AD-A068 794 COLUMBIA UNIV DOBBS FERRY N Y HUDSON LABS F/G 17/1

CRITERIA FOR AN OPTIMUM RECEIVER FOR USE WITH A TEMPORALLY UNST--ETC(U)

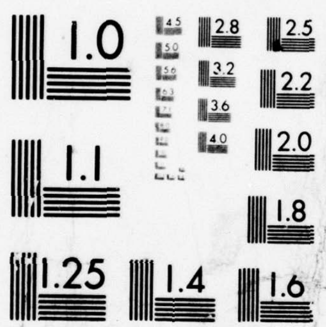
JAN 63 R E WILLIAMS NONR-266(84)

UNCLASSIFIED TM-64 NL

1 OF 1

AD
AO 68794

END
DATE
FILMED
6--79
DDC




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

# DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALIT
PRACTICABLE. THE COPY FURNISH
TO DDC CONTAINED A SIGNIFICAN
NUMBER OF PAGES WHICH DO NO
REPRODUCE LEGIBLY.

Hudson Laboratories
of
Columbia University
Dobbs Ferry, New York

Robert A. Frosch
Director

# CRITERIA FOR AN OPTIMUM RECEIVER FOR USE WITH A TEMPORALLY UNSTABLE MEDIUM.

by

Ross E. Williams

Technical Memorandum No. 64

14 Jan 63

25p.

July 24, 1962

Revised January 14, 1963

TM-64



This work was supported by the Office of Naval Research under Contract Nonr-266(84). Reproduction in whole or in part is permitted for any purpose of the United States government.

The problem of receiving intelligence through a medium whose properties are time varying has become an important one recently as communication techniques are extended to the upper atmosphere, or troposphere, and to the underwater domain of submarines. Both the troposphere and sea water are time varying in a number of ways, and the time variations have a wide range of characteristic periods. Some of these are long, such as the diurnal and seasonal variations of underwater sound velocity profiles. Others are quite short and may result from magnetic storms and solar flares, which affect the ion density of the troposphere, or the wave motion of the seawater surface and the relative motion of different underlying layers of water. This paper will be concerned with such short-time temporal instabilities, namely, those that have an opportunity to distort the time base of a transmitted or received signal (waveform) during the time duration of the coded signal. Such distortions, even though slight, can be expected to play havoc with the usual type of matched filter or correlator receiver,[1] which is actually an optimum receiver when temporal distortions are not present. Moreover, as signal codes of longer duration are used in an attempt to put more energy into the medium and attain more processing gain in the receiver, the liklihood of temporal distortions occurring within the time duration of the signal code becomes more pronounced. It will be shown that the optimum receiver in the presence of temporal distortions is a particular form of an adaptive receiver.

The communication problem treated here is considered in the general sense; it includes the transmission of a predetermined waveform coded in any manner, and the reception of a temporally distorted version of this waveform plus additive white Gaussian noise. Thus, it pertains to message transmissions for purposes of sending intelligence from one point to another as well as to radar and sonar systems designed principally for detection. Actually the detection aspect of an optimally encoded waveform

---

[1] P. M. Woodward, Probability and Information Theory, with Applications to Radar (Pergamon Press, Oxford, 1953), Chap. 4.

will be emphasized in this paper; no attempt will be made to treat the choice of codes or the best utilization of channel capacities in a Shannon sense.

A second important cause of temporal incoherence, in addition to the vagaries of the medium, becomes apparent when we consider target detection by reflection of a long duration encoded signal. A target, to be worthy of its mission, obviously has the capability of changing its location in time and of accelerating, and most targets can do both during the time they are illuminated by a radar or sonar transmitter. The effect of such target motion is to multiply the time base of the reflected waveform by a constant scale factor, in the case of uniform relative motion, or by a variable scale factor, in the case of accelerated motion. The former effect produces the Doppler-distorted return signal for which satisfactory techniques have been developed. The latter effect appears as an unwanted temporal distortion of the signal-a "rubberizing" of the time base which is quite analogous to that imposed by the time variations, e. g., ocean surface motion, of the medium.

