LEVEL II

14 NUC-TN-665

NAVAL UNDERSEA RESEARCH AND DEVELOPMENT CENTER

San Diego, California 92132

1

MOST Project -3

6 A NONMATHEMATICAL EXPLANATION

OF

ADAPTIVE BEAMFORMING.

9 Technical note,

10 by
BURWELL GOODE
Sonar Techniques Division
Sensor & Fire Control Department

12 19 p.

D D C
RECEIVED
APR 3 1979
F

11 Mar 72

Technical Notes are working documents subject to revision or expansion and are not to be referenced in formal publications. They are intended for recording current pertinent technical information.

404 762

CONTENTS

i

1. INTRODUCTION

Optimum array processing, adaptive processing, and adaptive beamforming are terms which have been mentioned in the sonar literature with increasing frequency over the past decade. Are these three terms synonymous? The equations which describe the behavior of this type of signal processing can be long and involved. Stripped of the mathematical details, what are the underlying principles of optimum array processing, adaptive processing and adaptive beamforming? This paper presents answers to these questions without matrix algebra.

2. OPTIMUM ARRAY PROCESSING

Figure 1 presents a block diagram of an array processor system. The hydrophone array samples the signal and noise fields propagating through the medium. The processor combines the hydrophone outputs to steer a beam in a particular direction. Each hydrophone output contains signal and noise from the steering direction, noise from other directions, and nondirectional noise. The proportion of the beam output due to signal and noise from the steering direction is higher than
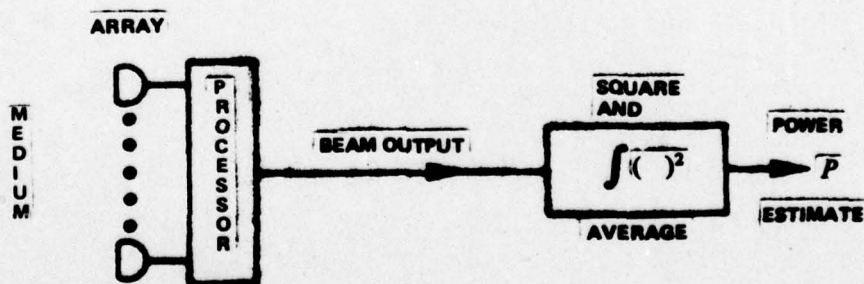
Figure 1. The Array Processor System

1

the corresponding proportion of the output of a single hydrophone. However, some of the beam output is still due to sources which come from a direction substantially different from the steering direction and some of it is nondirectional noise.

The power estimate P is a random variable. If the averaging time is short, P will have a large variance, and its probability density function (*pdf*) might look like the one in Figure 2. If the averaging time is longer, P will have a smaller variance, and its *pdf* might look like the one in Figure 3. The expected (or average) value of the power estimate $E(P)$ is the same in both cases.

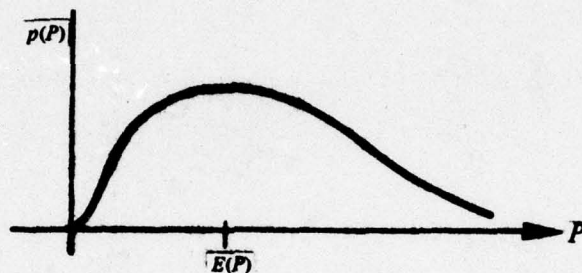Figure 2. Probability Density Function for Short Integration Time
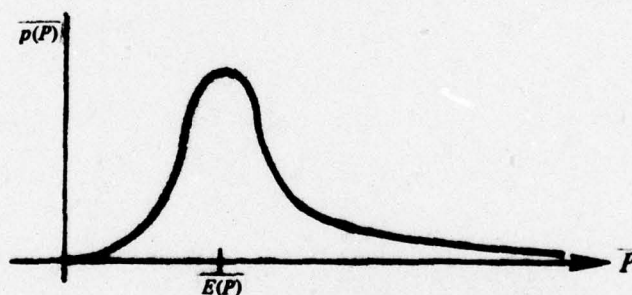

Figure 3. Probability Density Function for a Long Integration Time

Note that the area under the curves in Figures 2 and 3 is supposed to be equal to unity. In order to evaluate the behavior of the probability density function as the signal and noise fields change, the following section will consider the integration time to be fixed.

