Exact Maximum Likelihood Time Delay Estimation for Short Observation Intervals

Benoit Champagne, *Member, IEEE,* Moshe Eizenman, *Member, IEEE,* and Subbarayan Pasupathy, *Fellow, IEEE*

Abstract—**This paper presents an exact solution to the problem of maximum likelihood time delay estimation for a Gaussian source signal observed at two different locations in the presence of additive, spatially uncorrelated Gaussian white noise. The solution is valid for arbitrarily small observation intervals; that is, the assumption $T \gg \tau_c$, $|d|$ made in the derivation of the conventional "asymptotic" maximum likelihood (AML) time delay estimator, where τ_c is the correlation time of the various random processes involved and d is the differential time delay, is relaxed. The resulting "exact" maximum likelihood (EML) instrumentation is shown to consist of a finite-time delay-and-sum beamformer, followed by a quadratic postprocessor based on the eigenvalues and eigenfunctions of a one-dimensional integral equation with nonconstant weight. The solution of this integral equation is obtained for the case of stationary signals with rational power spectral densities. Finally, the performance of the EML and AML estimators are compared by means of computer simulations for a first-order autoregressive source signal and for values of T, τ_c, and d such that the condition $T \gg \tau_c$, $|d|$ is not satisfied. The results indicate that the AML estimator suffers a dramatic deterioration in performance (large error, bias, and standard deviation) as the ratio d/T increases from 0 to 0.1, making it essentially useless beyond this point. No such effect is observed with the EML estimator which has the best overall performance and whose mean-square error approaches the Cramér–Rao lower bound for large values of the signal-to-noise ratio (SNR). The results also indicate that for large SNR, considerable simplifications of the EML estimator are possible without any significant loss in performance.**

I. Introduction

During the last two decades, considerable attention has been given to the problem of time delay estimation (TDE), which consists of estimating the time differences between noisy, delayed replicas of a common random source signal [1]–[5]. This problem arises most naturally in the context of passive sonar where an array of sensors is used to monitor a propagating wave. In this case, the time delays between the various sensor output signals are equal to the time differences of arrival of the wavefront to the sensors. Estimation of these delays can therefore provide information about the direction of arrival of the wavefront. Besides this dominant application, TDE also occurs in other areas of science such as seismology, economics, and dendrochronology (historical dating from tree ring records) [6], to name a few.

One of the most popular methods of estimation used in connection with the TDE problem has been that of maximum likelihood (ML), largely because ML estimators are known to be asymptotically efficient in the limit $T \to \infty$, where T denotes the observation interval [7, p. 138]. The application of the ML method to the TDE problem is usually based on three fundamental assumptions [2]–[4], namely, constant delay, stationary processes, and long observation interval (i.e., $T \gg \tau_c$, $|d|$, where τ_c is the correlation time of the various random processes involved and d is the differential time delay.) Because they are derived under the assumption of a long observation interval, we refer to these ML estimators of time delay as "asymptotic" maximum likelihood (AML) estimators.

In many cases of practical interest, however, the assumption of long observation interval is inconsistent with other prevailing conditions. For example, when monitoring a rapidly moving source with a passive array of sensors, the assumptions of stationary processes and constant delay will only be satisfied over a limited time interval, and failure to take this fact into account may result in a serious deterioration of the delay estimator performance [8]. Another example occurs when the source signal under observation is a transient of short duration.[1] In this case, increasing the observation interval beyond the signal duration will eventually result in a performance deterioration rather than in an improvement. Finally, in other applications, external factors will prevent the observation interval from being sufficiently long. This may occur, for example, when working with prerecorded data sets.

While the above considerations justify the need for new TDE techniques that can be used when the condition $T \gg \tau_c$, $|d|$ is not satisfied, it is legitimate to wonder about the role played by the ML estimator in this "nonasymptotic" situation. There are, however, very important motivations for considering ML estimation even when the observation interval is short. First, as indicated in [10], there is actually a tradeoff between observation time and signal-to-noise ratio (SNR) in the TDE problem. Under appropriate conditions, one might therefore expect the ML estimator to be asymptotically optimal in the limit of large

Manuscript received September 2, 1989; revised June 4, 1990.

B. Champagne is with INRS-Télécommunications, Université du Québec, Verdun, Que., Canada H3E 1H6.

M. Eizenman and S. Pasupathy are with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada M5S 1A4.

IEEE Log Number 9143798.

[1] A discussion of the equivalence of transient and narrow-band signals in the context of signal processing can be found in [9].

1053-587X/91/0600-1245$01.00 © 1991 IEEE

SNR. Second, knowledge of the ML estimator (or, more exactly, of the associated log-likelihood instrumentation) can be used to evaluate various bounds on estimator performance, such as the Cramér–Rao or Barankin lower bounds [11, ch. 2]. Finally, the ML estimator provides fundamental insights that are essential in the development of simpler estimators, some of which follow as limiting cases of the ML estimator itself.

Despite its importance, the problem of ML TDE for short observation intervals has received relatively little attention in the signal processing literature, mostly because of its inherent complexity. Indeed, when the assumption of long observation interval is not satisfied, the problem becomes nonstationary in nature because of "end effects" and it is no longer possible to use the conventional Fourier techniques that previously made the analysis of the AML estimator tractable. Among the results available for nonstationary ML TDE, those of Stuller [12] are of particular interest here because they address specific issues related to end effects. Lourtie and Moura [13] also present general results on nonstationary ML TDE that are applicable in the case of short observation intervals. However, these studies formulate the ML estimator in terms of unsolved integral or partial differential equations and, as a result, they do not provide practical ways of realizing the ML estimator, nor do they compare its performance to that of the conventional AML estimator.

In this paper, we present an exact solution to the problem of ML TDE for a Gaussian source signal observed at two separated locations in the presence of additive, spatially uncorrelated Gaussian white noise. This solution is valid for arbitrarily short observation intervals; that is, the standard assumption $T \gg \tau_c$, $|d|$ made in the derivation of the AML estimator is relaxed. The exact maximum likelihood (EML) estimator is obtained in two steps: first, the factorization properties of optimum space–time processors [14] are used to transform the matrix integral equations defining the log-likelihood function into simpler scalar integral equations; second, these reduced integral equations are solved for the case of stationary signals with rational power spectral densities by generalizing a technique originally developed by Youla for a simpler class of integral equations [15]. In the last part of the paper, computer simulations are used to study the comparative performance of the EML, AML, and other related estimators. For the simulations, a first-order autoregressive source signal is used along with values of T, τ_c, and d violating the condition $T \gg \tau_c$, $|d|$.

II. Problem Formulation

We consider the family of vector random processes $\boldsymbol{x}(t; d)$, parametrized by the delay variable $d \in [d_-, d_+]$, and defined as follows:

$$\boldsymbol{x}(t; d) = \boldsymbol{a}(t; d) + \boldsymbol{n}(t), \qquad 0 \le t \le T \tag{1}$$

$$\boldsymbol{a}(t; d) = \begin{bmatrix} a(t) \\ a(t-d) \end{bmatrix}, \qquad \boldsymbol{n}(t) = \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \tag{2}$$

where $a(t)$, the source signal, and $n_i(t)$ $(i = 1, 2)$, the additive noise components, are zero mean uncorrelated stationary Gaussian random processes with known autocorrelation functions $R_a(\tau)$ and $R_n(\tau)$, respectively. The interval $[d_-, d_+]$ represents the *a priori* range of possible delay values. We assume that $a(t)$ possesses a power spectral density $G_a(\omega)$ which is a rational function of ω^2. That is,

$$R_a(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_a(\omega) e^{j\omega\tau}\, d\omega \tag{3}$$

$$G_a(\omega) = \frac{N(s^2)}{D(s^2)}, \qquad s = j\omega \tag{4}$$

where $N(\cdot)$ and $D(\cdot)$ are irreducible real coefficient polynomials of degrees m and n, respectively, satisfying the following conditions: a) any zero of $N(s^2)$ (interpreted as a function of the complex variable s) on the imaginary axis has even multiplicity; b) $D(s^2)$ has no zero on the imaginary axis; and c) $n > m$. These conditions ensure that $G_a(\omega)$ is a well-defined, absolutely integrable spectral density [15]. We further assume that the additive noise components $n_i(t)$ are white with unit spectral height, that is,

$$R_n(\tau) = \delta(\tau) \tag{5}$$

where $\delta(\tau)$ is the Dirac delta function.