An assumption will be made that signals arriving at the receiver via several different paths in the medium are resolvable even though overlapping. In other words, the various multipath arrivals are separated in time so that the output of a correlation receiver is a set of resolved correlation peaks corresponding to the various multipaths. Part of a proper encoding procedure is the choice of waveform which possesses a good ambiguity function and hence good time and Doppler resolution,[2] and therefore good multipath resolution. Moreover, the physical nature of the medium is often such, particularly in sound propagation through sea water, that the various arrivals are separated in time by appreciable fractions of a second, more than enough to resolve an optimally coded waveform of wide bandwidth.

The well-known optimum detector for a fully coherent signal corrupted by additive Gaussian noise is one which cross correlates the received

---

[2] Woodward, op. cit. (above, note 1), Chap. 7.

signal with a pre-recorded version of the transmitted signal. This result can be derived from a liklihood function as Woodward and others[3] have shown. As a next step to increasing the maximum range of detection, the transmitted waveform is lengthened (in the case of a peak-power limited transmitter) so as to couple more energy into the medium and to increase the frequency, or Doppler velocity, resolution for a moving target. However, in the presence of the temporal instabilities already discussed, extending the time duration of the signal becomes self-defeating after a point. Despite the increase in input signal-to-noise ratio obtained thereby, the greater the frequency resolution (resolution spacing = 1/T) possessed by a waveform by reason of its time duration T the more susceptible is its autocorrelation function to slight distortions of the time base. Thus, it is quite possible to "overdesign" a detection system by inadvertently building so much resolution into it that is fails to recognize a slightly distorted, target reflected signal in spite of an adequately high input signal-to-noise ratio. In fact, such an "overdesigned" system can easily fail to detect signals which would have been recognized unambiguously by simpler, less costly receivers which processed the received energy coherently over a shorter time span. Thus, it is obvious that an optimum detector should be matched somehow to the "coherence times" of the medium and any moving targets which may be present. It is less obvious exactly what form an optimum receiver, so matched, should take, although this problem has been treated in the radar case for a received signal distorted by multiplicative noise.[4,5] For certain sonar applications particularly, it is more pleasing and specific to treat the exact case of a time base distortion. The following paragraphs derive the form of an optimum receiver from the standpoint of a maximum liklihood detector when time base distortions are present.

We will consider a received signal $f(t)$ which is comprised of additive white Gaussian noise $n(t)$ and a distorted version $s[k_o t - \tau_o - \tau(t)]$

---

[3] C. W. Helstrom, Statistical Theory of Signal Detection (Pergamon Press, Oxford, 1960), p. 64, 73.

[4] R. Rojas, private communication.

[5] P. Bello, "Joint estimation of delay, Doppler, and Doppler rate," IRE Trans., P.G.I.T., IT-6, 330 (1960).

of the transmitted signal $s(t)$ :

$$f(t) = n(t) + rs[k_o t - \tau_o - \tau(t)] \quad . \tag{1}$$

The delay parameter $\tau_o$ represents the round-trip time delay for propagation out to the target and back, averaged over the signal duration. The term $\tau(t)$ introduces the time base distortion, which in itself is a function of time. The $k_o$ term, a constant, accounts for the time base compression or expansion resulting from the target's average speed relative to the transmitter-receiver location. The quantity $r$ is a real constant which represents attenuation of the transmitted signal. The $r$ might more generally have been considered complex and time varying, in which case its phase would represent a phase shift of the signal carrier and the time fluctuation could account for scintillation effects. While such considerations are applicable to radar, it is the contention here that phase shifts of the carrier do not adequately represent the usual time base distortions in underwater signal propagation.

