INSPECTION COPY

INSTITUTE OF ACOUSTICS

UNDERWATER ACOUSTICS GROUP

PROCEEDINGS OF THE CONFERENCE

ADAPTIVE PROCESSING

HELD AT:

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING,
LOUGHBOROUGH UNIVERSITY OF TECHNOLOGY,
ENGLAND.

14th JUNE 1977

INSTITUTE OF ACOUSTICS

UNDERWATER ACOUSTICS GROUP

PROCEEDINGS OF THE CONFERENCE

ADAPTIVE PROCESSING

HELD AT:

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING,
LOUGHBOROUGH UNIVERSITY OF TECHNOLOGY,
ENGLAND.

14th JUNE 1977

Institute of Acoustics

25 Chambers Street
Edinburgh
EH1 1HU

Telephone 031 - 225 2143

This Conference was organised by the Underwater Acoustics Group of the Institute of Acoustics.

The Committee was as follows:

Chairman:	Dr. D.E. Weston
Treasurer:	Dr. A.R. Pratt
Secretary:	Dr. B.V. Smith
Members:	Dr. D.J. Dunn
	Mr. R.B. Mitson
	Dr. J.S.M. Rusby

Printed at the University of Bath

CONTENTS

Opportunities and Limitations of Adaptive Array Processing

BY

Peter M. Schultheiss

Yale University

1. Introduction

This paper is concerned with the detection of (and extraction of information from) Gaussian signals observed in a background of Gaussian noise. If detection is to be possible at all, there must be significant differences between the statistical properties of signal and noise about which the designer is knowledgeable. In the typical passive sonar problem--our primary interest here--the signal originates from a spatially concentrated source and therefore generates a more or less coherent wavefront. The noise often originates from many small, spatially scattered sources which do not generate one coherent wavefront. The existence of a coherent wavefront, which can be identified through the use of an array of sensors, therefore establishes the presence of a signal and provides immediate information concerning certain of its features (e.g. the direction from which it comes). These elementary notions lead to the detector shown in Fig. 1a.

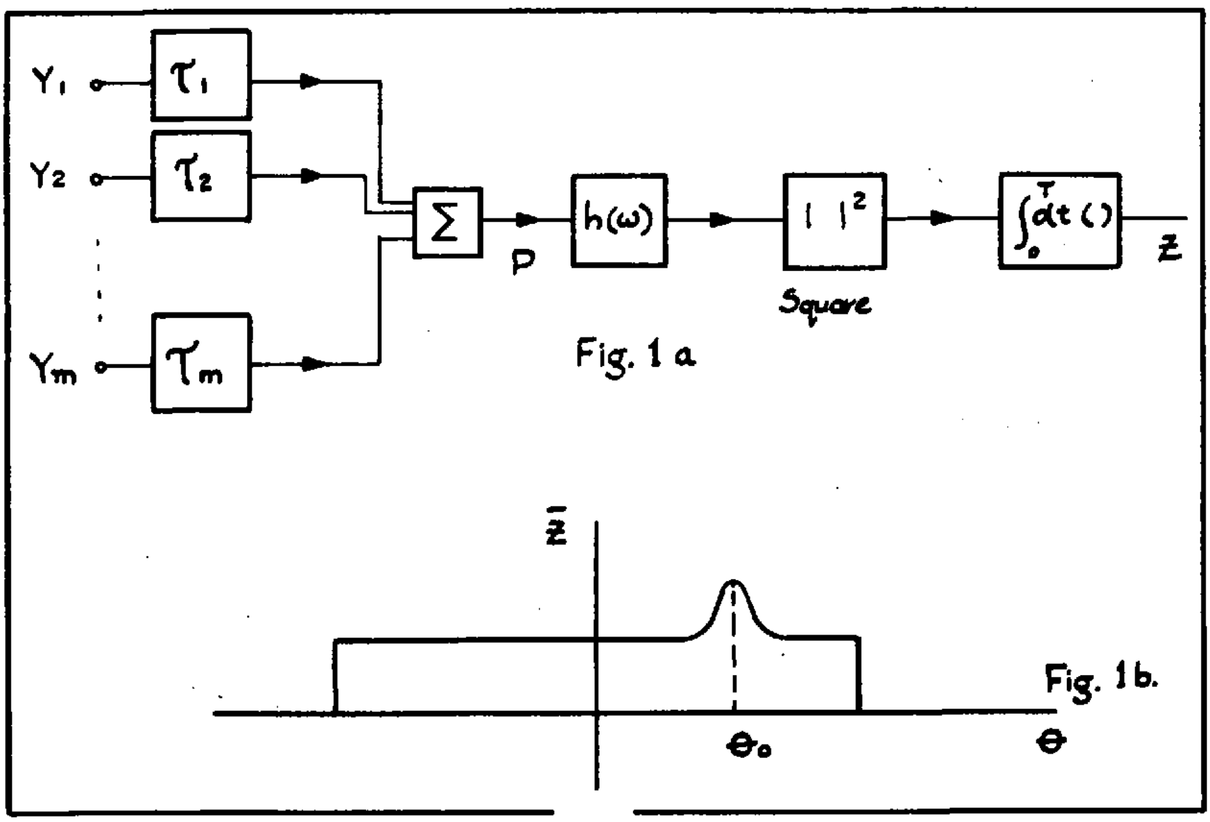

Fig. 1 a

Fig. 1 b.

If the set of delays is adjusted properly (to match the incoming signal wavefront), the signal components add coherently whereas the noise components do not. As one adjusts the relative delays one obtains a pattern such as that shown in Fig. 1b. The location of the peak is readily calibrated in terms of signal bearing.

The instrumentation of Fig. 1a has been used for many years and is now generally known as a "conventional detector". Its basic rationale is to use the coherence of the signal wavefront to generate maximum signal level at the summation point P. This presumably enhances the signal to noise ratio at P and therefore facilitates detection.

