COMMUNICATION SCIENCES

AND

ENGINEERING

X. PROCESSING AND TRANSMISSION OF INFORMATION*

Academic and Research Staff

Prof. P. Elias
Prof. R. G. Gallager
Prof. M. E. Hellman
Prof. E. V. Hoversten
Prof. R. S. Kennedy
Prof. C. E. Shannon
Dr. Jane W. S. Liu

Graduate Students

J. R. Clark
D. L. Cohn
J. G. Himes
J. E. Just
B. K. Levitt
G. Q. McDowell
P. F. Moulton
B. Neumann
R. S. Orr
J. L. Ramsey
J. H. Shapiro
T. A. Welch
H. P. H. Yuen

A. WAVEFORM ESTIMATORS FOR THE TURBULENT ATMOSPHERIC OPTICAL CHANNEL

1. Introduction

It is generally agreed[1] that an efficient optical receiver for the turbulent atmospheric channel must utilize the spatial diversity inherent in the distorted wavefront impinging on its aperture. By using a simplified "coherence-area" channel model,[2] such a receiver can be implemented by an array of detectors disposed over the aperture, with the outputs of the detectors processed individually and subsequently combined. An optimized multiple-area processor of this kind certainly performs at least as well as a similarly optimized single-area processor composed of one large detector; however, the latter has the clear-cut advantage of simplicity.

In this report we shall derive the optimum multiple-area processor and single-area processor for continuous demodulation of optical signals, using direct detection (no heterodyning). We utilize the "coherence-area" channel model of Kennedy and Hoversten,[2] assuming arbitrary slow fading and wideband Gaussian background noise. Maximum a posteriori probability (MAP) is the criterion for optimization. As an additional result, it is shown that if background noise is not present and the average energy of the received signal is sufficiently large, the optimum multiple-area processor is essentially equivalent to the optimum single-area processor. Similar results can be stated for more general channel models, with correlated fading; a detailed treatment will be given elsewhere.

2. Channel and Detector Models

Our channel model is that of Kennedy and Hoversten[2]; essentially, it describes the channel as a collection of independent (not necessarily Rayleigh), equal-strength diversity paths. The signalling duration T is assumed to be short compared with the

*This work was supported in part by the National Aeronautics and Space Administration (Grant NGL 22-009-013).

coherence time of the channel. Thus the fading factor for each path is an independent constant random variable. This factor, usually taken to be log-normal in amplitude and Gaussian in phase, will not be constrained to any particular distribution.

An array of D direct detectors, each of area A_c (assumed small relative to spatial variations of the received field), fills the receiving aperture. The complex envelope of the signal impinging on the j^{th} detector, integrated over the detector surface, is

$$y_j(t) = A_c z_j s[t, a(t)] + n_j(t), \qquad t \in [0, T]. \tag{1}$$

The transmitted field is linearly polarized, and n_j is the relevant polarization component of the background noise. $n_j(t)$ is assumed to be wideband Gaussian noise of finite average power. The channel model[2] specifies the background noise field to be essentially white in space, ensuring the independence of the n_j, $j = 1, \ldots, D$. z_j is a complex fading factor, identically distributed for all j. We presume, of course, that the amplitude and phase of the fading are coherent over areas of equal size – an assumption easily eliminated at the expense of simplicity.

We now make the essential assumption that the background noise is of small intensity per temporal mode. This guarantees that the output of the j^{th} direct detector is a Poisson process,[3] conditioned on z_j, with mean at time T,

$$\beta(s, s)\, v_j + \beta \overline{P}_n T, \tag{2}$$

where

$$\left.\begin{aligned} \beta &= A_c \eta / h\nu = \text{constant} \\ \eta &= \text{quantum efficiency} \\ h &= \text{Planck's constant} \\ \nu &= \text{mean carrier frequency} \\ v_j &= A_c |z_j|^2 \\ (a, b) &= \int_0^T a(t)\, b^*(t)\, dt \\ \overline{P}_n &= \text{total average noise power.} \end{aligned}\right\} \tag{3}$$