Let $\boldsymbol{x}(t)$ $(0 \le t \le T)$ be a particular realization (i.e., an observation) of the process $\boldsymbol{x}(t; d^*)$, where d^* represents the true value of the unknown delay parameter. The TDE problem then consists of finding a statistically reliable estimate of d^* from the observation $\boldsymbol{x}(t)$. By definition, the ML estimate $\hat{d}_{\mathrm{ML}}$ of d^* is the value of d at which the log-likelihood function of the observed data, $\ln \Lambda(\boldsymbol{x}; d)$, defined below, attains an absolute maximum. That is,

$$\ln \Lambda(\boldsymbol{x}; \hat{d}_{\mathrm{ML}}) \ge \ln \Lambda(\boldsymbol{x}; d), \qquad \text{all } d \in [d_-, d_+]. \tag{6}$$

The log-likelihood function of the observed data $\boldsymbol{x}(t)$ $(0 \le t \le T)$, represented simply by $\boldsymbol{x}$, can be obtained by means of the following series representation [16, ch. 6]:

$$\ln \Lambda(\boldsymbol{x}; d) = \frac{1}{2} \{ l_1(\boldsymbol{x}; d) - l_2(d) \} \tag{7}$$

$$l_1(\boldsymbol{x}; d) = \sum_{i=1}^{\infty} \frac{\lambda_i}{1 + \lambda_i} \left\{ \int_0^T \Phi_i^{\mathrm{tr}}(t)\, \boldsymbol{x}(t)\, dt \right\}^2 \tag{8}$$

$$l_2(d) = \sum_{i=1}^{\infty} \ln (1 + \lambda_i). \tag{9}$$

The superscript $^{\mathrm{tr}}$ in (8) denotes transposition. λ_i and $\Phi_i(t)$, which also occur in the Karhunen–Loève expansion of the vector process $\boldsymbol{a}(t; d)$ in (2), are the eigenvalues and (real) normalized vector eigenfunctions, respectively, associated with the autocorrelation matrix $\boldsymbol{R}_a(\tau; d)$ of $\boldsymbol{a}(t; d)$. They satisfy the following integral equations:

$$\int_0^T \boldsymbol{R}_a(t - u; d)\, \Phi_i(u)\, du = \lambda_i \Phi_i(t), \qquad 0 \le t \le T \tag{10}$$

$$\int_0^T \Phi_i^{tr}(t)\, \Phi_j(t)\, dt = \delta_{ij} \tag{11}$$

where δ_{ij} is the Kronecker delta. Note that λ_i and $\Phi_i(t)$ are actually functions of the delay parameter d.

In order to evaluate the log-likelihood function $\ln \Lambda(\boldsymbol{x}; d)$ and thus determine the ML estimate $\hat{d}_{ML}$, it is first necessary to solve the integral equations (10) and (11) for the λ_i and the $\Phi_i(t)$. Using a standard technique [11, p. 205], it is rather easy to show that under the asymptotic condition $T >> \tau_c$, $|d|$, where τ_c is the correlation time (or inverse bandwidth) of the process $a(t)$ in (2), possible solutions to (10) and (11) are given by ($i = 0, \pm 1, \cdots$)

$$\lambda_i = 2G_a(\omega_i), \qquad \omega_i = \frac{2\pi i}{T} \tag{12}$$

$$\Phi_i(t) = (2T)^{-1/2} e^{j\omega_i t} \begin{bmatrix} 1 \\ e^{-j\omega_i d} \end{bmatrix}. \tag{13}$$

Here, the eigenfunctions are complex valued and the superscript ^{tr} in (11) must be interpreted as a complex transpose operation. Real, properly normalized eigenfunctions can be obtained from (13) by considering the real and imaginary parts individually and multiplying them by an appropriate scaling factor. Substituting these real eigenfunctions together with their eigenvalues λ_i (12) into (8) and (9), it can be shown that the log-likelihood function (7) reduces to the conventional AML estimator, which consists of a generalized cross correlator with Hannan–Thomson prefilters [4].

However, when the observation interval is short, i.e., when the condition $T >> \tau_c$, $|d|$ is not satisfied, λ_i and $\Phi_i(t)$ specified by (12), (13) are no longer solutions of (10), (11) and, as we will see in Section V, their use in (8), (9) may result in unreliable delay estimates. In the next two sections, we present an explicit method that can be used to find the exact solutions of (10) and (11) for any value of T, regardless of how small it is.

III. Dimensionality Reduction

The first step in the solution of the matrix integral equations (10) and (11) is to transform them into equivalent scalar integral equations. This can be achieved by applying one of the factorization properties of optimum space-time processors presented in [14]. The property of interest [14, property 1′] essentially states the following: since the vector process $\boldsymbol{a}(t; d)$ in (2) is related to the scalar process $a(t)$ by a linear operation, the eigenfunctions $\Phi_i(t)$ in (10) can be obtained by applying the same linear operation to the properly normalized eigenfunctions of a scalar integral equation whose eigenvalues are precisely the λ_i. There are at least two important reasons for proceeding this way rather than considering (10) and (11) directly. First, solving a scalar integral equation is conceptually simpler than solving a matrix integral equation. Second, this approach provides simple, yet very important structural information about the eigenfunctions $\Phi_i(t)$ and the log-likelihood processor (7)–(9).

In the case $0 \le d \le T$, [14, property 1′] gives the following results: Define the weighting function

$$\rho(t) = \begin{cases} 1, & -d < t < 0 \\ 2, & 0 < t < T - d \\ 1, & T - d < t < T \end{cases} \tag{14}$$

and let λ_i and $\psi_i(t)$ be the eigenvalues and eigenfunctions of the scalar integral equation

$$\int_{-d}^{T} R_a(t - u)\, \psi_i(u)\, \rho(u)\, du = \lambda_i \psi_i(t), \qquad -d \le t \le T \tag{15}$$

with the $\psi_i(t)$ satisfying the orthonormality condition

$$\int_{-d}^{T} \psi_i(t)\, \psi_j(t)\, \rho(t)\, dt = \delta_{ij}. \tag{16}$$

Then, the functions

$$\Phi_i(t) = \begin{bmatrix} \psi_i(t) \\ \psi_i(t - d) \end{bmatrix}, \qquad 0 \le t \le T \tag{17}$$

are normalized eigenfunctions of $\boldsymbol{R}_a(\tau; d)$ with eigenvalues λ_i. That is, λ_i and $\Phi_i(t)$ obtained through (14)–(17) satisfy (10) and (11) (this can be verified by direct substitution). Of course, $\boldsymbol{R}_a(\tau; d)$ will have other eigenfunctions that cannot be generated in this way. However, as explained in [14], these eigenfunctions all have zero-eigenvalue and therefore, in light of (8) and (9), they need not be considered when evaluating the log-likelihood function (7).

Once we have determined the λ_i and $\Phi_i(t)$ for a given value of d in $[0, T]$, it is a simple matter to obtain the λ_i and $\Phi_i(t)$ corresponding to $-d$. Indeed, using (10) and (11), it can be verified easily that (the dependence upon d has been introduced temporarily)

$$\lambda_i(-d) = \lambda_i(d) \tag{18}$$

$$\Phi_i(t; -d) = \begin{bmatrix} \psi_i(t - d; d) \\ \psi_i(t; d) \end{bmatrix} \tag{19}$$

where $\lambda_i(d)$ and $\psi_i(t; d)$ are obtained through (14)–(16). Reference [14, property 1′] could also be used in a similar way to handle the case $|d| > T$. However, since no additional difficulties are involved and since $|d| > T$ does not represent a very practical situation (no overlap between the source signal components observed at the two sensors), we shall limit ourselves to the case $|d| \le T$. The problem of solving the reduced integral equation (15) is addressed in the next section.

We now look at the processor configuration that results from the specific structure of the eigenfunctions $\Phi_i(t)$ in (17). Substituting (17) in (8), the following expression can be obtained for the data-dependent term $l_1(\boldsymbol{x}; d)$ of the log-likelihood function (7):

$$l_1(\boldsymbol{x}; d) = \sum_{i=1}^{\infty} \frac{\lambda_i}{1 + \lambda_i} \left\{ \int_{-d}^{T} \psi_i(t)\, y(t)\, dt \right\}^2 \tag{20}$$

where the scalar process $y(t)$ is given in terms of the components of the observed vector process $\boldsymbol{x}(t) = [x_1(t), x_2(t)]^{tr}$ by

$$y(t) = \begin{cases} x_2(t+d), & -d < t < 0 \\ x_1(t) + x_2(t+d), & 0 < t < T-d \\ x_1(t), & T-d < t < T. \end{cases} \quad (21)$$

The resulting processor configuration is shown in Fig. 1. It consists of two specialized subprocessors. The first one, referred to as a finite-time beamformer, transforms the observed vector process $\boldsymbol{x}(t)$ $(0 \le t \le T)$ into the scalar process $y(t)$ $(-d \le t \le T)$ by means of (21). The second one, referred to as a quadratic postprocessor, computes $l_1(\boldsymbol{x}; d)$ from $y(t)$ by means of (20) and finally performs bias compensation and scaling to generate the log-likelihood function (7).

Many interesting observations can be made regarding the finite-time beamforming operation (21). First, its output $y(t)$ will in general be discontinuous at $t = 0$ and $t = T - d$. These discontinuities, or "end effects," are a direct consequence of the fact that the observation interval $[0, T]$ is finite. Second, the middle term on the right-hand side of (21), whose contribution to the integral in (20) is most important when $T >> d$, corresponds to a simple coherent delay-and-sum beamforming operation. Finally, the operation (21) is robust in the sense that it is independent of the signal statistics. In other words, the finite-time beamformer in Fig. 1 is entirely determined by the geometry of the problem and is insensitive to modeling errors in the source signal $a(t)$.