There are clearly other ways to attack the problem. Instead of using signal coherence to increase the signal level one might use noise coherence to cancel part of the noise. This is not always feasible. If the noise is independent from sensor to sensor, noise cancellation is not possible and the procedure of Fig. 1a appears to be optimal. On the other hand, if the spatial noise coherence is strong (and different from that of the signal) noise cancellation may well lead to greater improvement in signal to noise ratio than signal enhancement.

From a practical point of view the main problem with this second approach is that the designer rarely has very accurate information concerning the spatial structure of the noise. The obvious response to this difficulty is to gather information about actual environmental conditions while the system is in operation and to adapt various system parameters to match the observed conditions.

Since environmental conditions are likely to be quite variable and unpredictable one is almost forced into such adaptive procedures if one hopes to improve performance significantly beyond the level set by a conventional processor. However, the improvement is bought at a price: Increased complexity. The number of parameters to be adjusted in a fully adaptive system for a large array could be very high indeed and there is a real question whether the cost is commensurate with the potential improvement. This is the theme of the present paper. We shall seek to distinguish between features in signal and noise statistics which lend themselves to adaptation and those which do not. For this purpose we are not so much concerned with the properties of particular adaptive algorithms as with the maximum improvement which could be achieved with any algorithm. Fortunately it is not too difficult to set bounds on this maximum improvement: The ideal adaptive system would ultimately converge to the optimum system designed under full knowledge of signal and noise statistics. For our purposes we therefore need only compare the performance of the conventional system with that of the optimum system and identify signal and noise features which cause the difference between the two to be significant. In those cases where meaningful improvements are possible we must then inquire whether the relevant signal or noise features are readily identifiable, or whether their isolation requires configurations of great inherent complexity.

2. Basic Theory

We are concerned with the detection of a Gaussian signal in Gaussian noise. According to a well-known result in detection theory, the detector which maximizes detection probability for a given false alarm rate forms the likelihood ratio

$$\text{L.R.} = \frac{P(\underline{y})/S+N)}{P(\underline{y}/N)} \tag{1}$$

and compares it with a threshold. $p(\underline{y}/S+N)$ is the probability density of the data vector $\underline{y}$ when signal (as well as noise) is present, $p(\underline{y}/N)$ is the corresponding probability density in the absence of signal . Instead of working with L.R. one can, of course, use any monotone function of L.R. and compare it with an appropriately modified threshold. Because of the exponential form of the Gaussian distribution log L.R. immediately suggests itself as the most convenient test statistic.

If P is the covariance matrix of the signal component of $\underline{y}$ and Q that of the noise component

$$z = \log L.R. = \underline{y}^*[Q^{-1} - (P+Q)^{-1}]\underline{y} \qquad (2)$$

$\underline{y}$ is, in general, a complex data vector and the symbol ( )* denotes the conjugate-transpose of the bracketed quantity.

Eq. (2) is quite general: $\underline{y}$ may be any convenient representation of the data such as time samples, Fourier coefficents, or expansion coefficients associated with any complete orthonormal set. Note that we must represent the time function received at each of M sensors in this fashion. The dimension of $\underline{y}$ is therefore M times the number of coefficients required to represent each time function. This is likely to be a very large number and the matrix inversions required by Eq. (2) are therefore extremely cumbersome. However, with careful choice of the representation one can often circumvent this difficulty. In particular, it is almost invariably true in passive sonar problems that the observation time T is large compared with the correlation times of signal and noise. Under these conditions it is easy to prove that Fourier coefficients associated with different frequencies are uncorrelated so that a representation in terms of Fourier coefficients leads to block diagonal covariance matrices P and Q. Then Eq. (2) reduces to the simpler form

$$z = \sum_{k=1}^{n} \underline{y}_k^*[(N_kQ_k)^{-1} - (S_kP_k + N_kQ_k)^{-1}]\underline{y} \tag{3}$$

$\underline{y}_k$ is the vector of Fourier coefficients associated with frequency $\omega_k = 2\pi k/T$. Its dimension is M, the number of sensors. Q_k is the covariance matrix of the noise Fourier coefficients at frequency ω_k, normalized so that $Tr(Q_k) = M$. N_k is the noise spectrum at ω_k. Similarly P_k is the normalized covariance matrix of signal Fourier coefficient and S_k describes the signal spectrum. The normalization has been introduced to separate spectral features of signal and noise (described by S_k and N_k) from spatial features (characterized by P_k and Q_k).

If the number of sensors is at all large, the matrix inversions of Eq. (3) still require elaborate computational procedures. Analytical results can be obtained in two practically important cases

a) $S_k/N_k << 1$ (low signal to noise ratio at all frequencies).

In that case

$$(S_kP_k + N_kQ_k)^{-1} \simeq (N_kQ_k)^{-1}[I - S_kP_k(N_kQ_k)^{-1}] \tag{4}$$

Substituting (4) into (3):

$$z \simeq \sum_{k=1}^{n} \frac{S_k}{N_k^2} \underline{y}_k^* Q_k^{-1} P_k Q_k^{-1} \underline{y}_k \tag{5}$$

Since P is non-negative definite it can be factored as follows

$$P_k = B_k B_k^* \tag{6}$$

B_k is an M x M matrix. Substituting (6) into (5)

$$z = \sum_{k=1}^{n} \frac{S_k}{N_k^2} (\underline{y}_k^* Q_k^{-1} B_k)(B_k^* Q_k^{-1} \underline{y}_k)$$

$$\equiv \sum_{k=1}^{n} \frac{S_k}{N_k^2} \underline{w}_k^* \underline{w}_k = \sum_{k=1}^{n} \frac{S_k}{N_k^2} \sum_{i=1}^{M} |w_{ki}|^2 \qquad (7)$$