The classical detection theory problem assumes that the statistics of the noise field are known. That assumption means that the processor designer knows the exact probability density function of the power estimate P when signal is absent. The detection problem is to determine whether signal is present. Signal in the steering direction shifts the probability density function curve to the right. The probability density function for noise only is sketched in Figure 4, and the probability density function (*pdf*) for signal plus noise is sketched in Figure 5.

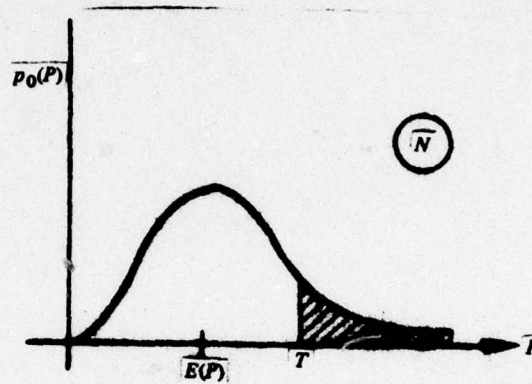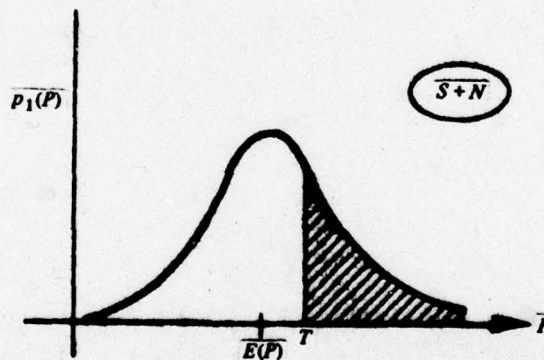Figure 4. Power *pdf* for Noise Alone

Figure 5. Power *pdf* for Signal Plus Noise

The processor can not directly observe a shift in the *pdf*. The operator (or an automated decision computer) must make a decision as to whether or not signal is present, based upon a single number P'. That number P' is a sample of the random variable P drawn from one of the distributions shown in Figures 4 and 5.

The classical detection theory problem is equivalent to the following problem: Given two probability distributions with the *pdf*'s shown in Figure 4 and 5, decide whether the sample P of the random variable P came from the distribution with *pdf* $p_0(P)$, or whether it came from the distribution with *pdf* $p_1(P)$. Clearly, this can not be done with certainty. Statistical decision theory tells us to set a threshold T and decide that signal is present if $P > T$. Otherwise, it is decided that signal is absent. The probability that the signal is correctly detected is equal to the shaded area under the curve in Figure 5. The probability of a false alarm is shown by the shaded area under the curve in Figure 4.

In this kind of detection problem it is not possible to know with certainty whether or not the signal is present because there is no finite range of P such that *pdf* is zero and the other is nonzero. The subject of where to set the threshold involves a value judgement as to the relative importance of correct detections and false alarms. For a given processor, when the threshold for a given false alarm probability is set, the detection probability is set simultaneously. However, the amount by which a signal shifts the *pdf* to the right varies for different processors. A normalized measure of the spread between the *pdf* of the output power when signal plus noise is present and the *pdf* of the output power for noise alone is the signal-to-noise ratio. The expected value of the output power when signal plus noise is present is denoted by $E(P/S + N)$. The expected value of the output power when noise alone is present is denoted by $E(P/N)$. The difference between the two expected power outputs is a measure of the change caused by the addition of the signal. In order to compare different types of processors, the difference of the means is normalized by the square root of the variance of the output power when noise alone is present, Var (P/N). Thus, the signal-to-noise ratio is defined as

$$ G = \frac{E(P/S + N) - E(P/N)}{[\text{Var}\,(P/N)]^{1/2}} $$

The denominator is a measure of the spread of the *pdf*. A change in the mean is easier to detect if the fluctuation in output power is small. This detectability can be measured either by the signal-to-noise ratio or by the probabilities of detection and false alarm.