As seen from (7), (9), and (20), the quadratic postprocessor has the same structure as a conventional log-likelihood processor based on the eigenvalues λ_i and scalar eigenfunctions $\psi_i(t)$, except that the eigenfunctions are now normalized according to (16) which features the discontinuous weighting function $\rho(t)$. In practice, because of the asymptotic behavior of the λ_i in the limit $i \to \infty$, only a finite number N of terms need to be included in (9) and (20). If the observation interval T is short and the spectrum $G_a(\omega)$ converges to zero sufficiently rapidly as ω tends to $\pm\infty$, this number can actually be very small, therefore resulting in computational simplifications. This was indeed our primary motivation for using a series expansion approach to the evaluation of the log-likelihood function (7). Other possible realizations of the quadratic postprocessors are indicated in [14].

Finally, it is of interest to study the behavior of the data-dependent term $l_1(\boldsymbol{x}; d)$ (20) in the limiting case of large signal power. To begin, we observe from (15) that multiplying $R_a(t)$ by a constant factor has the effect of rescaling all the eigenvalues λ_i by the same factor. Therefore, if the signal power $R_a(0)$ is sufficiently large, the dominant eigenvalues will be much larger than one and, for these eigenvalues, we will have $\lambda_i/(1 + \lambda_i) \approx 1$. Making this approximation in (20), we obtain

$$l_1(\boldsymbol{x}; d) = \sum_{i=1}^{\infty} \left[\int_{-d}^{T} \psi_i(t)\, y(t)\, dt \right]^2. \quad (22)$$

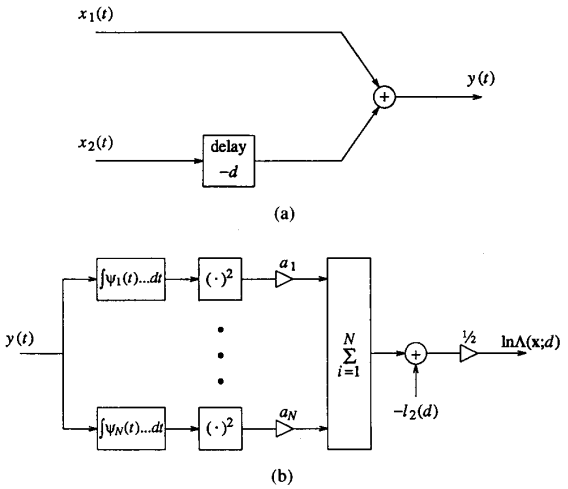


Fig. 1. Configuration of the EML processor. (a) Finite-time beamformer ($x_i(t) = 0$ for $t < 0$ and $t > T$, $i = 1, 2$). (b) Quadratic postprocessor ($a_i = \lambda_i/(1 + \lambda_i)$, $i = 1, \cdots, N$).

Equation (22) can be further simplified by means of the completeness relation [17, p. 266]

$$\rho(t) \sum_i \psi_i(t)\, \psi_i(u) = \delta(t - u) \quad (23)$$

which is satisfied by the eigenfunctions $\psi_i(t)$ in (15), (16). The result is

$$l_1(\boldsymbol{x}; d) = \int_{-d}^{T} \frac{y^2(t)}{\rho(t)}\, dt. \quad (24)$$

The right-hand side of (24) is a nonconventional measure of the total signal energy at the output of the finite-time beamformer. From a practical point of view, (24) presents two major advantages over (20). First, it does not require *a priori* knowledge of the signal statistics, i.e., it is independent of $R_a(\tau)$. This is particularly important for applications in which such knowledge is unavailable.[2] Second, because it does not require the computation of the eigenfunctions $\psi_i(t)$, its implementation is considerably simpler than that of (20). It remains to be determined whether or not (24) can actually be used as a reasonable approximation to (20), and if so, under what conditions. This will be investigated experimentally in Section V.

IV. Solution of the Reduced Integral Equation

Many techniques are available for the determination of the eigenvalues and eigenfunctions of a stationary process $a(t)$ whose autocorrelation function $R_a(\tau)$ satisfies (3), (4) (see [11, p. 187] for references). Unfortunately, none of them can be applied directly to solve the reduced integral equation (15) because of the nonconstant weighting function $\rho(u)$ appearing under the integral sign. In this section, we develop a new algorithm that can be used to solve

[2]We note, however, that only $l_1(\boldsymbol{x}; d)$ has been simplified to a form independent of $R_a(\tau)$. The complete determination of the log-likelihood function (7) still requires the knowledge of the bias term $l_2(d)$ (9), which depends in a specific way on the eigenvalues of $R_a(\tau)$.

(15) when $R_a(\tau)$ is of the type (3), (4) and $\rho(t)$ is given by (14).

We begin by stating three properties that are satisfied by the eigenvalues λ_i and eigenfunctions $\psi_i(t)$ of the integral equation (15). These properties will later help in clarifying some aspects of the proposed solution algorithm for (15). The proofs of these properties can be found in Appendix A.

Property 1: The eigenvalues λ_i are bounded. More precisely, $0 < \lambda_i < 2\gamma$, where

$$\gamma = \max \{G_a(\omega): -\infty < \omega < \infty\}. \tag{25}$$

Property 2: The eigenfunctions $\psi_i(t)$ are continuous on the interval $[-d, T]$.

Property 3: Let $L_2(\rho)$ denote the Hilbert space of square integrable functions defined over the interval $[-d, T]$, with scalar product given by

$$(f, g) = \int_{-d}^{T} f(t)\, g(t)\, \rho(t)\, dt, \qquad \text{all } f, g \in L_2(\rho). \tag{26}$$

There exists a complete orthonormal set of eigenfunctions $\{\psi_i(\cdot)\}$ in $L_2(\rho)$, such that each eigenfunction in the set satisfies the symmetry relation

$$\psi_i(t) = \epsilon_i \psi_i(T - d - t) \tag{27}$$

where ϵ_i, referred to as the symmetry index of $\psi_i(t)$, is either equal to $+1$ or -1.

The next theorem is at the basis of the solution algorithm for the reduced integral equation (15). Before stating it, however, we need to introduce some notation. Let $\psi_i(t)$ be a solution of (15) and define the auxiliary functions $\psi_{ij}(t)$ ($j = 1, 2, 3$; $-\infty < t < \infty$) as follows:

$$\psi_{ij}(t) = \begin{cases} \psi_i(t), & t \in I_j \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

where $I_1 = (-d, 0)$, $I_2 = (0, T - d)$ and $I_3 = (T - d, T)$. Denote by $\Psi_{ij}(s)$ the bilateral Laplace transforms (BLT) of the auxiliary functions $\psi_{ij}(t)$, that is,

$$\Psi_{ij}(s) = \int_{-\infty}^{\infty} \psi_{ij}(t) e^{-st}\, dt. \tag{29}$$

Finally, let $D^+(s)$ and $D^-(s)$ be real coefficient polynomials of degree n defined by the conditions:

$$D(s^2) = D^+(s)\, D^-(s) \tag{30}$$

$$D^+(s) = D^-(-s) \tag{31}$$

$$D^+(s) = 0 \quad \text{implies Re } (s) < 0 \tag{32}$$

where Re (s) denotes the real part of s. The factorization of $D(s^2)$ as in (30)–(32) is sometimes referred to as a canonical factorization [18, p. 41].

Theorem: a) Let λ_i and $\psi_i(t)$ be solutions of (15) and suppose that $\psi_i(t)$ satisfies (27). Then, there exist real coefficient polynomials $P_i(s)$ and $Q_i(s)$, of degrees equal at most to $n - 1$ and $2n - 1$, respectively, such that

$$\Psi_{i1}(s) = \frac{e^{sd} D^+(s)\, P_i(s) - Q_i(s)}{\lambda_i D(s^2) - N(s^2)} \tag{33}$$

$$\Psi_{i2}(s) = \frac{Q_i(s) + \epsilon_i e^{-s(T-d)} Q_i(-s)}{\lambda_i D(s^2) - 2N(s^2)} \tag{34}$$

all singularities in (33) and (34) being removable. Moreover,

$$\Psi_{i3}(s) = \epsilon_i e^{-s(T-d)} \Psi_{i1}(-s). \tag{35}$$

b) Conversely, suppose we can find polynomials $P_i(s)$ and $Q_i(s)$ as above, and numbers λ_i ($0 < \lambda_i < 2\gamma$) and ϵ_i ($\epsilon_i = \pm 1$) such that the right-hand sides of (33) and (34) have only removable singularities. Let $\psi_i(t) = \psi_{i1}(t) + \psi_{i2}(t) + \psi_{i3}(t)$, where $\psi_{ij}(t)$ are the inverse BLT of $\Psi_{ij}(s)$ given by (33)–(35), the region of convergence being the whole s plane. Then λ_i and $\psi_i(t)$ are solutions of (15) and $\psi_i(t)$ satisfies (27).