Here the vector $\underline{w}_k$ is defined as

$$\underline{w}_k = B_k^* Q_k^{-1} \underline{y}_k \qquad (8)$$

w_{ki}, its i^{th} component, is a linear combination of the data associated with frequency ω_k. Consider now one component of the i sum in the last version of Eq. (7).

$$z_i = \sum_{k=1}^{n} \frac{S_k}{N_k^2} |w_{ki}|^2 \approx \frac{T}{2\pi} \int_{-\infty}^{\infty} \left| \frac{\sqrt{S(\omega)}}{N(\omega)} w_i(\omega) \right|^2 d\omega \qquad (9)$$

To reach the integral form of this equation we have assumed that the signal and noise spectra vary slowly over intervals of the order $2\pi/T$. We can now use Parseval's theorem to replace the frequency integral by an equivalent time integral and the optimum instrumentation therefore assumes the form shown in Fig. 2

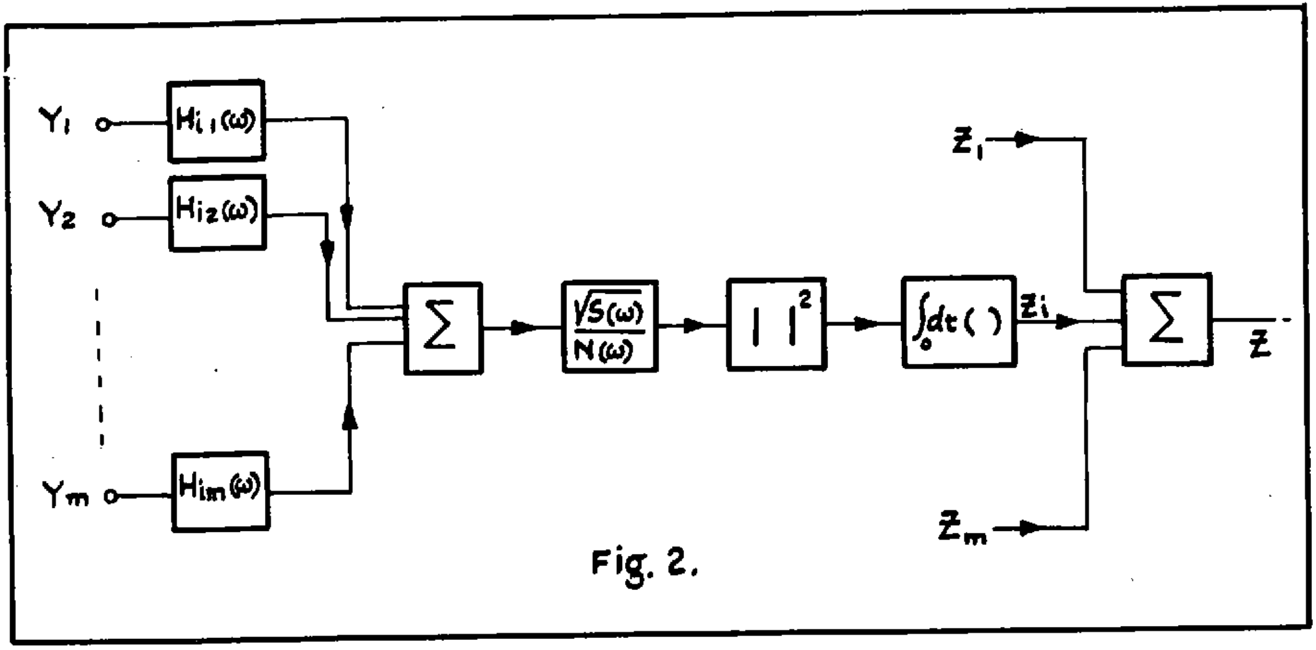

Fig. 2.

The frequency sensitive filters $H_{ij}(\omega)$ are specified by

$$H_{ij}(\omega_k) = (\underline{b}^*_{ki}Q_k^{-1})_j \tag{10}$$

where $\underline{b}_{ki}$ is the i^{th} column of B_k.

Note that the complete instrumentation of Fig. 2 has M channels $(z_1 \ldots z_M)$ whose outputs are added. The final smoothing integration is the same for each and can be made common. There are M identical filters $\sqrt{S(\omega)}/N(\omega)$ which compensate for spectral differences of signal and noise. To adapt for such differences is therefore a relatively simple matter. On the other hand there are M^2 filters $H_{ij}(\omega)$ whose purpose is to use spatial properties of signal and noise in optimal fashion. Unless M is very small, full spatial adaptation will therefore be an extremely complex task.

b. Coherent signal wavefront

The second important case in which useful analytical results can be obtained is that of a point source signal, which generates a planar wavefront if it is in the far field, a spherical wavefront if it is in the near field. In either case the signal waveshapes received at different sensors are identical except for a delay. P_k is then simply an outer product of a vector with itself.

$$P_k = \underline{V}_k^* \underline{V}_k \tag{11}$$

the "steering vector" $\underline{V}_k$ has elements

$$(V_k)_i = e^{j\omega_k \tau_i} \tag{12}$$

τ_i is the travel time of the signal from the source to the i^{th} sensor.