3. Optimum Demodulators

A modulator performs a no-memory operation on a message a(t) to produce the transmitted signal s(t, a(t)). The generalization to operations with memory is straightforward and will be discussed in this report. To estimate a(t), $t \in [0, T]$, we use as observables the number of "counts" $\{N_j\}_{j=1}^{D}$ recorded by each detector in $(0, T]$, and the times of occurrence of these counts $\{t_{ij}\}_{i=1}^{N_j}{}_{j=1}^{D}$. Now for the j^{th} detector, the joint conditional probability density function of the vector $\underline{t}_j = \{t_{ij}\}_{i=1}^{N_j}$ and N_j can be written[4-6]

$$p(\underline{t}_j, N_j | v_j) = \begin{cases} e^{-\beta[v_j(s,s)+\overline{P}_nT]} \beta^{N_j} \prod_{i=1}^{N_j} \left[v_j|s(t_{ij})|^2+\overline{P}_n\right], & N_j > 0 \\ e^{-\beta[v_j(s,s)+\overline{P}_nT]}, & N_j = 0 \end{cases} \tag{4}$$

This is, of course, implicitly conditioned on the modulation a(t). Averaging over the fading, we get

$$p(\underline{t}_j, N_j) = \int_0^\infty p(\underline{t}_j, N_j | v_j)\, p(v_j)\, dv_j. \tag{5}$$

Since the diversity paths are assumed to be independent, the joint probability density function of the vectors $\underline{t} = \{\underline{t}_j\}_{j=1}^{D}$ and $\underline{N} = \{N_j\}_{j=1}^{D}$ is just the product

$$p(\underline{t}, \underline{N}) = \prod_{j=1}^{D} p(\underline{t}_j, N_j). \tag{6}$$

Combining (4), (5), and (6), and taking logarithms, we find that

$$\ln p(\underline{t}, \underline{N}) = \sum_{j=1}^{D} \ln \left\{\beta^{N_j} e^{-\beta\overline{P}_nT} \int_0^\infty Q_j(v)\, p(v)\, dv\right\}, \tag{7}$$

where

$$Q_j(v) = e^{-\beta(s,s)v} \prod_{i=1}^{N_j} \left[v|s(t_{ij})|^2+\overline{P}_n\right]. \tag{8}$$

We have dropped the subscript on v because the fading factors z_j are assumed to be identically distributed for all j (equal strength diversity).

For a(t) a real, zero-mean, Gaussian random process, the MAP estimate $\hat{a}_k$ of its k^{th} Karhunen-Loève coefficient satisfies the equation[7]

$$\hat{a}_k = \lambda_k \left. \frac{\partial \ln p(\underline{t}, \underline{N})}{\partial a_k} \right|_{a(t)=\hat{a}(t)}, \tag{9}$$

where λ_k is the variance of a_k. The derivative of the expression in Eq. 7 is found to be[7]

$$\frac{\partial \ln p(\underline{t}, \underline{N})}{\partial a_k} = \sum_{j=1}^{D} \sum_{i=1}^{N_j} A_{ij} \frac{\partial |s(t_{ij})|^2}{\partial a(t_{ij})} \phi_k(t_{ij}) - B \frac{\partial(s,s)}{\partial a_k}, \tag{10}$$

where ϕ_k is the k^{th} eigenfunction in the expansion of a(t), and

$$A_{ij} = \frac{\int_0^\infty \frac{v}{v|s(t_{ij})|^2 + \overline{P}_n} Q_j(v)\, p(v)\, dv}{\int_0^\infty Q_j(v)\, p(v)\, dv} \tag{11}$$

$$B = \beta \sum_{j=1}^{D} \frac{\int_0^\infty vQ_j(v)\, p(v)\, dv}{\int_0^\infty Q_j(v)\, p(v)\, dv}. \tag{12}$$

Defining the MAP interval estimate of a(t) as $\hat{a}(t) = \sum_k \hat{a}_k \phi_k(t)$, and also defining $\frac{\partial(s,s)}{\partial a(t)} = \sum_k \frac{\partial(s,s)}{\partial a_k} \phi_k(t)$, we get for the estimator equation, with $K_a(t,u) = E(a(t)a(u))$,

$$\hat{a}(t) = \sum_{j=1}^{D} \sum_{i=1}^{N_j} K_a(t, t_{ij}) A_{ij} \frac{\partial |s(t_{ij})|^2}{\partial \hat{a}(t_{ij})} - B \int_0^T K_a(t,\tau) \frac{\partial(s,s)}{\partial \hat{a}(\tau)} d\tau. \tag{13}$$