A further normalization is sometimes used to allow a comparison of performance under different signal and noise conditions. The signal-to-noise ratio at the output of the processor is divided by the signal-to-noise ratio at the input to the processor (that is, at the output of a single sensor) to produce the *array processing gain A*:

4

$$A = \frac{G_{out}}{G_{in}}$$

In classical detection theory the noise is assumed to be stationary. This means that the statistical properties of the noise do not change with time. For example, it means that the *pdf* of the power estimate P at the beam output will not change from one hour to the next. If the signal duration is long, then, in principle, a large output signal-to-noise ratio G can be obtained with the integration of the output power estimate over a long time period. The long averaging time results in a low variance Var(P/N) of the power estimate P, facilitating the detection of small differences between expected output power with and without signal.

Stationarity is a relative concept. Nothing is stationary in terms of geological time. On the other hand, sea noise is usually stationary from one second to the next. Sea noise is composed of surface noise, biological noise, shipping, and other man-made noise. An hour is a long time for biological noise to remain stationary, and ten minutes is a long time for shipping noise to remain stationary. The orientation of a sonar platform (e.g., submarine) can change significantly in less than a minute, which will effect the relative directional nature of the noise observed by the sonar array.

Because of the limited time during which the signal and noise fields can be regarded as stationary, signal processing involves more than just time averaging. When either the signal duration or the nonstationarities of the signal and noise fields limit the possible averaging time, there is a lower limit on the variance of the power estimate. Further improvement in performance can only be obtained by more sophisticated techniques to accentuate the difference between the probability density functions of the power estimates, with and without signal. If the signal and noise fields are assumed stationary for a certain length of time, with known statistical properties, classical detection theory describes how to maximize the difference between the output *pdf*'s, with and without signal, which will maximize the probability of detection for a fixed false alarm rate. If the noise field is Gaussian, the property of the noise field which is required for optimum array processing is the cross-power spectral matrix. When noise alone is present, this matrix gives the correlations between all possible pairs of hydrophones as a function of frequency.

Optimum processing theory is valuable in two ways: 1) possible improvement in performance over conventional processing can be calculated for given signal and noise statistics; and, 2) possible receiver structure is suggested. The work of Bryn (1) and Vanderkulk (2) showed that for isotropic noise, little improvement could be achieved over conventional processing performance.

5

However, the work of Edelblute (3), Kinnison (4), Shapard (5), Kneipfer (6), and others, has shown that sonars are confronted with highly anisotropic noise fields under certain operating conditions, and optimum processing can result in a marked improvement in performance over conventional processing.

There are three principle limitations to the optimum processing theory: 1) the theory often leads to filters which are too complex to be implemented within the economic limitations of the problem; 2) the exact direction of the signal field may not be known; 3) very little is usually known about the noise field when the system is being built. The problem of uncertainty as to the signal direction is met by forming adjacent beams steered in different directions. However, if the design is based on inaccurate knowledge of the statistics of the signal and noise fields, there is no assurance that the "optimum" processor will perform better than the standard beamformer. In many array problems the statistics of the noise field change, and it is impossible to design a fixed processor which is optimum according to the criteria of classical detection theory.

3. ADAPTIVE PROCESSING

Even though the statistics of a noise field are unknown, the parameters of the statistical distribution can, in principle, be estimated if the statistics are essentially stationary over some observation interval. For example, if the noise field is Gaussian, its statistical distribution is completely determined by the cross-power spectral matrix as a function of frequency. A signal processing system could be designed to estimate the cross-power spectral matrix of the array at a number of frequencies. These matrices could be used to form beams which would be optimum if the estimated cross-power spectral matrices were the true ones. This approach is sometimes called the matrix inversion approach because the estimated matrices must be inverted before the beam outputs can be formed. Since the beam outputs are not fixed linear combinations of the input data, but involve a weighting of the input data by the estimated inverse of the cross-power spectral matrices, this approach is called adaptive. The term "adaptive" refers to systems which operate differently in different environments and usually implies an improvement over systems which perform the same simple operations on the input data regardless of the environment.*