Substituting (11) into (3) we can now use the standard matrix identity

$$(Q + \underline{V}^*\underline{V})^{-1} = Q^{-1} - \frac{Q^{-1}\underline{V}\underline{V}^*Q^{-1}}{1 + \underline{V}^*Q^{-1}\underline{V}} \tag{13}$$

A few lines of algebra yield

$$z = \sum_{k=1}^{n} \frac{S_k/N_k^{\,2}}{1 + \frac{S_k}{N_k}\underline{V}_k Q_k^{-1}\underline{V}_k} (\underline{y}_k^* Q_k^{-1}\underline{V}_k)(\underline{V}_k^* Q_k^{-1}\underline{y}_k) \tag{14}$$

The similarity with Eq. (7) is striking. The principal difference is that the M x M matrix B_k is replaced by the M vector $\underline{V}_k$. As a result the optimal instrumentation has only one of the M channels appearing in Fig. 2. The appropriate block diagram is shown in Fig. 3. The implications for adaptive processing are of obvious importance. When the signal wavefront is coherent, spatial adaptation requires the control of only M rather than M^2 filters. For moderate M full adaptation is now at least a realistic possibility. In practice, of course, signals are never generated by perfect point sources. What is lost by working with Fig. 3 rather than Fig. 2? Straightforward but

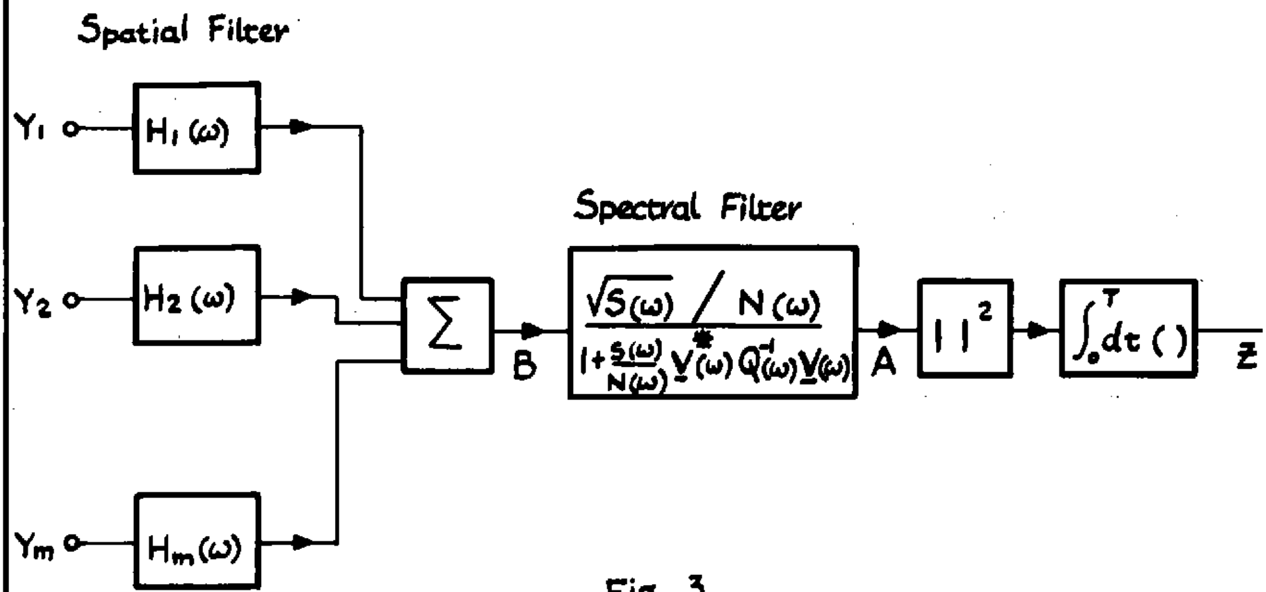

Fig. 3.

tedious calculation show[1] that the sacrifice is small as long as the angle subtended by the source at the receiver is much smaller than a receiver beam-width. This is very generally the case in practical sonar problems. Hence there is little incentive to go to a structure more complex than Fig. 3 as the starting point for adaptive procedures.

We note in passing that the spectral filter in Fig. 3 contains the spatial noise matrix Q. At low signal to noise ratios its effect is clearly small, but at high signal to noise ratios the spatial and spectral filtering operations are no longer uncoupled, with obvious negative implications for the complexity of the best adaptive processor.

We also note that Fig. 3 is optimal for detection. Suppose one is interested in maximizing signal to noise ratio or obtaining the best possible replica of the signal at a point such as A (perhaps in order to facilitate

---

[1] Bangs, W.J. Array Processing with Generalized Beamformers. Ph.D. dissertation, Yale University, 1971.

extraction of information from the signal). One finds[1] that the resulting structure is identical with Fig. 3 up to point B. Precisely the same spatial filtering operation needs to be performed, only the spectral filter changes somewhat. It is apparent, therefore that the spatial adaptation process is basically the same, regardless whether the ultimate objective is detection or the estimation of one or more signal parameters.

3. Array Gain

To estimate the improvement in performance potentially available through adaptation we must compare the performance of the conventional system with that of the optimal system. Since we are primarily concerned with the effect of spatial processing we shall characterize performance by the "array gain", the signal to noise ratio in an incremental frequency interval at point B (or A) divided by the signal to noise ratio in a similar frequency interval at the array input.

Let $\underline{H}(\omega_k)$ be the vector of filter gains at frequency ω_k, Then the amplitude of the k^{th} frequency component at point B is $\underline{H}^*(\omega_k)\underline{y}_k$. Therefore the ratio of signal power to noise power at point B is

$$(S/N)_0 = \frac{\underline{H}^*(\omega_k)\overline{\underline{y}_k^{(s)}y_k^{(s)*}}\,\underline{H}(\omega_k)}{\underline{H}^*(\omega_k)\overline{y_k^{(n)}y_k^{(n)*}}\,\underline{H}(\omega_k)} = \frac{S(\omega_k)}{N(\omega_k)}\,\frac{\underline{H}^*(\omega_k)P(\omega_k)\underline{H}(\omega_k)}{\underline{H}^*(\omega_k)Q(\omega_k)\underline{H}(\omega_k)} \tag{15}$$

The superscripts (s) and (n) denote signal and noise components respectively and the overbar represents a statistical average. Dividing (15) by the input signal to noise ratio one obtains the array gain

[1]Edelblute, D.J., Fisk, J.M. and Kinnison, G.L. Criteria for Optimum Signal Detection Theory for Arrays, JASA 41, 199. Jan. 1967.

$$G(\omega) = \frac{\underline{H}^*(\omega)P(\omega)\underline{H}(\omega)}{\underline{H}^*(\omega)Q(\omega)\underline{H}(\omega)} \tag{16}$$