It is clear that a nontrivial estimator results only when some sort of energy modulation is used; this is consistent with the fact that we have disallowed heterodyning. Note also that the first term in (13) is essentially a digital filter, and the second, an analog filter, with the data entering the second term only through the coefficient B. If a modulation system with memory is used; that is, if the input to a no-memory modulator is the output $x(t) = \int_0^T h(t,u)\, a(u)\, du$ of a linear filter, then it can be shown that (13) still applies, with $K_a(t,u)$ replaced by $h_a(t,u) = \int_0^T h(t,y)\, K_a(u,y)\, dy$ and the derivatives taken with respect to $\hat{x}(t) = \int_0^T h(t,u)\, \hat{a}(u)\, du$.[7] Thus for the case that we are studying, we see that linear filtering and MAP interval estimation commute — a familiar result in the estimation of a deterministically modulated Gaussian message in Gaussian noise, where the received data are $r(t) = s(t, a(t)) + n(t)$.[7]

Instead of many small detectors, we can constrain our receiver to use one large detector, covering many coherence areas. The suboptimum MAP receiver that results is equivalent to summing directly the individual detector outputs of the multiple-area processor, before subsequent processing (incoherent combining). The estimator equation remains the same in form; however, the fading variables v_j appear only in the sum

$$x = \sum_{j=1}^{D} v_j.$$

We denote this new estimate $\hat{a}_1(t)$; it satisfies

$$\hat{a}_1(t) = \sum_{i=1}^{N} K_a(t, t_i)\, {}_1A_i \frac{\partial |s(t_i)|^2}{\partial \hat{a}_1(t_i)} - {}_1B \int_0^T K_a(t, \tau) \frac{\partial (s, s)}{\partial \hat{a}_1(\tau)} \, d\tau, \tag{14}$$

where $\{t_i\}_{i=1}^{N}$ is the set of occurrence times of the counts (note that this set is simply a re-indexed version of the set $\underline{t}$ introduced before), and

$$N = \text{number of counts in } (0, T] = \sum_{j=1}^{D} N_j \tag{15}$$

$${}_1A_i = \frac{\int_0^\infty \frac{x}{x|s(t_i)|^2 + D\overline{P}_n} Q(x)\, p(x)\, dx}{\int_0^\infty Q(x)\, p(x)\, dx} \tag{16}$$

$${}_1B = \beta \frac{\int_0^\infty xQ(x)\, p(x)\, dx}{\int_0^\infty Q(x)\, p(x)\, dx} \tag{17}$$

$$Q(x) = e^{-\beta(s, s)x} \prod_{i=1}^{N} \left[x|s(t_i)|^2 + D\overline{P}_n\right]. \tag{18}$$

The probability density p(x) is given by the D-order convolution of identical probability density functions

$$p(x) = \mathop{*}_{j=1}^{D} p(v_j) = \mathop{*}_{j=1}^{D} p(v). \tag{19}$$

Equations 13 and 14 represent the optimum multiple and single-area demodulators for the channel and detector assumptions that we have made. Their applicability is limited, though, by the awkward forms of the coefficients A_{ij}, B, ${}_1A_i$, and ${}_1B$. Fortunately, revealing simplifications can be made. It can be shown that the coefficients are directly related to conditional least-squares estimates of the fading intensity. Specifically, $B = \beta \sum_1^D \hat{v}_j$, where $\hat{v}_j$ is the least-squares estimate of v_j, given the data $\underline{t}_j$ and N_j. On the other hand, apart from a factor, A_{ij} is the least-squares estimate of v_j, given all of the data from the j^{th} detector <u>except</u> t_{ij}. Similar statements can be made about ${}_1A_i$ and ${}_1B$.

Further simplifications can be made by expanding the product in $Q_j(v)$ in a sum

$$Q_j(v) = e^{-\beta(s,s)v} \prod_{i=1}^{N_j} \left[|s(t_{ij})|^2 v + \overline{P}_n\right]$$

$$= e^{-\beta(s,s)v} \sum_{\ell=0}^{N_j} c_\ell(\underline{t}_j, N_j)\, \overline{P}_n^{N_j-\ell}\, v^\ell. \tag{20}$$

The coefficient $c_\ell(\underline{t}_j, N_j)$ is the sum of $\binom{N_j}{\ell}$ different ℓ-order products of factors $|s(t_{ij})|^2$, with i not repeated in a given term of the sum. For example, $c_o = 1$, $c_1 = \sum_{i=1}^{N_j} |s(t_{ij})|^2$, $c_2 = |s(t_{1j})|^2\, |s(t_{2j})|^2 + |s(t_{1j})|^2\, |s(t_{3j})|^2 + \ldots + |s(t_{2j})|^2\, |s(t_{3j})|^2 + |s(t_{2j})|^2\, |s(t_{4j})|^2 + \ldots$, etc. Using (20), the integrals in A_{ij} and B can be expressed in terms of derivatives of the moment-generating function of v, which in turn are related to the counting distribution (conditioned on the message) associated with the fading signal in the <u>absence</u> of background noise.[3] Equation 18 can be expanded in a similar manner to simplify ${}_1A_i$ and ${}_1B$.