The proof of this theorem is given in Appendix B. Based on this theorem, it is now easy to develop an algorithm for the determination of the eigenvalues and eigenfunctions of the reduced integral equation (15). Essentially, we must determine the coefficients of the polynomials $P_i(s)$ and $Q_i(s)$ by requiring that the right-hand sides of (33) and (34) have only removable singularities. This will translate into a $3n \times 3n$ system of homogeneous linear equations which has nonzero solutions only for certain values of the λ_i, which are precisely the desired eigenvalues. We now describe in details the various steps of the solution algorithm for (15):

1) Determine the canonical factors $D^+(s)$ and $D^-(s)$ of $D(s^2)$ as in (30)–(32).

2) Find the K_1 distinct zeros $s_{1k}(\lambda)$, $k = 1, \cdots, K_1$, of the polynomial $\lambda D(s^2) - N(s^2)$ and let m_{1k} denote their respective multiplicities. Similarly, find the K_2 distinct zeros $s_{2k}(\lambda)$, $k = 1, \cdots, K_2$, of the polynomial $\lambda D(s^2) - 2N(s^2)$ and let m_{2k} denote their respective multiplicities. Observe that the zero configurations of both these polynomials are symmetric about the real and the imaginary axis in the s plane.

3) Define the functions

$$g_1(s) = e^{sd} D^+(s)\, P(s) - Q(s) \tag{36}$$

$$g_2(s, \epsilon) = Q(s) + \epsilon e^{-s(T-d)} Q(-s) \tag{37}$$

where $\epsilon = \pm 1$ and where $P(s)$ and $Q(s)$ are yet undetermined real coefficient polynomials of degrees $n - 1$ and $2n - 1$, respectively, i.e.,

$$P(s) = p_0 + p_1 s + \cdots + p_{n-1} s^{n-1} \tag{38}$$

$$Q(s) = q_0 + q_1 s + \cdots + q_{2n-1} s^{2n-1}. \tag{39}$$

4) Consider the following system of $4n$ complex homogeneous linear equations in the $3n$ real unknown coefficients $p_0, \cdots, p_{n-1}, q_0, \cdots, q_{2n-1}$:

$$g_1^{(l)}(s_{1k}(\lambda)) = 0, \qquad k = 1, \cdots, K_1, \quad l = 0, \cdots, m_{1k} - 1 \tag{40}$$

$$g_2^{(l)}(s_{2k}(\lambda), \epsilon) = 0, \qquad k = 1, \cdots, K_2, \quad l = 0, \cdots, m_{2k} - 1 \tag{41}$$

where the superscript (l) denotes the lth derivative with respect to s.

5) Using the symmetries present in the zero configurations $\{s_{1k}(\lambda)\}$ and $\{s_{2k}(\lambda)\}$ and in the functions $g_1(s)$ and $g_2(s, \epsilon)$, simplify this system into an equivalent one consisting of $3n$ real homogeneous linear equations in the unknowns $p_0, \cdots, q_{2n-1}$. Let this new system be represented by the matrix equation

$$A(\lambda, \epsilon)X = 0 \tag{42}$$

where $A(\lambda, \epsilon)$ is a $3n \times 3n$ real matrix and $X = [p_0, \cdots, q_{2n-1}]^{\text{tr}}$.

6) Find the roots (λ_i, ϵ_i), with $0 < \lambda_i < 2\gamma$ and $\epsilon_i = \pm 1$, of the equation

$$\det A(\lambda, \epsilon) = 0. \tag{43}$$

The λ_i so obtained are the eigenvalues of (15) and the ϵ_i are the symmetry indices of the corresponding eigenfunctions.

7) For each pair (λ_i, ϵ_i), determine the solution space of the equation

$$A(\lambda_i, \epsilon_i)X = 0. \tag{44}$$

For simplicity, assume that the solution space has dimension one (the general case needs only trivial modifications) and let $X_i = [p_{0,i}, \cdots, q_{2n-1,i}]^{\text{tr}}$ be an arbitrary nonzero element of the solution space.

8) Let

$$P_i(s) = p_{0,i} + \cdots + p_{n-1,i}s^{n-1}, \tag{45}$$

$$Q_i(s) = q_{0,i} + \cdots + q_{2n-1,i}s^{2n-1}. \tag{46}$$

Then, we have

$$\psi_i(t) = \begin{cases} c_i \sum_{k=1}^{K_1} \text{Res}\left[\dfrac{D^+(s)\,P_i(s)e^{s(t+d)}}{\lambda_i D(s^2) - N(s^2)}, s_{1k}(\lambda_i)\right], \\ \qquad -d < t < 0 \\ c_i \sum_{k=1}^{K_2} \text{Res}\left[\dfrac{Q_i(s)e^{st}}{\lambda_i D(s^2) - 2N(s^2)}, s_{2k}(\lambda_i)\right], \\ \qquad 0 < t < T - d \end{cases} \tag{47}$$

where c_i is a yet undetermined coefficient and where the notation Res $[f(s), s_0]$ is used to denote the residue of the function $f(s)$ at one of its pole s_0. For $T - d < t < T$, $\psi_i(t)$ can be obtained from (47) by making use of the symmetry relation (27).

9) In order to determine c_i, simply substitute the above expressions for $\psi_i(t)$ into (16) and perform the appropriate integration.

This completes the algorithm. Note that in general, steps (6) and (9) will have to be carried out numerically.

This algorithm generalizes Youla's technique [15] for the determination of the eigenvalues and eigenfunctions of $R_a(\tau)$ in the absence of the weighting function $\rho(t)$. Besides the various modifications needed to account for $\rho(t)$, there is one major distinction between our approach and that of Youla which deserves some explanation. What makes Youla's technique efficient is the use of a symmetry relation, equivalent to (27) for $d = 0$, that reduces by half the number of unknown eigenfunction coefficients. However, the derivation of this relation in [15] relies upon certain supplementary assumptions concerning the multiplicity of the eigenvalues. Rather than making supplementary assumptions, we have chosen a different approach: since Property 3 asserts the existence of eigenfunctions satisfying the symmetry relation (27), we developed our solution algorithm so that it gives precisely these eigenfunctions.

Some interesting observations can be made about the eigenfunctions $\psi_i(t)$. First, although they are continuous as stated by Property 2, (47) reveals that their behavior in the middle interval $0 < t < T - d$ is different from that in the end intervals $-d < t < 0$ and $T - d < t < T$. This is another consequence of the fact that the observation interval is finite. Second, the existence of symmetries in the eigenfunctions, as stated by Property 3, is a very desirable feature that can be exploited at the implementation level to reduce by half the computational load. Finally, as a consequence of (47), it follows that the auxiliary functions $\psi_{ij}(t)$ associated with $\psi_i(t)$ by means of (28) satisfy specific constant-coefficient linear homogeneous ordinary differential equations. More precisely, we have:

Property 4: Let λ_i and $\psi_i(t)$ be solutions of (15) and let $\psi_{ij}(t)$ be defined as in (28). Then

$$[\lambda_i D(\Delta^2) - \rho(t)\,N(\Delta^2)]\psi_{ij}(t) = 0, \qquad t \in I_j \tag{48}$$

where Δ represents the time derivative operator d/dt.

Practically, this property implies that the eigenfunctions $\psi_i(t)$ can be generated on-line by means of a recursive equation. In this case, the solution algorithm still needs to be used to determine the eigenvalues λ_i and the initial conditions necessary to start the recursion.

As a final remark, we mention that the above solution algorithm for (15) could easily be extended to a larger class of weighting functions consisting of all the positive step functions that are symmetrical about the middle point $(T - d)/2$ of the interval $[-d, T]$. This might actually find an application in the more general TDE problem with an arbitrary number $M \geq 2$ of equidistant sensors.

V. Computer Simulations

Computer simulations were used to study the comparative performance of the EML, AML, and other related time delay estimators when the asymptotic condition $T \gg \tau_c, |d|$ is not satisfied. Our primary concern was not to provide an exhaustive characterization of the performance of the various processors, but rather to determine what kind of improvement, if any, can be achieved by using EML instead of a more conventional method such as AML. Accordingly, we focused our attention on a simple low-pass system and only a few values of the parameters of interest, such as bandwidth, signal-to-noise ratio and true delay, were considered. Nevertheless, the results are quite significant and clearly show that substantial im-

provements in performance are possible with EML when the observation interval is short. In this section, we briefly describe the simulations and discuss the results obtained.

A. Simulation Details

The signal $a(t)$ was modeled as a stationary first-order Gauss–Markov process with autocorrelation function and power spectral density

$$R_a(\tau) = Pe^{-\alpha|\tau|} \tag{49}$$

$$G_a(\omega) = \frac{2\alpha P}{\omega^2 + \alpha^2} \tag{50}$$

respectively. P represents the mean-square value of the signal $a(t)$ while α is a measure of its bandwidth (-3 dB point of $G_a(\omega)$). In what follows, we refer to the dimensionless quantity αT as the time-bandwidth product.