---

*The term "adaptive" arises from a heirarchy in our understanding of the processor. It is convenient to think of a linear weighting of the input data, with the weights determined adaptively. The adaptive process for finding the weights is a highly nonlinear combination of the input data. The algorithm for finding the weights is a fixed computational procedure. Taken as a whole, the algorithm and the linear weighting together comprise a fixed nonlinear processor. Although it is a good idea to remember that the adaptive weights are functions of the input data, it is confusing to look at an explicit formula for the beam output in terms of constants and the input data.

There are several ways to estimate a parameter associated with a statistical distribution. One possible way is to use the data sample to compute a *sufficient statistic* for the parameter. A statistic is sufficient if, when the statistic is computed, the calculation of other statistics with the same data will not improve the estimate of the parameter. Beamforming can be interpreted in this context. The output of the beam can be interpreted as an estimate of the signal in the steering direction, and the signal can be interpreted as a parameter of the statistical distribution of the outputs of the array. Under certain reasonable assumptions, the best estimate of the signal is a linear weighting of the outputs of the array. This best estimate of the signal is a sufficient statistic for the signal. Again under reasonable assumptions, the optimum detector can be obtained from the best estimator if the signal estimates are squared and averaged.

The optimum weights used in the linear combination of the array outputs to achieve the best estimate of the signal are functions of the inverse of the cross-power spectral matrix of the array. When the statistics of the noise field are not known, then an attempt can be made to learn enough about the noise field to estimate the optimum weights. One method is to estimate the cross-power spectral matrix, invert it, and then calculate what the optimum weights would be if the estimated cross-power spectral matrix were the true one. However, there are $K(K+1)/2$ different elements in the cross-power spectral matrix for an array of K hydrophones. For one beam only K weights are needed. If only one beam is to be formed, it is easier from a computational point of view and better from a statistical point of view to estimate the K beamforming weights directly, rather than to compute them from the K^2 elements of the estimated cross-power spectral matrix. That is, it is sufficient to estimate K weights instead of $K(K+1)/2$ matrix elements. Only when the number of adaptive beams approaches K does inversion of the estimated cross-power spectral matrices become attractive.

Several algorithms can be used to estimate the optimum beamforming weights. Stochastic approximation algorithms provide the best convergent properties in a stationary environment; however, they perform poorly in a nonstationary environment. Steepest descent adaptive algorithms are similar to stochastic approximation; their convergent properties are weaker in a stationary environment, but they can perform reasonably well in a slowly varying statistical environment. These algorithms form beam outputs (signal estimates) during adaptation and use the beam outputs in the iterative adjustment of the weights. The use of the beam output in the iterative algorithm is called performance feedback. Algorithms of this type are called "closed-loop" adaptive processing techniques. Estimation of the cross-power spectral matrices does not involve performance feedback and therefore techniques which compute the inverse of estimated cross-power spectral matrices are called "open-loop" adaptive processing techniques. (There are open-loop techniques which

7

attempt to find the beamforming weights directly, and closed-loop techniques have been proposed for finding the inverse of the cross-power spectral matrices.) A single-beam-closed-loop adaptive array processor system is shown in Figure 6.
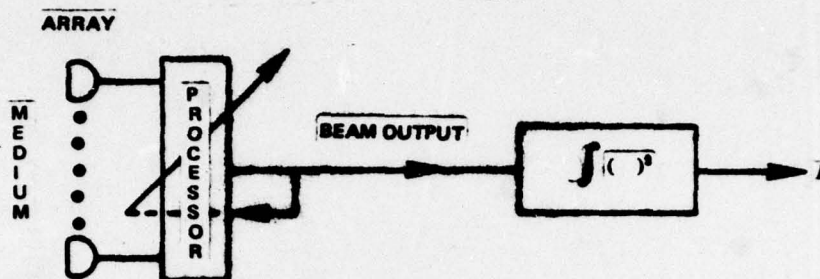