The vector of filter functions $\underline{H}$ is completely general. We are mainly interested in the conventional processor of Fig. 1 and the optimal processor. In the conventional case $\underline{H}^*$ is simply a set of delay operators which align the signal components, i.e. $\underline{H}$ is the $\underline{V}$ vector defined by Eq. (12). Then (suppressing the frequency variable ω)

$$G_{conv} = \frac{\underline{V}^*\underline{V}\underline{V}^*\underline{V}}{\underline{V}^*Q\underline{V}} = \frac{M^2}{\underline{V}^*Q\underline{V}} \tag{17}$$

On the other hand the optimal processor uses

$$\underline{H} = Q^{-1}\underline{V} \tag{18}$$

and the array gain becomes

$$G_{opt} = \frac{\underline{V}^*Q^{-1}\underline{V}\underline{V}^*Q^{-1}\underline{V}}{\underline{V}^*Q^{-1}QQ^{-1}\underline{V}} = \underline{V}^*Q^{-1}\underline{V} \tag{19}$$

If the noise is spatially white (uncorrelated from sensor to sensor) $Q = I$ and

$$G_{opt} = G_{conv} = M \tag{20}$$

Thus the primitive conventional processor is optimal and no adaptation can improve performance. In practice we frequently encounter situations where major fractions of the noise power are spatially uncorrelated. An array gain of M therefore serves as a useful benchmark. Adaptation is worth considering if the conventional detector falls far below this performance level and/or the optimal detector promises an array gain well in excess of M. In the light of Eq. (19) the latter possibility arises primarily when the Q matrix is near-singular.

Since noise coherence is the key feature distinguishing (17) from (19) we begin with the extreme case: A noise containing a strong coherent component (an interference). If the noise power is the same at each sensor the normalized noise covariance is now

$$Q_k = \varepsilon I + (1-\varepsilon)\underline{V}_I\underline{V}_I^* \tag{21}$$

$\underline{V}_I$ is the steering vector of the interference and ε is the fraction of the total noise power which is spatially incoherent. Frequency dependences of ε and $\underline{V}_I$ have been suppressed for simplicity of notation. By direct computation from Eq. (17)

$$G_{conv} = \frac{M^2}{\varepsilon M + (1-\varepsilon)|\underline{V}^*\underline{V}_I|^2} \tag{22}$$

$|\underline{V}^*\underline{V}_I|^2$ is the (normalized) interference power appearing in a beam steered on the signal. It is proportional to M^2 and (except for the isolated frequencies for which a far field interference might lie in a null of the beam pattern) can easily be the dominant term in the denominator of Eq. (22). Thus the conventional array gain can be substantially smaller than M. The optimum array gain is easily calculated from Eq. (19), using the matrix identity (13).

$$G_{opt} = \frac{1}{\varepsilon}[M - \frac{\frac{1-\varepsilon}{\varepsilon}}{1 + \frac{1-\varepsilon}{\varepsilon}M}|\underline{V}^*V_I|^2] \tag{23}$$

If the interferences is at all strong

$$\frac{1-\varepsilon}{\varepsilon} M >> 1 \tag{24}$$

and Eq. (23) becomes

$$G_{opt} \simeq \frac{1}{\varepsilon}[M - \frac{1}{M}|\underline{V}^*\underline{V}_I|^2] \tag{25}$$

If the interference is separated from the signal by more than a beamwidth $|\underline{V}^*\underline{V}_I|^2 \ll M^2$ so that

$$G_{opt} \simeq \frac{M}{\varepsilon} \tag{26}$$

Thus one can achieve a performance almost equal to that attainable when the interference is absent entirely. Substantial gains can therefore be made by adapting to such a noise field. Nor is the required instrumentation terribly complex. All one needs is a good replica of the interference and an estimate of its location, which can be obtained by adapting one beam (for a far field interference) or the outputs of a few widely scattered sensors (for a near field interference). The estimate, properly delayed, can then be subtracted from each sensor output to generate essentially interference-free data. Adaptation is therefore particularly promising for combating strong noise sources located near the observing platform.

The above argument is easily generalized to an environment containing two coherent noise components

$$Q_k = \varepsilon I + \frac{1-\varepsilon}{2} \underline{V}_1\underline{V}_1^* + \frac{1-\varepsilon}{2} \underline{V}_2\underline{V}_2^* \tag{27}$$

$\underline{V}_1$ and $\underline{V}_2$ are the steering vectors of the two coherent noise components. For simplicity their power is assumed to be equal. The equivalent of Eq. (23) is now [for $\frac{1-\varepsilon}{2\varepsilon} M \gg 1$]

$$G_{opt} = \frac{1}{\varepsilon}\left\{M - \frac{1}{M}\underline{V}^*\underline{V}_1 - \frac{\frac{1-\varepsilon}{2\varepsilon}\left[|\underline{V}^*\underline{V}_2|^2 + \frac{1}{M^2}|\underline{V}^*\underline{V}_1|^2|\underline{V}_1^*\underline{V}_2|^2 - \frac{1}{M}\underline{V}^*(\underline{V}_1\underline{V}_1^*\underline{V}_2\underline{V}_2^* + \underline{V}_2\underline{V}_2^*\underline{V}_1\underline{V}_1^*)\underline{V}\right]}{1 + \frac{1-\varepsilon}{2\varepsilon}\left(M - \frac{1}{M}\underline{V}_2^*\underline{V}_1\right)}\right\} \tag{28}$$

Eq. (28) furnishes several interesting insights:

(1) If $|V_1^* V_2|$ is small (i.e. the two interfering sources are well separated from each other)

$$G_{opt} \approx \frac{1}{\varepsilon}\{M - \frac{1}{M}\underline{V}^*\underline{V}_1 - \frac{1}{M}\underline{V}^*\underline{V}_2\} \tag{29}$$

If both are well separated from the signal one can again approach the performance of Eq. (26).

