$. The subspace spanned by the p largest eigenvectors is termed the signal subspace. In the case of MUSIC this is equivalent to finding the value of θ which ensures that $\underline{h}(\theta)$ is furthermost from the vector space spanned by the L-p smallest eigenvectors of $\hat{R}_x$. This latter subspace is often called the noise subspace (we continue with this terminology but note that it is misleading since the noise vectors generally span the full L dimensional vector space).

MUSIC and the general family of eigenvector methods are suboptimal. They do generally have the advantage of better performance than Fourier based methods and produce results via a one-dimensional scan, so the output is in a similar format to that of a Fourier based method and makes searching for peaks a simple task. However these method's performance degrades in multi-path environments, when signals have correlated amplitude terms.

Another class of algorithms are those based on maximum likelihood (not to be confused with the algorithm proposed by Capon [5], which is also sometimes called maximum likelihood). We develop these algorithms for the deterministic signal model (3) and (4). The basis of these algorithms is to solve (3) to obtain an expression for the amplitudes A, namely

$$A = XH(H^{\dagger}H)^{-1} \tag{5}$$

Then the error between the measured data X and the model estimate $AH^{\dagger}$ is formed as

$$E = X - AH^{\dagger} = X(I - H(H^{\dagger}H)^{-1}H^{\dagger}) = XP^{\perp}$$

where $P^{\perp}$ is the orthogonal projection matrix. The norm of this matrix is then minimised with respect to the parameters in H, *i.e.* the angles of arrival θ. The function to be minimised is

$$P^{\perp}X^{\dagger}XP^{\perp} = tr\{P^{\perp}\hat{R}_x\} = tr\{(X - AH^{\dagger})^{\dagger}(X - AH^{\dagger})\} \tag{6}$$

where $tr\{\ \}$ is the trace operator. Remember that the trace of a matrix is equal to the sum of its eigenvalues. Thus these maximum likelihood methods require one to project the estimated data correlation matrix onto the space spanned by the vectors $\underline{h}(\theta_k)$ $k=1,2, \ldots p$ and measure the size of the resulting error. The p directions θ_k are selected so as to minimise this error, requiring a p dimensional optimisation. This increase in dimensionality enables the algorithms to cope with multi-path problems but does introduce a large computational loading. This loading can be mitigated but using an iterative scheme. Such algorithms perform the multi-dimensional search via a series of one-dimensional scans.

The equation (5) represents the least squares solution of (3). This least squares solution only coincides with the maximum likelihood solution if the background noise is white. If the background noise is coloured and the noise correlation matrix R_w is known, then the maximum likelihood solution is given by the weighted least squares solution of

$$\tilde{X} = R_w^{-1/2}\, X = R_w^{-1/2}\, A\, H + R_w^{-1/2}\, W$$

where $R_w^{-1/2} R_w^{-1/2} = R_w^{-1}$.

4. HIGH RESOLUTION DOA ESTIMATION IN COLOURED NOISE

In this section we discuss algorithms for estimating DOAs which are designed to function in non-white noise scenarios, specifically in cases where the noise correlation matrix is unknown. There are far fewer algorithms designed to function under these conditions. The maximum likelihood formulation requires one to maximise the function

$$p(X) = \frac{1}{\pi^{Np}\left|R_w\right|^N} e^{-N\, tr\{R_w^{-1}\hat{R}_w\}} = \frac{1}{\pi^{Np}\left|R_w\right|^N} e^{-tr\{R_w^{-1} W^{\dagger} W\}}$$

where we now assume that not only are the DOAs and amplitudes of signals unknown parameters but also the noise correlation matrix R_w is unknown. Taking the logarithm of $p(X)$ and omitting constant terms gives [6]

$$L(X) = log(p(X)) = -N\, log(|R_w|) - tr\{R_w^{-1}\, W^{\dagger} W\} \tag{7}$$

Differentiating with respect to R_w, and setting to zeros gives that the maximum likelihood estimator of R_w is

$$\hat{R}_w = \frac{1}{N} W^{\dagger} W$$

Substituting this into (7) and again omitting constant terms, yields

$$L(X) = log(\ |W^{\dagger} W|\ )$$

Thus this maximum likelihood estimator is obtained by minimising the function

$$log(\ |W^{\dagger} W|\ ) = log(\ |(X\text{-}AH^{\dagger})^{\dagger}\,(X\text{-}AH^{\dagger})|\ ) \tag{8}$$

Care should be taken when interpreting this operation, since the determinant of a matrix is the product of its eigenvalues, and if only one of the eigenvalues is zero then the determinant is zero. However here the correct interpretation of the determinant is the product of only the non-zero eigenvalues.