Figure 6. A Closed-Loop Adaptive Array Processor

Adaptive Detection Theory

The adaptive detection problem assumes some knowledge about the signal but none about the noise. The idea of an adaptive array processor is that the system somehow measures or "learns" enough about the noise field to keep the response low in the directions of loud noises. A stochastic process with certain temporal characteristics is called a signal if it impinges on the array from a conical spatial angle called the "main beam," but the same temporal process is called a noise if it comes from outside the cone—see Figure 7. Other temporal processes are regarded as noise regardless of their directional properties. It would be desirable to design a processor which could detect signals anywhere inside the main beam but reject any noise, including anything outside the main beam. If the solid angle of directions from which it is possible for the signal to arrive is larger than one main beam, then adjacent main beams may be used to complete the coverage. The array designer has some control over the width of the main beam, but the aperture of the array, the number of sensors, and the cost and complexity of the processor place limitations on the beamwidth. Although it is desired that the processor reject noise outside the main beam, for most noise fields there will remain some nonzero probability that the noise outside the main beam will be mistaken for signal inside the main beam (a false alarm).
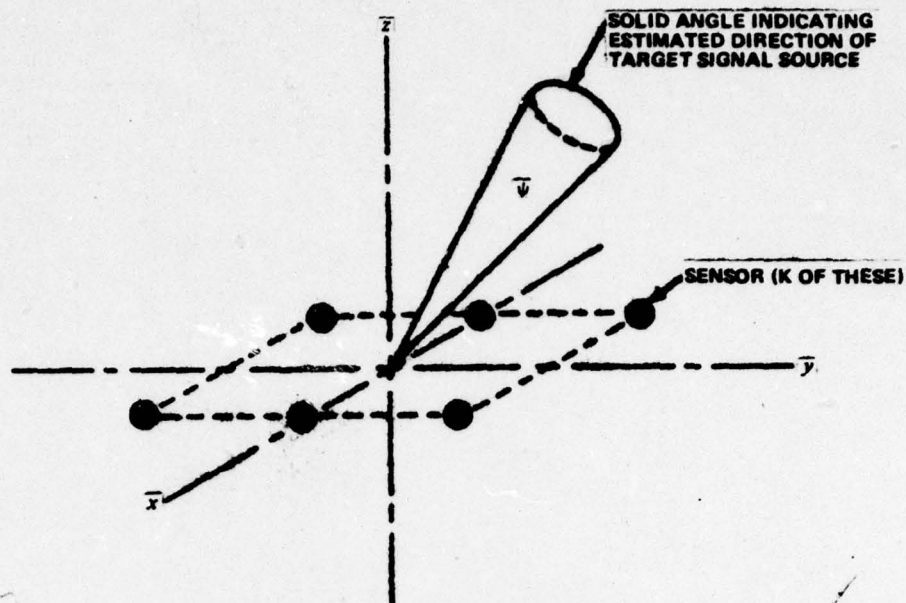
8

Figure 7. Array and Desired Main Beam

The heuristic criterion for an adaptive array detector is that the system detect signals in the main beam but reject any noise, including anything outside the main beam, without prior knowledge of the noise statistics. This criterion is substantially different from the criterion of the classical statistical detection theory described earlier. Classical statistical detection theory requires the knowledge of statistics which describe the temporal and spatial character of the noise field, as well as the signal field. The adaptive spatial detection criterion allows a description of the statistics of the signal field but not of the noise field.* An optimum solution with this criterion is not known. The threshold detection feature which is prescribed by statistical detection theory is sometimes used in adaptive detection. However, there is a substantial difference between an adaptive detector and an optimum detector in the classical sense. Classical statistical detection theory recognizes that the detector output (which is to be compared with the threshold) will fluctuate, but the only admissible reasons for the fluctuations are the stochastic nature of the input signal and noise field or a fundamental

---

*Complete statistical knowledge of the signal is not required. In some proposed applications, only the shape of the wavefront is assumed to be known.