(2) If $\underline{V}_1 \approx \underline{V}_2$ (interferences nearly coincident)

$$G_{opt} \approx \frac{1}{\varepsilon}\{M - \frac{1}{M}\underline{V}^*\underline{V}_1\} \tag{30}$$

Thus the two nearby interferences act as one strong interference which can be eliminated with ease.

(3) If the interferences are well separated from each other and from the signal, an obvious procedure for eliminating them is to steer the array on each interference with a separate set of delays and use the interference replicas thus generated to cancel both interferences from each sensor output. Alternatively one can, of course, combine all of these operations into an instrumentation of the form of Fig. 3. The amplitude and phase characteristics demanded of the filters $H_i(\omega)$ are now quite complex and vary rapidly with frequency. The adaptive filter must therefore come quite close to the optimum before satisfactory performance is achieved.

(4) Once can clearly extend the argument to more than two interferences. In fact, one can view an arbitrary noise field as generated by a large number of small, widely scattered point sources. If their number exceeds M one can, of course, not cancel them perfectly and there will be a residual error.

If no M of these sources contribute a significant fraction of the total noise power the value of the entire cancellation process becomes questionable and one suspects that the simple conventional processor may not be too far from the optimum. In the next section we shall examine this central issue from a different and perhaps more intuitively appealing point of view.

4. Space Frequency Analysis

Here we confine our attention to far field signals and noises. To avoid cumbersome trigonometric manipulations we assume that our receiving array is linear. Suppose a sinusoidal signal of frequency ω_0 is incident at an angle θ on an array of length L, as suggested by Fig. 4. If the signal

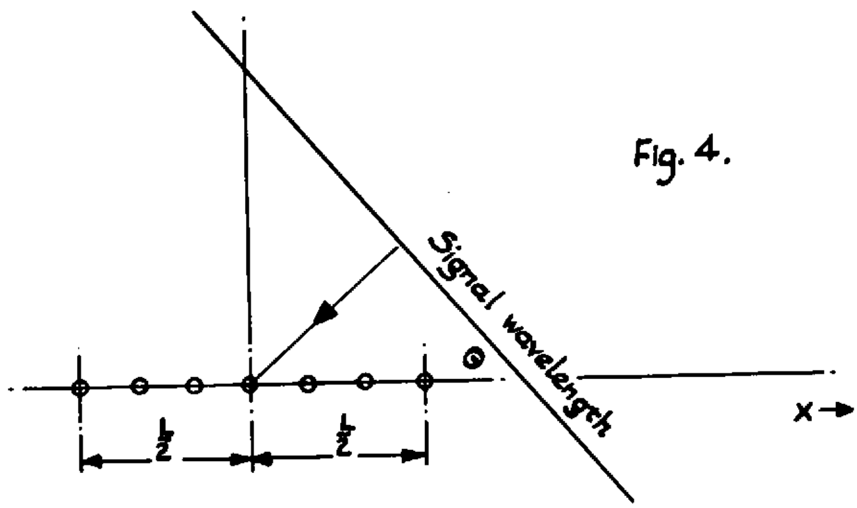

Fig. 4.

received at the origin of coordinates is

$$S_0 = Ae^{j\omega_0 t} \tag{31}$$

then the signal received at point x on the array is

$$s(x) = Ae^{j\omega_0(t+\frac{x}{c}\sin\theta)}$$

$$= S_0e^{j\omega_0\frac{x}{c}\sin\theta} \equiv S_0e^{j\nu_s x} \tag{32}$$

Here

$$\nu_s = \frac{\omega_0}{c}\sin\theta \tag{33}$$

is the space frequency (wave number) of the signal. c is the velocity of sound.

The important point in Eq. (32) is that, viewed as a function of x, the signal is a sinusoid characterized by the frequency ν_s (which specifies the direction of arrival). Suppose, for the moment, that the noise is spatially white (uncorrelated from sensor to sensor). We are now dealing with a spatial version of the classical detection problem: Detect a known wave-shape (a sinusoid) in a background of white noise. The solution is well known: Crosscorrelate the received noisy waveshape y(x) with a replica of the known signal. In our case the required test statistic is therefore

$$z = \int_{-\frac{L}{2}}^{\frac{L}{2}} y(x)e^{-j\nu_s x}dx = \int_{-\frac{L}{2}}^{\frac{L}{2}} y(x)e^{-j\omega_0\frac{x}{c}\sin\theta}\,dx \tag{34}$$

$(x/c)\sin\theta$ is the time shift of the signal at point x on the receiving array relative to that at the origin. Hence the operation specified by Eq. (34) aligns the signal components at all points on the receiving array. It is nothing more than a space-continuous version of the conventional detector whose optimality in a white noise environment is therefore obvious from elementary detection theory.

It is useful to push the space-time analogy somewhat further. If the noise is not white but the observation interval is large compared with the noise correlation interval, elementary detection theory tells us that the best processor prewhitens the noise and then crosscorrelates with an appropriately modified replica of the signal. Figure 5 gives the spatial version of this procedure. $N(\nu)$ is the spatial noise spectrum. We may think of it as being

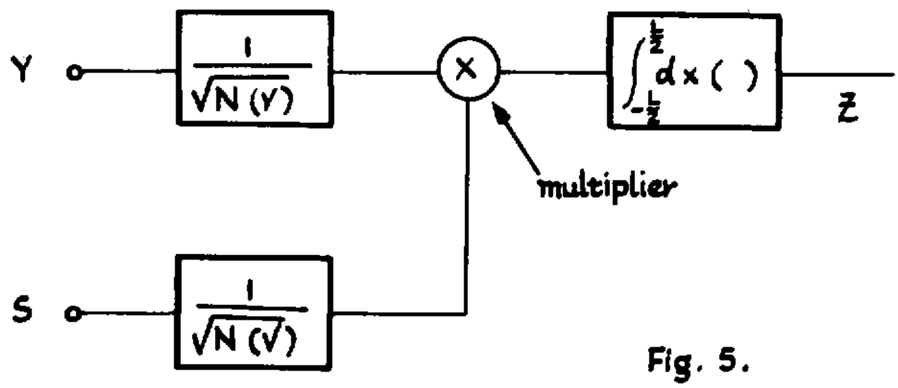

Fig. 5.