In order to implement the simulations on a digital computer, the observation interval $[0, T]$ was divided into N_s subintervals of length $T_s = T/N_s$. The true value of the delay parameter was then chosen as a multiple of T_s, i.e., $d^* = k^* T_s$ where k^* is an integer. To generate the sample values of the sensor output signals $x_1(t)$ and $x_2(t)$ at the time instants $t_n = (n - 1/2)\, T_s$ $(n = 1, \cdots, N_s)$, we proceeded as follows. First the sequence $a(t_n)$ $(n = 1 - k^*, \cdots, N_s)$ was obtained by passing a Gaussian white noise sequence through an appropriate first order IIR linear filter. Two other Gaussian white noise sequences $e_1(t_n)$ and $e_2(t_n)$, with variance

$$E[e_i^2(t_n)] = \frac{1}{T_s} \tag{51}$$

were also generated (these correspond to samples of fictitious band limited white noise processes $e_1(t)$ and $e_2(t)$ with common power spectral density $G_e(\omega) = 1$ for $|\omega| < \pi/T_s$ and 0 otherwise). Finally, the three independent sequences $a(t_n)$, $e_1(t_n)$ and $e_2(t_n)$ were combined according to[3]

$$\begin{aligned} x_1(t_n) &= a(t_n) + e_1(t_n) \\ x_2(t_n) &= a(t_{n-k^*}) + e_2(t_n). \end{aligned} \tag{52}$$

From the above considerations, it follows that the signal-to-noise ratio (SNR) at the sensor outputs, which we simply define as the signal mean-square value divided by the noise mean-square value, is given by

$$\text{SNR} = PT_s. \tag{53}$$

Five different methods were used to generate time delay estimates from the simulated data $x_1(t_n)$ and $x_2(t_n)$. They are listed as follows together with their abbreviations: 1) exact maximum likelihood (EML); 2) high SNR version of EML (HEML); 3) asymptotic maximum likelihood (AML); 4) AML with prewindowing (WAML); 5) maximization of tapered cross-correlation estimate (MTCC).

For each method, a time delay estimate was first obtained by maximizing the corresponding likelihood function over the restricted set of delay values

$$d = kT_s, \qquad k \in \{k^* - 8, \cdots, k^* + 8\}. \tag{54}$$

This estimate was then refined by applying a three-point quadratic interpolation formula as in [19]. Important simulation details for each of the methods listed above as well as the motivation for Methods 4 and 5 are given in the next paragraphs.

The likelihood function used to generate the EML estimate was obtained via (7), (9), (20), and (21), with the infinite summations in (9) and (20) replaced by finite ones extending from $i = 1$ to a preset value $i = N$, and with the integral over t in (20) approximated by a Riemann sum with step size T_s. The eigenvalues $\lambda_i(d)$ and the sample values $\psi_i(t_n; d)$ of the eigenfunctions were precomputed using the algorithm of Section IV and stored along with the (truncated) bias term $l_2(d)$.

The HEML estimate was obtained by maximizing the likelihood function $l_1(\boldsymbol{x}; d)$ (24), without using the bias term $l_2(d)$ (9). This later modification was suggested by independent computations revealing the weak dependence of $l_2(d)$ on the parameter d. In light of (24), the HEML estimate is therefore completely independent of the signal statistics.

The AML estimate was obtained by maximizing the asymptotic log-likelihood function

$$L(\boldsymbol{x}; d) = \sum_{i=1}^{M} \frac{2G_a(\omega_i)}{2G_a(\omega_i) + 1} |X_i|^2, \qquad \omega_i = \frac{2\pi i}{T} \tag{55}$$

$$X_i = (2T)^{-1/2}[1, e^{j\omega_i d}] \int_0^T \boldsymbol{x}(t) e^{-j\omega_i t}\, dt \tag{56}$$

which can be obtained by the technique outlined in Section II. The integral in (56) was evaluated via radix-2 FFT. In order to make our comparison of EML and AML meaningful, we chose $M = N/2$ as the upper limit of summation in (55). This choice, based on experimental considerations, ensured that the corresponding instrumentations had comparable bandwidths.

The WAML method was similar to the AML method except that the data sequences $x_1(t_n)$ and $x_2(t_n)$ were multiplied by a Hanning window function $w(t_n) = \sin^2(\pi t_n/T)$ prior to computing the FFT in (56). This was done as an attempt to limit certain performance degradations observed with the AML estimator and caused by end effects (this is discussed further in Section V-B).

Finally, the MTCC estimate was obtained by maximizing the "tapered" cross-correlation estimate, which is given by [20]

$$\hat{R}_{12}(d) = \frac{1}{T - d} \int_0^{T-d} x_1(t)\, x_2(t + d)\, dt \tag{57}$$

[3]The problem addressed in this paper might as well have been formulated in the discrete-time domain. Indeed, as long as d^* is a multiple of T_s, the continuous-time derivations of Sections III and IV admit direct discrete-time counterparts. While the discrete-time approach seems more natural for simulations and DSP implementations, the continuous-time approach more accurately reflects the analog nature of the physical signals monitored at the sensor outputs. In any case, provided the sampling rate is sufficiently high, both approaches are equivalent.

for $0 \leq d \leq T$. Again, this method was considered because it can potentially eliminate problems caused by end effects.

For each choice of system parameters αT, N_s, SNR, d^*/T, and N considered, 500 independent simulations were run and at the end, the various statistical performance indicators of interest were computed. These were the percentage of anomalous time delay estimates and the bias and standard deviation of the nonanomalous estimates. Based on Ianniello [19], an estimate $\hat{d}$ was said to be anomalous if $|\hat{d} - d^*| > \tau_c/2$, where τ_c, the correlation time of the source signal $a(t)$, is defined here as

$$\tau_c = \frac{1}{R_a^2(0)} \int_{-\infty}^{\infty} R_a^2(\tau)\, d\tau = \frac{1}{\alpha}. \tag{58}$$

Conversely, $\hat{d}$ was said to be nonanomalous if $|\hat{d} - d^*| \leq \tau_c/2$. In order to simplify the discussion in Section V-B, we shall adopt the following notational convention: the symbols η, b, and σ, with one of the subscripts EML, HEML, AML, WAML, and MTCC, will be used to represent, respectively, the percentage of anomalous delay estimates and the bias and standard deviation of the nonanomalous estimates, for the corresponding method of estimation indicated by the subscript.

B. Results

The results presented below were obtained with the following values of the system parameters:

$$\alpha T = 4$$
$$N_s = 32 \quad (\text{i.e., } T_s = T/32)$$
$$d^*/T_s = 0, 2, 4, \cdots, 16$$
$$\text{SNR} = 0, 10, 20 \text{ dB}$$
$$N = 32.$$

The above range for d^*/T_s is quite sufficient to illustrate the practical problems that arise with conventional TDE techniques when the condition $T >> d^*$ is not satisfied. Observe that $d^*/T_s = 16$ corresponds to the case when the true delay is equal to half of the observation interval.

Tables I–III show the percentage of anomalous time delay estimates as a functions of d^*, for each of the five estimators considered and for SNR equals 0, 10, and 20 dB, respectively.[4] These tables reveal that the EML estimator achieves the lowest percentage of anomalous estimates for all values of SNR and d^* considered (except for SNR = 0 dB and $d^* = 2$, where η_{AML} is slightly lower than η_{EML}). It is also observed that for a fixed value of d^*, η_{EML} decreases as SNR increases, while for a fixed value of SNR, η_{EML} generally increases as d^* increases. Both these observations are intuitively very conceivable: in the first case, as SNR $\rightarrow \infty$, only the signal component is present at the sensor outputs and since this signal is random in nature, it should be possible to determine d^* exactly (at least in principle; in practice, the estimation accuracy may be limited by finite precision arithmetic); in the second case, as d^* increases, the effective period of time over which perfect correlation is possible between the signal components at the sensor outputs decreases, making it more and more difficult to estimate d^* accurately in the presence of noise.

The behaviors of η_{AML} and η_{EML} as functions of d^* and SNR are quite different. Indeed, although both quantities are similar for $d^* \leq 2T_s$, η_{AML} suddenly begins to exceed

[4]Because of the statistical uncertainty resulting from the finite number of simulations used (i.e., 500), a percentage of anomalous estimates corresponding to a number of anomalous trials between 1 and 4 is simply indicated as $<1\%$.