Only statistical knowledge is available to the observer for $n(t)$ and $\tau(t)$ in Eq. (1). Their forms as explicit functions of t are not known, but they can be described in terms of the probability distributions of their amplitudes:

$$W_o(n) = \frac{1}{\sqrt{2\pi}\,\sigma_n} e^{-\frac{n^2}{2\sigma_n^2}} , \tag{2}$$

and

$$W_o(\tau) = \frac{1}{\sqrt{2\pi}\,\sigma_\tau} e^{-\frac{\tau^2}{2\sigma_\tau^2}} \tag{3}$$

for any time $t$ .

The probability distribution upon $\tau$ is assumed to be Gaussian not only because of its simplicity and its frequent close approximation to actual probability distributions in nature, but because by the central limit theorem the actual distribution will tend to a Gaussian shape if a number of independent and random factors contribute to the $\tau(t)$ distortion at any given time. In tropospheric and underwater propagation over long distances this is the case, for the resultant time base distortion is a cascading of many spatially separated distorting effects. For example, a principal mode for transmission of underwater sound over long distances is the RSR mode (refraction-surface reflection) which entails multiple surface or near surface reflections before arrival at the receiver. The sea wave amplitudes and phases at the spatially separated surface reflections can be considered independent, and reflection from each of the moving surfaces acts sequentially as a time base distorter upon the propagated signal. Thus, the resultant distortion can be thought of as an operator which is the product of a number of other independent operators cascaded to produce the resultant effect.

Equations (2) and (3) give the probability distributions at any time $t$, but the joint probability distributions for all times included in the signal duration are actually of interest. For this it is convenient to use Shannon's concept of number of "degrees of freedom" in a waveform. Shannon showed that a waveform of duration $T$ containing frequencies with a bandwidth $W$ has $2WT$ degrees of freedom. In other words, any such waveform can be fully specified by a set of real time samples $\frac{1}{2}W$ sec apart taken over the time span $T$, or by a set of complex frequency components spaced $1/T$ cps apart summed over the bandwidth $W$. Since the samples are statistically independent, the joint probability distribution for any arbitrary noise sample can be written as a multivariate normal distribution without cross-correlation terms:

$$W_o[n(t)] = \left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right)^{2WT} e^{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{2WT} n_i^2\left(\frac{i}{2W}\right)}, \qquad (4)$$

or from Eq. (1)

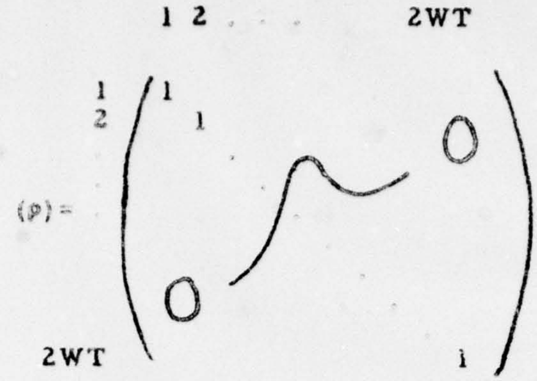
$$W_o[n(t)] = \left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right)^{2WT} e^{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{2WT}[f_i - rs_i(k_o, \tau_o, \tau_i)]^2} . \quad (5)$$

Equation (5) can be rewritten more compactly by considering the 2WT independent components $\tau_i$ to form a vector in a signal space of 2WT orthogonal dimensions. Then Eq. (5) becomes

$$W_o(\vec{n}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right)^{2WT} e^{-\frac{1}{2\sigma_n^2}[\vec{f} - r\vec{s}(k_o, \tau_o, \vec{\tau})]^2} , \quad (6)$$

where the arrow above a quantity indicates a vector. For convenience, we will describe $W_o(\tau)$, Eq. (3), in the same space of 2WT dimensions even though the number of independent components in the vector $\tau$ is $2W_\tau T$, where $W_\tau$ is the bandwidth of frequencies involved in $\tau(t)$. Since $\tau(t)$ arises in the first place because of physical motion of the propagating medium and the target, its frequencies must be band limited to a low pass region descriptive of these phenomena. Hence, it is reasonable to assume that $W_\tau \ll W$, the signal bandwidth. Then a description of $\tau$ in a space of 2WT dimensions will require cross-correlation coefficients between the various $\tau_i$ components which are no longer independent. In particular, if the $\tau_i$ are specified as time samples of $\tau(t)$ at $\frac{1}{2}W$ intervals, then the normalized correlation matrix[6] for $\tau$ will be of the form