generated by a large number of infinitesimalplane wave contributions coming from all possible directions. Figure 6 gives a typical pattern for a far field noise concentrated in a direction to the left of broadside. ω_0/c is the

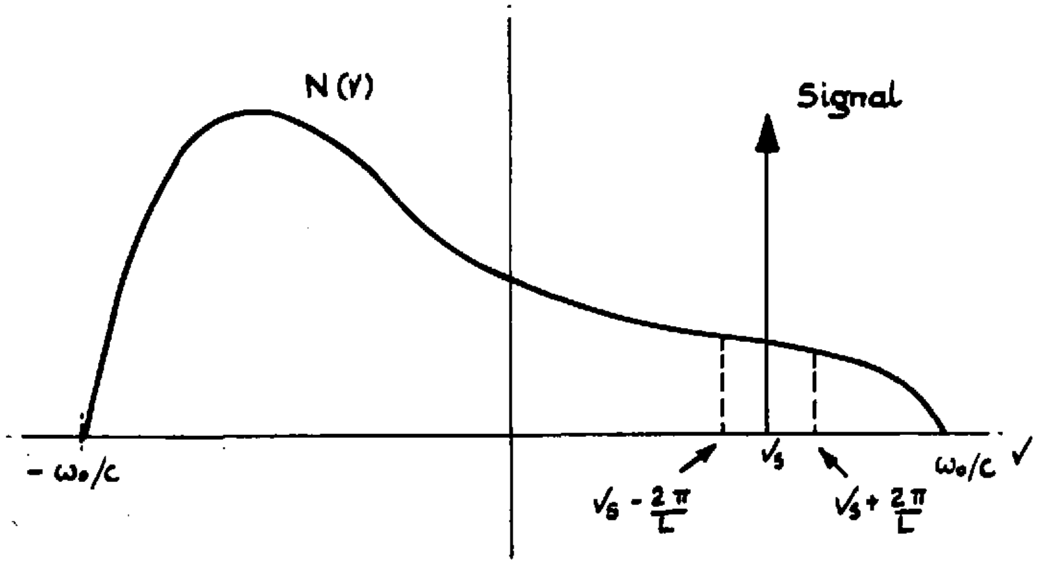

Fig. 6.

space frequency associated with plane waves incident from the endfire direction and is therefore an absolute bound on space frequencies which can be associated with a far field noise source.

If the noise spectrum $N(\nu)$ happens to have the shape of Fig. 6 and the signal is located as suggested in that figure, we can draw an interesting conclusion. The integrator in Fig. 5 is a spatial low pass filter of bandwidth $2\pi/L$. Since s is a pure sinusoid of frequency ν_s it follows that only those frequency components of y which lie within $\pm 2\pi/L$ of ν_s contribute significantly to the output z. If the noise spectrum is essentially flat over this band (as suggested by Fig. 6) the prewhitening operation becomes unnecessary. One can therefore conclude:

If the array length is large compared with the noise correlation distance and if the noise spectrum is essentially flat over $(\nu_s - 2\pi/L, \nu_s + 2\pi/L)$ the conventional detector is nearly optimal.

The first of these conditions rules out major components of the noise field concentrated in a space frequency interval much smaller than a beamwidth and therefore confirms our previous observation that these will be among the most logical features to be exploited by adaptive schemes.

The observation that far field signals and noises are simply spatial sinusoids suggests that it may be profitable to work with spatial Fourier coefficients to represent the received data. The received sound field is characterized completely by the set of Fourier coefficients

$$c_n = \int_{-\frac{L}{2}}^{\frac{L}{2}} y(x) e^{-j\frac{2\pi n}{L}} dx \qquad (35)$$

Comparison with Eq. (34) shows that this is simply the output of a conventional beamformer steered in the direction

$$\sin\theta_n = 2\pi n \frac{c}{\omega_0} \frac{1}{L} = \frac{\lambda_0 n}{L} \tag{36}$$

λ_0 is the acoustic wavelength. By proper choice of the time frequency origin one can always align the signal with one of these directions. [See Fig. 7].