The conditional counting distribution is a function of the fading parameters and $\beta(s,s)$ only, and can be evaluated exactly in certain cases.[3] For complex Gaussian (Rayleigh) fading, the distributions are Bose-Einstein or negative binomial; for complex Gaussian (Rician) fading, the distributions are Laguerre; for log-normal fading, the distributions can be approximated by Laguerre distributions.[8] Thus the demodulator structure can be specified exactly in many cases of importance.

The simplified coefficients contain weighted sums of "signal" and "noise" terms, as reflected in (20) — in essence, an algebraic separation of the effects of the fading and the effects of the noise. This result is a manifestation of the fact that the Poisson process at the output of a given photodetector is the sum of a "fading signal" process and an <u>independent</u> "noise" process. In the general case of arbitrary Gaussian background noise, the detector outputs are not Poisson, and the separation into independent signal and noise components is not possible.[3] A joint probability density function paralleling Eq. 4 exists,[5] but appears to be inaccessible analytically, except for trivial modifications of the pure Poisson case.

4. Equivalence of Multiple- and Single-Area Processors

If there is no background noise, we can make a significant additional simplification. A_{ij} becomes

$$A_{ij} = \frac{1}{|s(t_{ij})|^2}, \tag{21}$$

and it can be shown that B is given by

$$B = \frac{1}{(s,s)} \sum_{j=1}^{D} (N_j+1) \frac{p(N_j+1)}{p(N_j)}, \tag{22}$$

where p(k) is the conditional counting distribution discussed above. If all the N_j are consistently large (that is, if the average energy of the fading signal is sufficiently large), the ratio of probabilities ≈ 1, and

$$B \approx \frac{N+D}{(s,s)}. \tag{23}$$

Equations 21 and 23, together with Eq. 13, reveal that the optimum multiple-area processor is equivalent to a single-area processor, under the conditions stated. The second term in (13) does not use data from individual photodetectors, and the double indexing in the first term, along with the double summation, can be changed to single indexing and a single summation. $N = \Sigma N_j$ terms are evaluated at the N occurrence times and summed, without regard for the detector in which the data originated.

It should be mentioned that even if the $\{N_j\}$ are not large enough to justify (23), the data need be identified as to origin only in the time-independent coefficient B. Also, we see from (22) that the multiple-area processor is precisely equivalent to a single area processor, regardless of the magnitudes of the $\{N_j\}$, whenever the expression $(N_j+1)\,p(N_j+1)/p(N_j)$ is linear in N_j. This is the case for zero-mean Gaussian (Rayleigh) fading because, as we have mentioned, p(k) is Bose-Einstein (geometric).

J. R. Clark

References

1. E. V. Hoversten, "The Atmosphere as an Optical Communication Channel," 1967 IEEE International Convention Record, Vol. 15, Pt. 11, pp. 137-145.

2. R. S. Kennedy and E. V. Hoversten, "On the Atmosphere as an Optical Communication Channel," IEEE Trans., Vol. IT-14, pp. 716-725, September 1968.

3. S. Karp and J. R. Clark, "Photon Counting: A Problem in Classical Noise Theory," Technical Report, NASA Electronics Research Center, Cambridge, Mass. (to be published in 1970); also, submitted for publication to IEEE Trans. on Information Theory.

4. I. Bar-David, "Communication under the Poisson Regime," IEEE Trans., Vol. IT-15, pp. 31-37, January 1969.

5. R. L. Stratonovich, Topics in the Theory of Random Noise (Gordon and Breach, Science Publishers, New York, 1963), Vol. 1, Chap. 6.

6. E. Parzen, Stochastic Processes (Holden-Day, Inc., San Francisco, 1962).

7. H. L. Van Trees, Detection, Estimation, and Modulation Theory (McGraw-Hill Book Co., New York, 1968), Vol. 1, Chap. 5.

8. S. Karp and J. R. Clark, "Approximations for Lognormally Fading Optical Signals" (submitted to Proc. IEEE).