[6] W. B. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill Book Company, Inc., New York, 1958), p. 151-52.




where the diagonal elements are all one, and, as shown by the Gaussian shaped curve at the center of the matrix, the slightly off-diagonal elements will be less than unity, falling to zero as their separation from the main diagonal increases. This form of the correlation matrix shows that closely spaced sample points $\tau_i$ are highly correlated, with the correlation dropping to zero as the time between sample points increases. The shape of this decay depends upon the time constants of the physical causes of the temporal distortions; in the case of a sonar system, it would depend upon the sea state and acceleration capabilities of the target. Finally, the variance factor $\sigma_\tau^2$ is unchanged whether $W_0(\tau)$ is described in a space of $2W_\tau T$ or $2WT$ dimensions. The variance is indicative of the range of amplitude excursions in $\tau(t)$ and also depends upon sea conditions and target acceleration capability, but it is independent of the shape of the frequency spectrum of $\tau(t)$. Rewriting Eq. (3) as a multivariate normal distribution in a space of $2WT$ dimensions, we take account of the cross correlation of the $\tau_i$ terms as follows:[6]

$$W_0(\vec{\tau}) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_\tau}\right)^{2WT} \frac{1}{\sqrt{|\rho|}}\, e^{-\frac{\vec{\tau}\,(\rho)^c\,\vec{\tau}}{2\sigma_\tau^2\,|\rho|}}, \qquad (7)$$

where

$(\rho)^c$ = cofactor of the correlation matrix, and

$|\rho|$ = determinant of the correlation matrix .

To define an optimum receiver we follow Woodward's concepts of a maximum liklihood detector.[1] Certainly the best that any receiver could provide would be the a posteriori probability distribution $W(\vec{\tau}, k_o, \tau_o | \vec{f})$ of $\vec{\tau}$, $k_o$ and $\tau_o$, given the received signal $\vec{f}$, where

$$W(\vec{\tau}, k_o, \tau_o | \vec{f}) = c\, W_o(\vec{\tau}, k_o, \tau_o)\, L(\vec{f} | \vec{\tau}, k_o, \tau_o)$$

$$= c\, W_o(k_o, \tau_o)\, W_o(\vec{\tau})\, L(\vec{f} | \vec{\tau}, k_o, \tau_o) \quad . \qquad (8)$$

No further information could be obtained or desired from the received signal $\vec{f}$.

In Eq. (8) c is a constant and $W_o(\vec{\tau}, k_o, \tau_o)$ is the a priori probability density distributions of $\vec{\tau}$, $k_o$ and $\tau_o$ which can be factored according to the lower expression in Eq. (8) into two independent terms, $W_o(k_o, \tau_o)$, a function of target motion (position and relative speed), and $W_o(\vec{\tau})$, which includes the perturbing effects of the medium and acceleration effects of the target in the time T. $L(\vec{f} | \vec{\tau}, k_o, \tau_o)$ is the liklihood function, which is proportional to the conditional probability $W(\vec{f} | \vec{\tau}, k_o, \tau_o)$ considered as a function of $\vec{\tau}$, $k_o$ and $\tau_o$. The objective of an optimum receiver is to evaluate Eq. (8). Short of this, which is an impractical task for most practical hardware systems, a maximum liklihood receiver can be designed to recognize the peak in the $W(\vec{\tau}, k_o, \tau_o | \vec{f})$ curve plotted as a function of $\vec{\tau}$, $k_o$ and $\tau_o$. If the curve is symmetric about its peak, that triad of values $(\vec{\tau}, k_o, \tau_o)$ corresponding to the peak will be identical to the triad of expectation values $(\langle\vec{\tau}\rangle, \langle k_o\rangle, \langle\tau_o\rangle)$, and the maximum liklihood receiver will provide these expectation values as its output. A further important assumption is usually made that the a priori probability $W_o(\vec{\tau}, k_o, \tau_o)$ is slowly varying as compared with the more sharply peaked $L(\vec{f} | \vec{\tau}, k_o, \tau_o)$, or $W(\vec{f} | \vec{\tau}, k_o, \tau_o)$. This assumption is valid when time base distortions, and hence the $\vec{\tau}$ factor, do not exist.