9

change in the signal field. Because a fundamental change in the noise field is not allowed in classical detection theory, signal-to-noise ratio is a sufficient measure of the performance of a detector which is optimum in the classical sense. The following paragraphs show that the signal-to-noise ratio (or array gain) alone is not a sufficient measure of the performance of an adaptive detector.

If the noise field changes very slowly, or is essentially stationary, while the signal field changes abruptly, the signal can be detected if the output of the adaptive estimator is squared and integrated for a short time. Then, the adaptive processor does not change significantly during the detection interval, and the system performs in essentially the same way as a fixed linear estimator followed by a square-law detector. However, there are important applications to passive sonar by which the signal field changes slowly. When the rate of change in the signal field is comparable to the rate of change in the noise field, the adaptive processor will change as the signal changes. In this case the adaptive processor is a nonlinear system.

The adaptive detector, like the classical detector, forms a power estimate P for the signal in the beam. As in the case of the fixed processor, the adaptive detector depends on changes in the probability density function (*pdf*) of P to indicate the presence of a signal. However, the classical detection problem assumes that the input statistics due to noise never changes so that any shift in the *pdf* of P must be caused by the signal. The adaptive detection problem, on the other hand, allows changes in the noise statistics. Figures 8 and 9 show two possible probability density functions for the output power of a processor due to noise alone. If the threshold T is determined to yield a certain maximum false alarm probability under noise condition No. 1, the processor will have a significantly higher false alarm probability under noise condition No. 2. It is possible to find examples like Figures 8 and 9 with the same input noise power at the hydrophone. The change in the *pdf* is due to changes in the noise statistics, rather than a change in the input power.
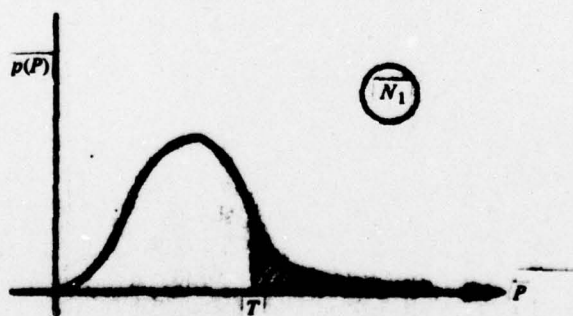

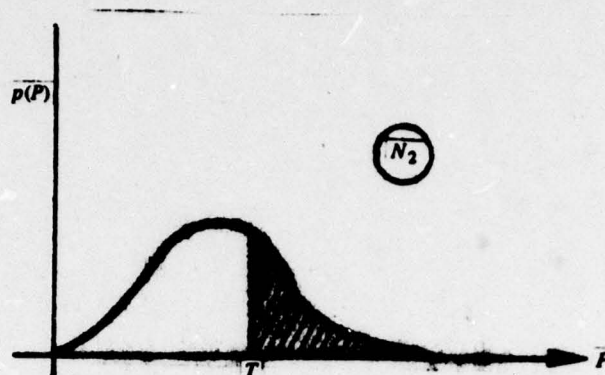Figure 8. Power *pdf* for Noise Condition No. 1

10

Figure 9. Power *pdf* for Noise Condition No. 2

Suppose the processor is designed to maximize array gain under noise condition No. 1. Then, the threshold T is set for a certain maximum false alarm rate under noise condition No. 1, as shown in Figure 8. The probability of detection is that area to the right of T illustrated under the curve shown in Figure 10. If the threshold remains fixed, the probability of false alarm increases as the noise field changes to condition No. 2. This can happen with an adaptive processor which is constantly maximizing array gain. Since it is assumed that the signal changes amplitude at approximately the same rate as the change in noise statistics, it is very difficult to adjust the threshold to maintain a fixed false alarm rate. Therefore, in addition to high array gain, the adaptive processor should minimize changes in the *pdf* of the output power that are due to changes in the noise statistics.