This result is not in itself novel. Reilly, Wong and Reilly [7] developed the same cost function from a Bayesian stand point. They treated the noise correlation matrix as a nuisance parameter in the estimation scheme and by employing a suitable prior distribution integrated it out. Here we present a more direct derivation which we believe is more amenable.

It is interesting to contrast the results in (6) and (8). We see that in the case of white noise the optimal estimates are those which minimise the sum of the projected eigenvalues, whilst in the case of unknown coloured noise the optimal parameters are those obtained by minimising the product of the projected eigenvalues.

Unfortunately the result (8) is far more difficult to compute than its counterpart (6). This arises because the estimation of the amplitudes A and angles of arrival θ no longer decouple, so that whereas in the white noise case we are able to write the A in terms of H alone, in the coloured noise case this is not possible. Further whilst H is described by p parameters, the amplitude matrix A contains a total of pN terms. So the minimisation has a total of $p(N+1)$ parameters to work over. Such a minimisation, whilst being in theory feasible, is in practice beyond the scope of existing algorithms and/or computer power for reasonable values of N and p.

To circumvent this problem a suboptimal approach can be adopted. Specifically one can assume that the amplitude terms take the white noise form given by (5). This yields the problem of minimising the cost function

$$|P^{\perp}\hat{R}_x P^{\perp}| \tag{9}$$

which is only a function of the angles of arrival.

A second method for estimating angles of arrival in coloured noise exists [8]. This method exploits the concept of Minimum Description Length (MDL) which has been successful exploited in other array processing problems [9]. MDL is an information theoretic criterion, where the objective is to express data in as smaller fashion as possible. Applying this criterion to the problem in hand Wax [8] developed a second cost function based on the MDL principle. This function requires the minimisation of the function

$$|P^{\perp}\hat{R}_x P^{\perp}||P\hat{R}_x P| = \left|\hat{R}_N\right|\left|\hat{R}_S\right| \tag{10}$$

where P is the projection matrix $H(H^{\dagger}H)^{-1}H^{\dagger}$. The matrices $\hat{R}_N$ and $\hat{R}_S$ are the projections of the correlation matrices into the noise and signal subspaces respectively. The determinant operator can be geometrically thought of as a measure of volume. Thus the cost function (10) can be regarded as minimising the product of the volumes of the noise and signal subspaces. In practice the volume of the signal subspace peaks at correct angles whilst the volume of the noise subspace drops. Intuitively the minimisation of (10) is unpleasing since we are trying to minimise the effect of the signal. From these observations we propose an alternative cost function

$$\frac{\left|\hat{R}_N\right|}{\left|\hat{R}_S\right|} \tag{11}$$

In this case the two components work with each other rather than against each other as in (10). Note that in these formulations the noise and signal subspaces provide different information. In the white noise case minimising in the noise subspace was equivalent to maximisation in signal subspace and consequently the two spaces yielded the same information, this is not the case here.

We note that both (7) and (9) can be written in the notation of (10) and (11) as

$$tr\left\{\hat{R}_N\right\} \tag{12}$$

$$\left|\hat{R}_N\right| \tag{13}$$

respectively. This leaves us with 4 cost functions which we seek to minimise, namely equations (10)-(13). Each method requires a minimisation in p dimensional space (although we bear in mind that (12) is amenable to searching via a series of one dimensional scans).

5. RESULTS

To help ensure parity when comparing algorithms we choose to use the same optimisation routine on all the algorithms merely changing the cost function. The results we present are compared with the Cramer-Rao lower bound (CRLB) for the deterministic model, as presented in [10]. That result was derived for the case of white noise but can be relatively simply modified by appropriate whitening of the parameters for a coloured noise environment [6].

For the sake of these simulations we refer to the algorithms as follows : (7) is called the trace algorithm, (9) RWR (Reilly Wong and Reilly) algorithm, (10) Wax algorithm and (11) is the Quotient method.

The simulations consisted of 100 Monte-Carlo simulations using different realisations of the noise signal. For each realisation all 4 algorithms were implemented and the results stored. These results were then performed over a range of Signal to Noise Ratios (SNR) and the mean squared error calculated.