but in the present case it is exactly the fact that $W_o(\vec{\tau})$ has an influence upon the location of the peak of $W(\vec{\tau}, k_o, \tau_o | \vec{f})$ which cannot be ignored that leads to the conclusion that an optimum receiver is an adaptive receiver rather than an ordinary coherent correlator. It is still true that $W_o(k_o, \tau_o)$, the a priori probability on location and speed of the target, is slowly varying over that range of $k_o$ and $\tau_o$ for which the remainder of the right side of Eq. (8) varies significantly, and this term will be ignored in establishing the peak of Eq. (8).

We must, therefore, find the peak of

$$W_o(\vec{\tau})\, L(\vec{f} \,|\, \vec{\tau}, k_o, \tau_o) \quad \text{or} \quad W_o(\vec{\tau})\, W(\vec{f} \,|\, \vec{\tau}, k_o, \tau_o)$$

in which

$$W_o(\vec{\tau})\, W(\vec{f} \,|\, \vec{\tau}, k_o, \tau_o) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_\tau}\right)^{2WT} \frac{1}{\sqrt{|\rho|}}\, e^{-\frac{1}{2\sigma_\tau^2} \frac{\vec{\tau}(\rho)^C \vec{\tau}}{|\rho|}} \times$$

$$\times \left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right)^{2WT} e^{-\frac{1}{2\sigma_n^2}[\vec{f} - r\vec{s}\,(k_o, \tau_o, \vec{\tau})]^2} , \qquad (9)$$

where we have used Eq. (7) for $W_o(\vec{\tau})$ and Eq. (6) for $W(\vec{f} \,|\, \vec{\tau}, k_o, \tau_o)$, since $W(\vec{f} \,|\, \vec{\tau}, k_o, \tau_o) = W_o(\vec{n})$ by virtue of Eq. (1).

Obviously Eq. (9) attains peak value for maximum positive value of the exponent. Rewriting Eq. (9),

$$W_o(\vec{\tau})\, W(\vec{f}\,|\,\vec{\tau}, k_o, \tau_o) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_\tau}\right)^{2WT} \frac{1}{\sqrt{|\rho|}} \left(\frac{1}{\sqrt{2\pi}\,\sigma_n}\right)^{2WT} e^{-\frac{1}{2\sigma_\tau^2}\frac{\vec{\tau}(\rho)^C\vec{\tau}}{|\rho|} -}$$

$$- \frac{1}{2\sigma_n^2}\left[\vec{f}\cdot\vec{f} + r^2\vec{s}(k_o, \tau_o, \vec{\tau})\cdot\vec{s}(k_o, \tau_o, \vec{\tau}) - 2r\vec{f}\cdot\vec{s}(k_o, \tau_o, \vec{\tau})\right].$$

(10)

The terms $f \cdot f = \int f^2(t)\, dt$ and $r^2\vec{s}(k_o, \tau_o, \vec{\tau})\cdot\vec{s}(k_o, \tau_o, \vec{\tau}) =$

$$= r^2 \int [s(k_o t - \tau_o - \tau(t))]^2 dt$$

merely represent the total energy and the total signal (coherent) energy, respectively, available at the receiver input. These are functions of the average noise power $\sigma_n^2$, and the transmitter power and the attenuation coefficient $r^2$, and in neither case are they considered to be variables. Hence, we are concerned with maximizing the remaining part of the exponential.