Figure 10. Power *pdf* for Signal Plus Noise No. 1

11

Some array systems and processors give larger *pdf* changes of output power than others. An array-processor system that presents an output power estimate P which is statistically insensitive to changes in the input noise statistics is called a *robust* system. An array-processor system that changes, as shown in Figure 11, is said to be *less robust* than the one shown in Figure 12. The design of the array and the design of the adaptive processor are both important for robustness. It is possible, in principle, to obtain a very robust adaptive array-processor system which obtains near-optimum array gain under slowly-varying noise statistics. However, some of the adaptive processors which have been considered obtain high array gain but are not robust. Thus, array gain alone is not a sufficient measure of the performance of an adaptive array-processing system.
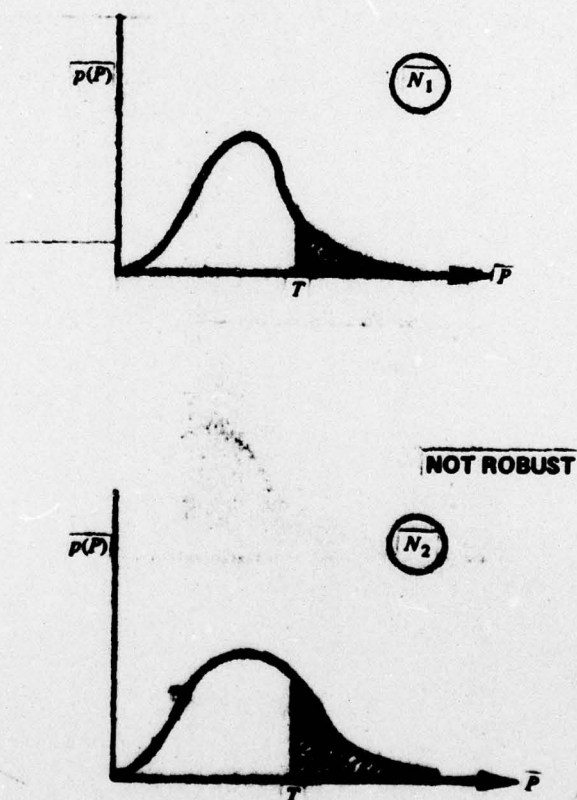
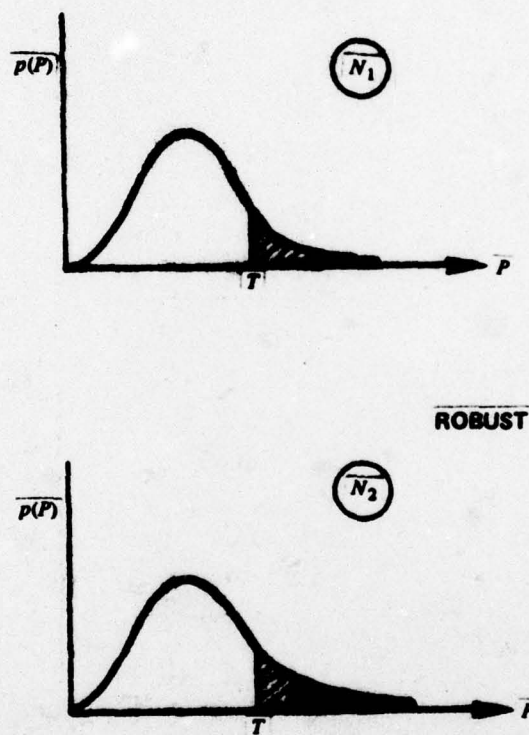Figure 11. Effect on *pdf* of Output Power Due to Changes in Noise Statistics

Figure 12. Effect on *pdf* of Output Power Due to Changes in the Noise Statistics

13

4. TERMINOLOGY

This paper has shown a clear distinction between the terms "optimum array processing," "adaptive processing," and "adaptive beamforming," although people on the periphery of the field sometimes confuse the first two terms. Since there is no known optimum adaptive array processor, the distinction is important. Optimum array processing assumes a complete knowledge of the statistical distributions of the noise field and the signal field. Adaptive processing is used when our statistical knowledge is incomplete.