TABLE I
PERCENTAGE OF ANOMALOUS TIME DELAY ESTIMATES AS A FUNCTION OF TRUE DELAY FOR SNR = 0 dB

True Delay d^*/T_s	Percentage of Anomalous Estimates				
	EML	AML	WAML	MTCC	HEML
0	9.8	10.0	10.0	19.0	42.2
2	10.2	9.4	11.2	18.0	34.4
4	12.4	16.2	17.6	21.0	39.4
6	14.8	24.2	23.2	22.4	40.4
8	17.2	28.8	36.4	27.4	49.2
10	14.0	32.2	46.2	24.2	48.0
12	20.8	40.2	63.4	31.8	56.4
14	20.0	37.4	64.4	30.0	58.6
16	19.2	38.6	70.2	32.2	67.2

TABLE II
PERCENTAGE OF ANOMALOUS TIME DELAY ESTIMATES AS A FUNCTION OF TRUE DELAY FOR SNR = 10 dB

True Delay d^*/T_s	Percentage of Anomalous Estimates				
	EML	AML	WAML	MTCC	HEML
0	0	0	0	1.0	<1
2	0	0	0	<1	<1
4	<1	2.2	3.2	1.6	<1
6	0	10.6	12.8	3.4	1.8
8	<1	16.0	26.4	6.2	2.0
10	<1	17.6	38.4	6.8	3.8
12	<1	28.0	55.6	9.2	5.4
14	1.0	26.4	63.0	8.0	9.2
16	1.0	34.0	71.8	10.2	15.8

TABLE III
PERCENTAGE OF ANOMALOUS TIME DELAY ESTIMATES AS A FUNCTION OF TRUE DELAY FOR SNR = 20 dB

True Delay d^*/T_s	Percentage of Anomalous Estimates				
	EML	AML	WAML	MTCC	HEML
0	0	0	0	<1	0
2	0	0	0	<1	0
4	0	<1	1.0	<1	0
6	0	9.0	8.8	<1	0
8	0	13.6	21.6	1.6	0
10	0	16.8	34.4	3.0	0
12	0	24.0	53.6	4.2	0
14	0	24.0	60.8	3.8	0
16	0	32.0	72.6	5.0	<1

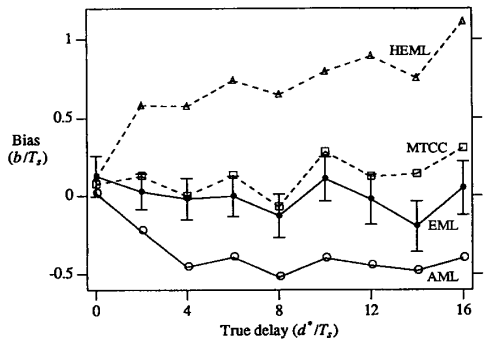


Fig. 2. Bias of nonanomalous time delay estimates as a function of true delay for SNR = 0 dB.

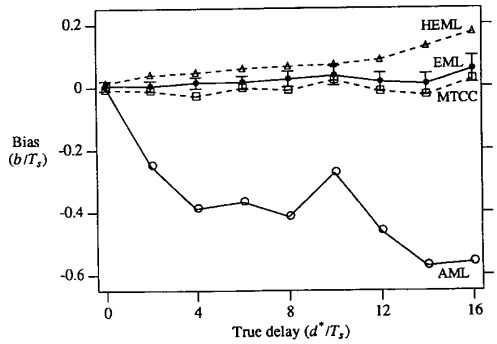


Fig. 3. Bias of nonanomalous time delay estimates as a function of true delay for SNR = 10 dB.

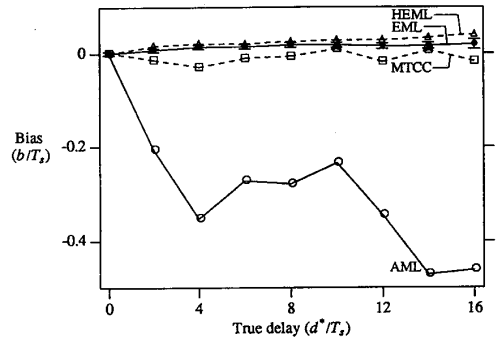


Fig. 4. Bias of nonanomalous time delay estimates as a function of true delay for SNR = 20 dB.

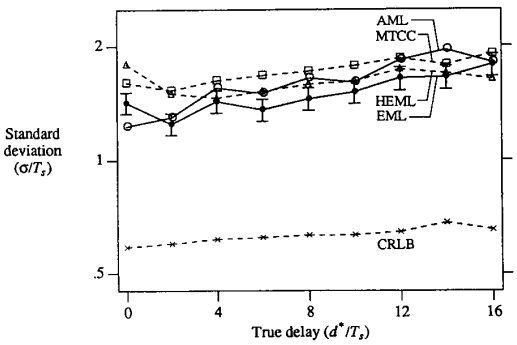


Fig. 5. Standard deviation of nonanomalous time delay estimates as a function of true delay for SNR = 0 dB.

η_{EML} around $d^* = 4T_s$, and for $d^* \geq 6T_s$, η_{AML} remains much larger than η_{EML}. What is even more striking is that this "threshold" phenomenon (in the d^*-domain) observed with AML does not disappear as the SNR increases. In contrast, such a threshold phenomenon is not observed with the EML estimator.

The WAML method, which was suggested as a possible approach to attenuate the threshold phenomenon observed with the AML method, does not produce the effect anticipated. In fact, the results for the WAML estimator are generally worse than those for the AML estimator. Although the introduction of a window function in the AML method may effectively lessen edge effects in some cases, it seems that multiplying $\boldsymbol{x}(t; d^*)$ in (1) by a nonconstant window function, and therefore destroying the simple structure of $\boldsymbol{a}(t; d^*)$, generally results in a loss of correlation between the two channels and hence a higher percentage of anomalous estimates. Thus, the WAML method will not be considered any further.

Interestingly, the results in Tables II and III indicate that at higher SNR, the simple MTCC method does not suffer as much as the AML method from a threshold phenomenon in the d^*-domain. However, we emphasize that η_{MTCC} is always larger than η_{EML}. As a result, the SNR required to make η_{MTCC} lower than a preset level is larger than that required for η_{EML}.

Finally, we note that η_{HEML} decreases to a level identical to that of η_{EML} as the SNR increases from 0 to 20 dB. This is consistent with the fact that the HEML estimator was obtained as an approximation to the EML estimator under the assumption of large signal power.

We now examine the statistical behavior of the nonanomalous time delay estimates. Fig. 2–4 show the sample bias of the nonanomalous estimates as a function of the true delay d^* for the EML, HEML, AML, and MTCC methods, and for SNR equals 0, 10, and 20 dB, respectively. The vertical line segments in these figures represent the 95% confidence intervals for the true bias of the EML estimator. The confidence intervals $[b_-, b_+]$ for the other estimators have been omitted for clarity of presentation, but they can be determined (within a good approximation) from the standard formula $b_\pm = b \pm 1.96\sigma/\sqrt{n_a}$, where n_a is the number of nonanomalous estimates, b is the sample bias, and σ is the sample standard deviation (see Figs. 5–7 for σ). For SNR = 0 dB (Fig. 2), the EML estimator is relatively unbiased (within experimental error) while the AML estimator is negatively biased and the HEML estimator is positively biased. The MTCC estimator is only slightly positively biased. Increasing the SNR (Figs. 3 and 4) eliminates the bias of the HEML and MTCC estimators but not that of the AML estimator which remains strongly biased even for SNR = 20 dB.

Figs. 5–7 show the sample standard deviation of the nonanomalous delay estimates as a function of the true

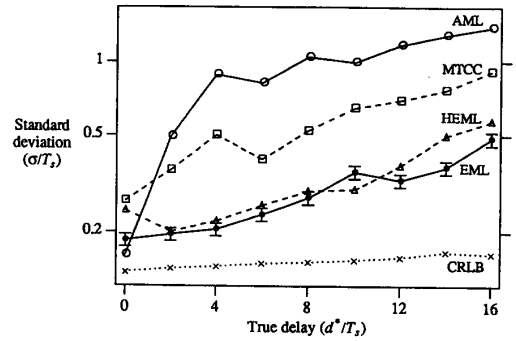


Fig. 6. Standard deviation of nonanomalous time delay estimates as a function of true delay for SNR = 10 dB.