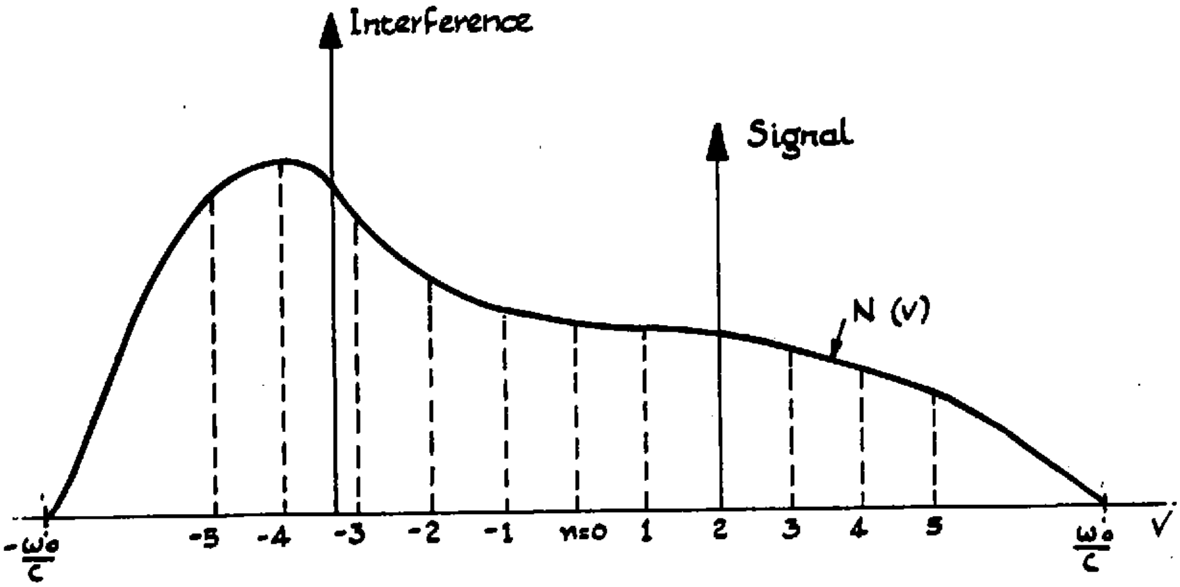

Fig. 7.

In that case all of the signal is contained in one beamformer output (n = 2 in our example). All of the other beams contain nothing but noise. They are useful only to the extent that their outputs are correlated with the noise on the signal beam. If a major fraction of the noise is concentrated in a small spatial region (an interference) the appropriate beam output will be strongly correlated with the noise on the signal beam and can therefore be used to reduce it in an adaptive procedure.

If there is no strong, spatially concentrated component of the noise, correlations with the noise on the signal beam will be weak and no single noise beam can achieve much reduction of that noise. This does not imply that major improvements might not be made through the use of many noise beams. Consider, for example, a spatially isotropic noise [Fig. 8]. The

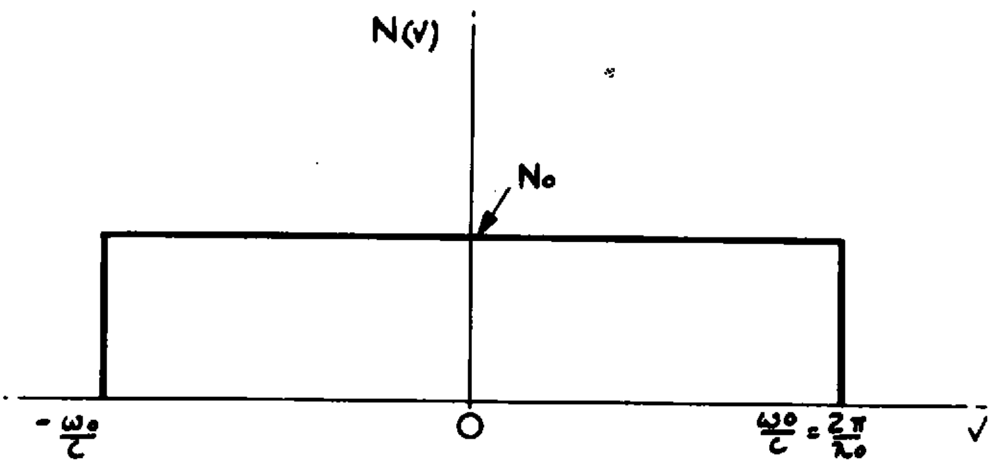

Fig. 8.

correlation distance of such a noise is of the order of an acoustic wavelength. If $\lambda_0 \ll L$ and $M \gg 1$ a basic theorem of functional analysis asserts that the eigenvalues of the spatial noise covariance NQ follow the pattern of the spatial spectrum $N(\nu)$. In our example there will therefore be a certain number ($\simeq L/\lambda_0$) of eigenvalues of magnitude near N_0 and all of the remainder will have magnitudes close to zero. The noise covariance matrix Q is therefore almost singular and Eq. (19) suggests a very large potential for improvement.[1] This appears to be in direct contradiction with our earlier assertion that the

[1] The phenomenon is often referred to as "superdirectivity" or "supergain".

conventional detector is near-optimal for a noise field whose spatial spectrum is essentially flat over an interval of $\pm 2\pi/L$ near the space-frequency of the signals. The dilemma is very real from a formal point of view, but not from a practical one. In claiming near optimality of the conventional detector we were working with local behavior, ignoring remote edge effects. It is precisely these edge effects which the formally optimal detector seeks to exploit. We have already observed that each beam provides little information about the noise on the signal beam so that many such beams must be used for any major improvement. We note from Eq. (36) that these beams no longer correspond to real angles once $n > L/\lambda_0$. In order to make frequencies above ω_0/c accessible we must spatially sample above the spatial Nyquist rate of $\lambda_o/2$. Thus the number of sensors and the associated problem of adaptation increases enormously. In terms of Fig. 3, we not only have a very large number of filters $H_i(\omega)$ to adapt, but their phase and amplitude characteristics must be controlled to a degree of precision which quickly becomes prohibitive.[1] The situation is even worse if one considers the locally generated (spatially white) noise which is inevitably present at each sensor. Since the isotropic noise component received at space frequencies above ω_0/c is very small, it is easily overwhelmed by the white noise and the beam represented by this space frequency is then virtually useless. From a practical point of view, therefore, the assertion that the conventional detector is near-optimal would be difficult to criticize.

The lesson to be learned for adaptive processing is that procedures which seek to realize superdirective behavior are subject to very severe

[1]For a discussion of the resulting sensitivity problem see: Cox, H., Sensitivity Considerations in Adaptive Beamforming. NATO Advanced Study Institute on Signal Processing, Loughborough 1972, p. 619.

limitations. Adaptation comes into its own when the noise field exhibits pronounced spatial features sufficiently concentrated so that they cannot be resolved by conventional beams. It becomes particularly attractive when the features are characterized by a small number of parameters (such as location and strength of a few interferring sources). In other words, adaptation is no substitute for a critical examination of the noise field in which one expects to operate and a careful selection of key noise field parameters about which one needs to gather information. The object of the present paper has been to provide some guidelines for this selection process.