The terms "adaptive array processing" and "adaptive beamforming" are often used interchangeably. Since "adaptive array processing" is a mouthful of syllables, "adaptive processing" is sometimes used as an abbreviation. Some vagueness results from this contraction, since adaptive processing can refer to the processing of a single channel as well as an array. Beamforming was the old way of filtering in space. "Adaptive beamforming" is a natural term to describe adaptive filtering in space. However, what is found adaptively may be some intermediate parameter, rather than the beam output. Furthermore, some adaptive array processing schemes do not form beams! "Adaptive array processing" is the more accurate term for describing the subject of this paper, but "adaptive beamforming" is the more popular term.

5. FURTHER RESEARCH AREAS

Some arrays seem to be more amenable to adaptive processing than others. At the present time the choice of array geometry is an art, not a science. A theory of optimum array geometry would be quite useful to the sonar engineer.

Clustering of hydrophones is another topic which merits study. Optimum array gain may increase as the number of hydrophones increase, but a larger number of time samples is also required to obtain the potential improvement. Thus, the speed of adaptation is inversely related to the number of input channels of the adaptive processor. Adaptive processing of conventionally steered clusters may be preferable to adaptive processing of the individual hydrophones.

The purpose of a passive multibeam array processing system is to map the noise field and focus attention on sources which might be signals (e.g., submarines). If the adaptive processing is narrowband, the number of beam-frequency combinations can be in the thousands. Operator attention can only be focused on a few of the beams. The purpose of adaptation is to reduce the

14

to the output power which results from noise outside the main beam. If the output of a beam is low, there is no need to adapt it because attention will not be focused on it. Only those beams whose output power crosses some threshold will need to be adapted. Therefore in light of the main purpose of a passive multibeam array processing system, it is recommended that a system be designed which will adapt the loudest beams. When a quiet beam becomes loud, an automatic monitor system would start adapting it; and when a loud beam becomes quiet, the monitor system would remove it from the group of adaptive beams. The operator should have the ability to designate a certain number of beams as adaptive. A statistical evaluation of such a system should be undertaken.

Attention should be focused on the cost of adaptive processing. Some adaptive processing systems are approximately equivalent in statistical performance, but differ considerably in complexity and cost. Some systems may perform a little better but cost much more than other systems. These considerations should be quantified whenever possible.

6. CONCLUSIONS

This paper has attempted to explain classical detection theory and adaptive detection theory without mathematics. Detection is the first thing to be done in sonar. There is also a classical estimation theory and corresponding adaptive approaches. No further discussion is possible at this point because there is a connection between detection and estimation which can only be explained mathematically. However, all of the statements in this paper can be proved mathematically.* This paper has been written for the busy man and the vaguely curious. It is hoped that those with the time and mathematical background will be motivated to sample the depth and diversity of the literature in the field.

---

*Most are proven in the doctoral dissertation by Burwell Brothers Goode, *Adaptive Sensor Array Processing*, NUC TP 259. The reference list of that report cites papers which complete the mathematical proofs that are only outlined in the dissertation.

15

REFERENCES

1. F. Bryn, "Optimum Signal Processing of Three-Dimensional Arrays Operating on Gaussian Signals and Noise," J. Acoust. Soc. Am., **34**, 289–297 (March 1962).
2. W. Vanderkulk, "Optimum Processing for Acoustic Arrays," J. Brit. IRE, **26**, 285–292 (October 1963).
3. D. Edelblute, J. Shapard, and G. Kinnison, "Array Gain Results from a 12-Hydrophone Experiment," Proc. of the Optimal/Adaptive Space Processing Seminar, NAVSHIPS 0967-368-4010, CONFIDENTIAL (12–14 August 1969, CONFIDENTIAL).
4. J. Sharpard et al., "Interim Report – 200 Channel Adaptive Beamforming Analysis of Passive At-Sea Data from BQS-6B Sonar Array on USS Sturgeon," NUC TN 392 (July 1970, CONFIDENTIAL).
5. R. Kneipfer and R. Hilt (Title classified), NUSC/NL Tech. Memo No. 2242-0032-71 (20 January 1971, SECRET-NOFORN).

INITIAL DISTRIBUTION