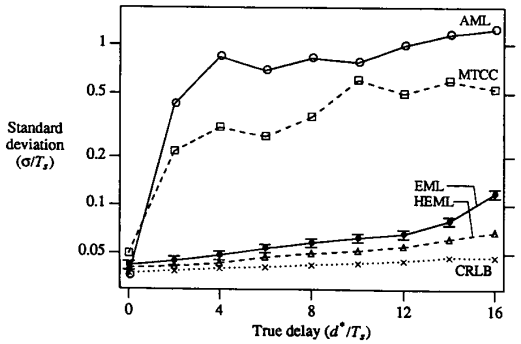


Fig. 7. Standard deviation of nonanomalous time delay estimates as a function of true delay for SNR = 20 dB.

delay d^* for the EML, HEML, AML, and MTCC methods, and for SNR equals 0, 10, and 20 dB, respectively. The vertical line segments in these figures represent the 95% confidence intervals for the true standard deviation of the nonanomalous EML estimates. The confidence intervals $[\sigma_-, \sigma_+]$ for the other estimators are given by $\sigma_\pm = \sigma \pm 1.96\sigma/\sqrt{2n_a}$ (again within a good approximation). Also shown in Figs. 5–7 is the Cramér–Rao lower bound (CRLB) on the variance of unbiased time delay estimators. The CRLB was computed by Monte Carlo simulations as follows. For each of the 500 independent experiments, the second derivative of the function $g(\xi) = \ln \Lambda(x; d^* + \xi)$ at $\xi = 0$, where $\ln \Lambda(x; d)$ is the version of the log-likelihood function (7) used to obtain the EML estimate, was evaluated numerically by means of the rapidly converging expansion

$$g''(0) = \frac{1}{T_s^2}\left\{-\frac{\pi^2}{3}g(0) + 2\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^2} \cdot [g(-nT_s) + g(nT_s)]\right\}. \quad (59)$$

This formula can be derived easily from the sampling theorem by assuming that $g(\xi)$ is band limited to $\pm 1/(2T_s)$ Hz. At the end of the 500 experiments, the Monte Carlo approximation to the CRLB, denoted σ^2_{CRLB}, was obtained as

$$\sigma^2_{\text{CRLB}} = -\langle g''(0)\rangle^{-1} \quad (60)$$

where $\langle \cdot \rangle$ indicates a sample average. This approach for computing the CRLB was motivated by the fact that the standard formulae available in the literature for the CRLB on time delay estimator variance [3], [4] are inapplicable here because the condition $T >> \tau_c$, d^* is not satisfied.

For SNR = 0 dB (Fig. 5), the four estimation methods give similar results. In this case, the standard deviations of the estimators exceed the CRLB by a factor comprised between 6.4 and 10.2 dB and come very close to the value of $2.31T_s$, which is the standard deviation of a random variable uniformly distributed within the interval $\pm\tau_c/2$. This indicates that time delay estimation is dominated by large errors, as confirmed by the results in Table I.

As the SNR increases from 0 to 10 dB (Fig. 6), the standard deviations of the EML, AML, and MTCC estimators, as a function of d^*, begin to depart quite significantly from each other. Although σ_{AML} is slightly lower than σ_{EML} at $d^* = 0$, the AML estimator experiences again a severe deterioration in performance as d^* increases from 0 to $4T_s$, and for $d^* \geq 4T_s$, σ_{AML} exceeds σ_{EML} by an averaged factor of 10.7 dB. σ_{MTCC}, although lower than σ_{AML} for all values of d^* except 0, still exceeds σ_{EML} by an averaged factor of 5.5 dB. It is also important to note that the variance of the EML estimator now exceeds the CRLB by only 2.6 dB at $d^* = 0$ and 5.4 dB at $d^* = 8T_s$ (in comparison to 7.7 and 7.3 dB, respectively, at SNR = 0 dB), indicating a trend towards optimality of the EML estimator as SNR increases. Finally, it is observed that σ_{HEML} is now very close to σ_{EML}.

The differences between σ_{EML}, σ_{AML}, and σ_{MTCC} are even more accentuated at SNR = 20 dB (Fig. 7). In this case, the results show that for $d^* \geq 2T_s$, σ_{AML} exceeds σ_{EML} by an averaged factor of 22.3 dB, while for σ_{MTCC}, the corresponding factor is 15.9 dB. In fact, for the AML estimator, increasing the SNR from 10 to 20 dB does not result in a very significant reduction of the standard deviation when $d^* \geq 2T_s$. The results also show that σ_{EML} approaches the CRLB within 1.0 dB at $d^* = 0$ and 2.8 dB at $d^* = 8T_s$. Hence, for such values of d^*, the EML estimator is nearly optimal. Finally, although Fig. 7 indicates that σ_{HEML} is slightly lower than σ_{EML}, it must be pointed out that b_{HEML} is always larger than b_{EML} in magnitude, so that the actual difference between the mean-square errors (about d^*) of the EML and HEML estimators is less than the difference between their variances.

Our final remark concerns the sensitivity of the EML method to modeling errors in $G_a(\omega)$. Since the HEML estimator is independent of $G_a(\omega)$ and its performance approaches that of the EML estimator in the limit of high SNR, it would seem that the EML estimator is relatively insensitive to errors in the assumed spectrum, at least at high SNR. As mentioned earlier, this is particularly important for applications in which complete knowledge of $G_a(\omega)$ is not available.

VI. Summary and Conclusions

In this paper, we have presented an exact solution to the problem of maximum likelihood time delay estimation between two sensors that is valid for arbitrarily short observation intervals. In other words, the standard assumption $T >> \tau_c$, $|d|$ made in the derivation of the conventional AML estimator has been relaxed. The resulting EML estimator has been shown to consist of a special finite-time beamformer, followed by a scalar log-likelihood processor based on the eigenvalues and eigenfunctions of a one-dimensional integral equation with nonconstant weight. The solution of this integral equation has been obtained for the case of stationary signals with rational power spectral densities. Finally, the performances of the EML, AML, and other related estimators have been compared by means of computer simulations for a first order autoregressive source signal and for system parameter values such that the condition $T >> \tau_c$, $|d|$ was not satisfied.

The results of these experiments have shown that the AML estimator suffers a dramatical deterioration in performance (large error, bias, and mean-square error) as the ratio d/T increases from 0 to about 0.1, making it essentially useless beyond this point. No such effect has been observed with the EML estimator which had the best overall performances. Moreover, for large SNR, the total mean-square error of the EML estimator came very close to the Cramér–Rao lower bound. Finally, the results have also shown that for large SNR, considerable simplifications of the EML estimator are possible without any significant loss in performance.

Appendix A

Proof of Property 1: Using (15), (16), and (3), it can be verified that

$$\lambda_i = \frac{1}{2\pi}\int_{-\infty}^{\infty} G_a(\omega)|A_i(\omega)|^2\, d\omega \tag{A1}$$

where

$$A_i(\omega) = \int_{-d}^{T} \psi_i(t)\,\rho(t)e^{-j\omega t}\, dt. \tag{A2}$$

Being the Fourier transform of a nonzero function, $A_i(\omega)$ must be different from zero in some interval (ω_1, ω_2), with $\omega_1 < \omega_2$, of the ω axis. Therefore, since $G_a(\omega)$ in (4) can only have isolated zeros, (A1) implies $\lambda_i > 0$. Next, since $G_a(\omega) = \gamma$ only at isolated values of ω, (A1) also implies

$$\lambda_i < \frac{\gamma}{2\pi}\int_{-\infty}^{\infty} |A_i(\omega)|^2\, d\omega. \tag{A3}$$

To complete the proof, simply observe that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |A_i(\omega)|^2\, d\omega = \int_{-d}^{T} \psi_i^2(t)\,\rho^2(t)\, dt \le 2. \tag{A4}$$

The equality on the left is a particular case of Parseval's relation, while the inequality on the right follows from (14) and (16). Combining (A3) and (A4), we finally conclude that $\lambda_i < 2\gamma$.

Proof of Property 2: Since $\lambda_i > 0$, (15) implies that

$$\psi_i(t) = \frac{1}{\lambda_i}\int_{-d}^{T} R_a(t-u)\,\psi_i(u)\,\rho(u)\, du. \tag{A5}$$

The continuity of $\psi_i(t)$ then follows immediately from that of $R_a(t)$.

Proof of Property 3: Consider the self-adjoint operators K and O defined by

$$[Kf](t) = \int_{-d}^{T} R_a(t-u)\, f(u)\,\rho(u)\, du, \qquad -d \le t \le T \tag{A6}$$

$$[Of](t) = f(T-d-t), \qquad -d \le t \le T \tag{A7}$$

where f is an arbitrary function in $L_2(\rho)$. Using the identities $R_a(\tau) = R_a(-\tau)$ and $\rho(t) = \rho(T-d-t)$, it is not difficult to verify that K and O commute, i.e., $KO = OK$. Hence, according to a well-known property of self-adjoint linear operators, there exists a complete orthonormal set of eigenfunctions common to both operators. For an eigenfunction $\psi_i(t)$ in that set, we have

$$\psi_i(T-d-t) = [O\psi_i](t) = \epsilon_i\psi_i(t) \tag{A8}$$

where ϵ_i is an eigenvalue of the operator O. However, since $O^2 = I$, where I is the identity operator in $L_2(\rho)$, the only possible eigenvalues of O are ± 1. This completes the proof.