The test data modelled two signals impinging on a 10 element uniform line array, from angles of 0° (Broadside) and 6°. These signals were modelled as perfectly coherent. They were immersed in coloured background noise, whose correlation matrix consisted of the elements :

$$r_{j,k} = 0.9|j-k| \, e^{-i(j-k)\pi/2}$$

which describes a background noise which has it last contribution at an angle of -30°.

a) b)

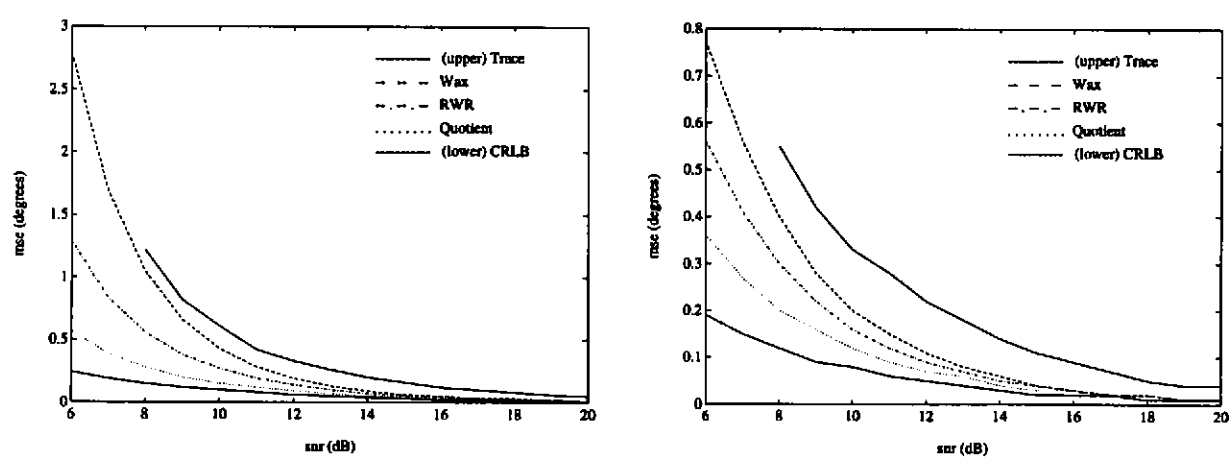

Figure 1 : *Mean Squared Error as a Function of SNR*

Figure 1 plots the means squared error for a range of SNRs, with 100 snapshots of data (*i.e. N=100*). Figure 1a) shows the results for the broadside signal, whilst Figure 1b) shows those for the signal from 6° (this convention is continued throughout this section). From these Figures we can see that the Quotient method performs better than either of the of other three methods. The trace algorithm fairs the worst, which is to be expected since the algorithm is optimised for the white noise case. Initially attempts were made to include MUSIC in these simulations, but the joint problems of working in a coloured noise environment with coherent signals meant that the algorithm failed.

Figure 2 shows the variation of the mean squared error as the number of snapshots is increased, whilst the SNR is fixed at 10dB. These results initially show similar trends with the ranking of the algorithms being similar to that in Figure 1. However as the number of snapshots gets large the performance of both the Wax and RWR algorithms plateaus out, this maybe indicative of inconsistency in the algorithms. In this region the performance of the trace algorithm continues to improve and surpasses the other two algorithms.

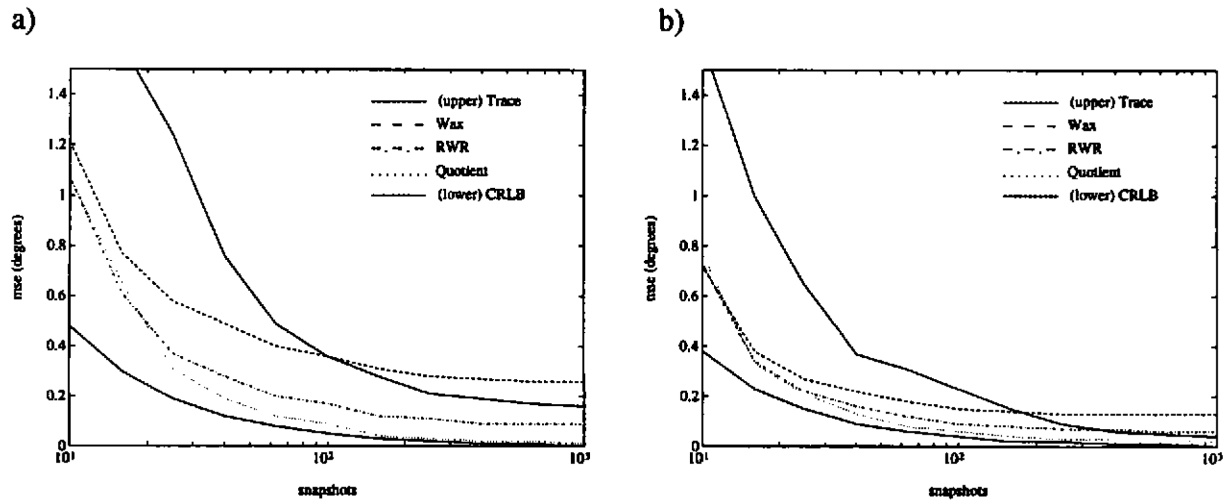

Figure 2 : *Mean Squared Error as a Function of Number of Snapshots*

Finally we plot the bias of the algorithms as a function of the SNR, using 100 snapshots. The results are shown in Figure 3. We again observe that the ranking of the algorithms is as before with the trace method generating generally less biased estimates than either the Wax or RWR algorithms.