$$\frac{r}{\sigma_n^2}\vec{f}\cdot\vec{s}(k_o, \tau_o, \vec{\tau}) - \frac{1}{2\sigma_\tau^2}\frac{\vec{\tau}(\rho)^C\vec{\tau}}{|\rho|} = \frac{r}{\sigma_n^2}\sum_{i=1}^{2WT} f_i\left(\frac{i}{2W}\right) s_i\left[k_o\left(\frac{i}{2W}\right) - \tau_o - \tau_i\left(\frac{i}{2W}\right)\right] -$$

$$- \frac{1}{2\sigma_\tau^2}\sum_{i,j=1}^{2WT}\frac{\tau_i \rho_{i,j}^C \tau_j}{|\rho|} \qquad (11)$$

The right side of Eq. (11) would be the ordinary correlation integral if all $\tau_i$ were set equal to zero and the second term were missing, corresponding to a complete lack of time base distortions. In the presence of time base distortions, it is necessary to maximize the sum of the two terms on the right side of Eq. (11) to define an optimum receiver. Obviously the first term, or correlation integral, should be made as large as possible consistent with a small value for the second term. A normal procedure in this direction would be to arrive at a proper choice for $\tau_o$ and $k_o$, the average delay and time base distortion factors, and then to optimize the correlation integral by adjusting the instantaneous distortion factors $\tau_i$ while at the same time keeping the second term suitably small. If the second term were not present, it would be possible to match exactly the $s_i$ to the $f_i$ in the correlation integral by adjusting the $\tau_i$, but reference to Eq. (1) would show that this procedure merely adapts the signal components $s_i$ to the noise $n_i$, which also has 2WT independent components. However, the presence of the second term in Eq. (11) prevents physically unlikely adjustments of the $\tau_i$. In particular, this term will become large, giving a small value to $W_o(\vec{\tau})$, Eq. (7), whenever the $\tau_i$ are varied so rapidly that the $\tau(t)$ bandwidth $W_\tau$ is exceeded. Thus, the correlation matrix $(\rho)$, which in the case of sonar is determined by sea conditions and target acceleration capability, will set restraints upon the choice of a 2WT set of components $\tau_i$ to maximize Eq. (11). In particular, the rapidity with which the $\tau_i$ are varied will be established by the correlation matrix $(\rho)$, and the amplitude of the $\tau_i$ variations will be indicated by the variance $\sigma_\tau^2$, which is also a function of sea state and target motion. When the average noise power $\sigma_n^2$ is small, the coefficient of the first term in Eq. (11) is large, and this term becomes relatively more important than the second. In this high signal-to-noise case it is, therefore, legitimate to allow the $\tau_i$ to assume even relatively unlikely sequences of values in the interest of maximizing the first term, even though the second term becomes more negative. This merely reflects the fact that when the noise power is low, there is a relatively high degree of confidence that whatever time base distortions occur in the signal are produced by the medium and not the additive noise, despite the lower a priori probability which may be associated with this type of medium distortion. Of course, the reverse is true when $\sigma_{\tau}^2$ is small, indicating relatively calm seas or the fact that targets which may be in the surveillance area have relatively small acceleration capability in the signal duration T, or when $\sigma_n^2$ is large, indicating the presence of a large noi contribution. In this case, the second term in Eq. (11) becomes more important and the tolerable departures of the $\tau_i$ from their zero mean are small. These considerations lead quite naturally to certain design forms for an adaptive receiver, which are discussed in Technical Memorandum No. 65.[7]

[7] Ross E. Williams, "The Design of Adaptive Receivers for Optimum Detection of Temporally Distorted Signals" (Hudson Laboratories, Columbia University Technical Memorandum No. 65, January 14, 1963).