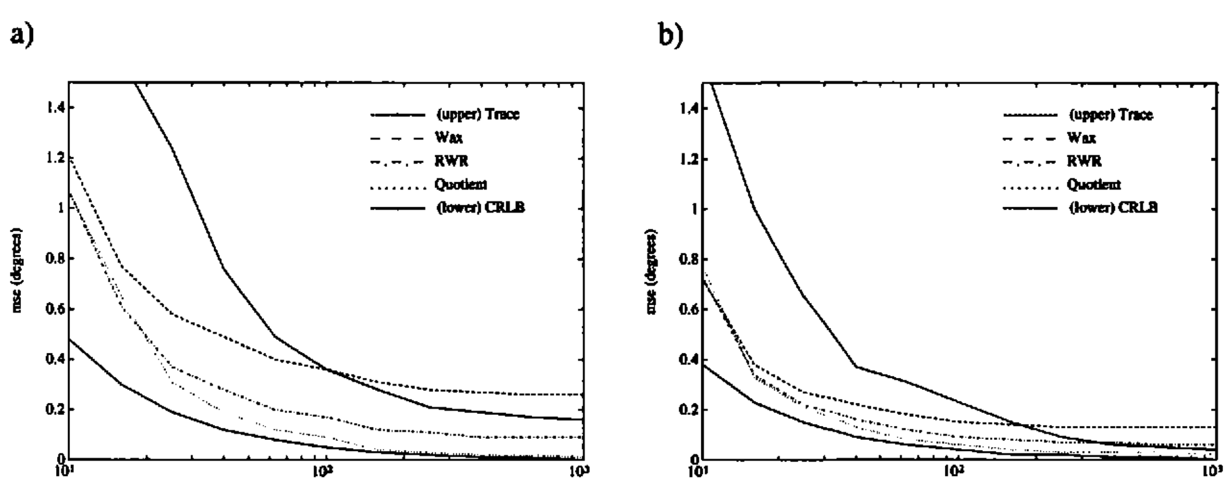

Figure 3 : *Bias as a Function of SNR*

6. CONCLUSIONS

In this paper we have discussed algorithms for estimating DOA in environments where the background noise is coloured with unknown statistics. The group of algorithms arising from this formulation are based on minimisations of determinants of projected correlation matrices. These algorithms, like the maximum likelihood algorithm for white noise, require multidimensional searches, resulting in relatively large computational loads. However the resulting algorithms are relatively robust to the presence of coloured noise and coherence in the signals. Of the algorithms discussed the heuristic Quotient algorithm show most promise.

7. REFERENCES

[1] Y. Bresler, A. Maxovski "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise",IEEE Trans. on Acoustics Speech and Signal Processing, Vol. 34, 1081-1089, 1986.
[2] I. Ziskin, M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection", IEEE Trans. on Acoustics Speech and Signal Processing, Vol. 36, 1553-60, 1988.
[3] I.J. Clarke, "Robustness of Eigen Based Analysis Techniques versus Iterative Adaption" RADAR 87, IEE Conference Publication No. 281, 84-88, 1987.
[4] R.O. Schmidt, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation", Ph.D. Thesis, Stanford University, 1982.
[5] J. Capon, "High Resolution Frequency-Wavenumber Spectrum Analysis", Proc. IEEE, Vol. 57, 1408-1418, 1969.
[6] R.G. Porges, "Robust Eigenvalue Decomposition Techniques for High Resolution Signal Estimation and Detection", Ph.D. Thesis, University of Southampton, 1993.
[7] J.P. Reilly, K.M. Wong, P.M. Reilly, "Direction of Arrival Estimation in the Presence of Noise with Unknown, Arbitrary Covariance Matrices", Proc. ICASSP '89, pp 2609-2612, 1989.
[8] M. Wax, T. Kailath, "Detection of Signals by Information Theoretic Criteria", IEEE Trans. on Acoustics Speech and Signal Processing, Vol. 33, 387-392, 1985.
[9] M. Wax, "Detection and Localisation of Multiple Sources in Noise with Unknown Covariance", Proc. ICASSP '95, pp 1345-8, 1995.
[10] P. Stoica, A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Roa Bound", IEEE Trans. on Acoustics Speech and Signal Processing, Vol. 37, 720-741, 1989.

ACKNOWLEDGEMENTS

The authors would like to that Prof. A. Parsons and Dr. A. Mc.Lean (DRA Winfrith) and P. Martinson (DRA Farnborough) for their technical assistance and DRA for funding this research.