Appendix B

Proof of Theorem 1: a) Consider the function

$$f(t) = \lambda_i\psi_{i1}(t) - \int_{-d}^{0} R_a(t-u)\,\psi_{i1}(u)\, du, \qquad -\infty < t < \infty. \tag{B1}$$

Denoting by $F(s)$ the BLT of $f(t)$, we have

$$F(s) = \lambda_i\Psi_{i1}(s) - \frac{N(s^2)}{D(s^2)}\Psi_{i1}(s), \qquad -\mu_0 < \mathrm{Re}\,(s) < \mu_0 \tag{B2}$$

where $\mu_0 = \min\{\mathrm{Re}\,(s)\colon D^-(s) = 0\}$. Let $\boldsymbol{D}^+$ and $\boldsymbol{D}^-$ be the differential operators defined by

$$\boldsymbol{D}^+ = D^+\left(\frac{d}{dt}\right), \qquad \boldsymbol{D}^- = D^-\left(\frac{d}{dt}\right). \tag{B3}$$

It can be shown that [15]

$$\boldsymbol{D}^+R_a(t) = 0, \qquad t > 0 \tag{B4}$$

$$\boldsymbol{D}^-R_a(t) = 0, \qquad t < 0. \tag{B5}$$

Hence, for $t < -d$, we have

$$\boldsymbol{D}^-f(t) = 0 \qquad (t < -d). \tag{B6}$$

Similarly, for $t > 0$, we have

$$\boldsymbol{D}^+f(t) = 0 \qquad (t > 0). \tag{B7}$$

For $-d < t < 0$, (28) and (15) imply

$$f(t) = \int_0^T R_a(t-u)\,\psi_i(u)\,\rho(u)\,du. \tag{B8}$$

Therefore

$$D^- f(t) = 0 \qquad (-d < t < 0). \tag{B9}$$

Now define ($i = 1, 2, 3$)

$$F_i(s) = \int_{J_i} f(t) e^{-st}\,dt \tag{B10}$$

where $J_1 = (-\infty, -d)$, $J_2 = (-d, 0)$, and $J_3 = (0, \infty)$. As a consequence of (B6), (B7), and (B9), it follows that

$$D^-(s)\,F_1(s) = e^{sd} P_1(s) \tag{B11}$$

$$D^-(s)\,F_2(s) = e^{sd} P_2(s) + P_3(s) \tag{B12}$$

$$D^+(s)\,F_3(s) = P_4(s) \tag{B13}$$

for some real coefficient polynomials $P_k(s)$ ($k = 1, \cdots, 4$) of degrees equal at most to $n-1$. Now, observing that $\Sigma F_i(s) = F(s)$ and using (B2), (30), (B11), (B12), and (B13), we finally obtain

$$\begin{aligned}[\lambda_i D(s^2) - N(s^2)]\Psi_{i1}(s) &= D(s^2)\,F(s) \\ &= D^+(s)\,D^-(s)[F_1(s) + F_2(s) + F_3(s)] \\ &= e^{sd} D^+(s)\,P_5(s) + Q_1(s)\end{aligned} \tag{B14}$$

for some real coefficient polynomials $P_5(s)$ and $Q_1(s)$ of degrees at most $n-1$ and $2n-1$, respectively. Proceeding in a similar way, it can be shown that

$$\begin{aligned}[\lambda_i D(s^2) - 2N(s^2)]\Psi_{i2}(s) &= Q_2(s) + e^{-s(T-d)} Q_3(s)\end{aligned} \tag{B15}$$

$$\begin{aligned}[\lambda_i D(s^2) - N(s^2)]\Psi_{i3}(s) &= e^{-s(T-d)} Q_4(s) + e^{-sT} D^-(s)\,P_6(s)\end{aligned} \tag{B16}$$

where $P_6(s)$ and $Q_k(s)$ ($k = 2, 3, 4$) are again real coefficient polynomials of degrees at most $n-1$ and $2n-1$, respectively.

In order to establish links between the various polynomials $P_k(s)$ and $Q_k(s)$, first consider the function

$$g(t) = \int_{-d}^T R_a(t-u)\,\psi_i(u)\,\rho(u)\,du, \qquad -\infty < t < \infty. \tag{B17}$$

The BLT of $g(t)$ is given by

$$G(s) = \frac{N(s^2)}{D(s^2)}[\Psi_{i1}(s) + 2\Psi_{i2}(s) + \Psi_{i3}(s)], \qquad -\mu_0 < \text{Re}\,(s) < \mu_0. \tag{B18}$$

Using (B3), (B4), and (15), it can be verified (exactly as we did for $f(t)$ in (B1)) that $g(t)$ in (B17) satisfies

$$D^- g(t) = 0, \qquad t < -d \tag{B19}$$

$$g(t) = \lambda_i \psi_i(t), \qquad -d < t < T \tag{B20}$$

$$D^+ g(t) = 0, \qquad t > T. \tag{B21}$$

As a result, $G(s)$ can also be expressed in the form

$$G(s) = e^{sd}\frac{P_7(s)}{D^-(s)} + e^{-sT}\frac{P_8(s)}{D^+(s)} + \lambda_i[\Psi_{i1}(s) + \Psi_{i2}(s) + \Psi_{i3}(s)]. \tag{B22}$$

Equating (B18) to (B22) and using (B14)–(B16) to eliminate the $\Psi_{ij}(s)$, we find

$$\begin{aligned}D^+(s)[P_5(s) + P_7(s)]e^{sd} + [Q_1(s) + Q_2(s)] \\ + [Q_3(s) + Q_4(s)]e^{-s(T-d)} \\ + D^-(s)[P_6(s) + P_8(s)]e^{-sT} = 0.\end{aligned} \tag{B23}$$

For this equation to be satisfied in the strip $-\mu_0 < \text{Re}\,(s) < \mu_0$, the polynomial coefficients of the various exponentials must vanish. From this requirement, we extract (among others) the relation

$$Q_1(s) + Q_2(s) = 0. \tag{B24}$$

In order to obtain a similar relation between $Q_2(s)$ and $Q_3(s)$, consider (27). In terms of the $\psi_{ij}(t)$ defined in (28), (27) implies

$$\psi_{i2}(t) = \epsilon_i \psi_{i2}(T-d-t) \tag{B25}$$

$$\psi_{i3}(t) = \epsilon_i \psi_{i1}(T-d-t) \tag{B26}$$

or, equivalently,

$$\Psi_{i2}(s) = \epsilon_i e^{-s(T-d)}\Psi_{i2}(-s) \tag{B27}$$

$$\Psi_{i3}(s) = \epsilon_i e^{-s(T-d)}\Psi_{i1}(-s). \tag{B28}$$

Combining (B15) with (B27), we obtain the desired relation between $Q_2(s)$ and $Q_3(s)$, namely,

$$Q_3(s) = \epsilon_i Q_2(-s). \tag{B29}$$

Equations (33) and (34) now follow easily from (B14) and (B15) by eliminating $Q_2(s)$ and $Q_3(s)$ with the help of (B24) and (B29) and setting $P_i(s) \equiv P_5(s)$ and $Q_i(s) \equiv Q_2(s)$. Because the $\Psi_{ij}(s)$ are analytic everywhere in the s plane, the singularities appearing in the right members of (33) and (34) are necessarily removable. Finally, note that (35) in the statement of the theorem is identical to (B28). This completes the first half of the proof.

b) Using the convolution theorem for the BLT, we have

$$\begin{aligned}\int_{-d}^T R_a(t-u)\,\psi_i(u)\,\rho(u)\,du &= \int_{-\infty}^{\infty} R_a(t-u)[\psi_{i1}(u) + 2\psi_{i2}(u) + \psi_{i3}(u)]\,du \\ &= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty}\left\{\frac{N(s^2)}{D(s^2)}[\Psi_{i1}(s) + 2\Psi_{i2}(s) + \Psi_{i3}(s)]\right\}e^{st}\,ds\end{aligned} \tag{B30}$$

where σ is any real number in the interval $-\mu_0 < \sigma < \mu_0$. However, by construction of the $\Psi_{ij}(s)$, the bracketed quantity under the integral sign in (B30) is equal to

$$\lambda_i \Psi_i(s) - e^{sd} \frac{P_i(s)}{D^-(s)} - \epsilon_i e^{-sT} \frac{P_i(-s)}{D^+(s)} \qquad \text{(B31)}$$

where $\Psi_i(s)$ is the BLT of $\psi_i(t)$. Therefore,

$$\int_{-d}^{T} R_a(t-u)\, \psi_i(u)\, \rho(u)\, du$$

$$= \lambda_i \psi_i(t) - \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{P_i(s)}{D^-(s)} e^{s(t+d)}\, ds$$

$$- \frac{\epsilon_i}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \frac{P_i(-s)}{D^+(s)} e^{s(t-T)}\, ds. \qquad \text{(B32)}$$

For $-d < t < T$, the two integrals on the left are zero. This shows that λ_i and $\psi_i(t)$ are solutions of (15). That $\psi_i(t)$ satisfies (27) is obvious by construction. This completes the proof.

REFERENCES

[1] V. H. MacDonald and P. M. Schultheiss, "Optimum passive bearing estimation in a spatially incoherent noise environment," *J. Acoust. Soc. Amer.*, vol. 46, pp. 37-43, 1969.

[2] E. J. Hannan and P. J. Thomson, "The estimation of coherence and group delay," *Biometrika*, vol. 58, pp. 469-481, 1971.

[3] W. R. Hahn and S. A. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 608-614, Sept. 1973.

[4] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 4, pp. 320-327, Aug. 1976.

[5] G. C. Carter, Ed., Special issue on time delay estimation, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, June 1981.

[6] R. V. Foutz, "Estimation of a common group delay between two multiple time series," *J. Amer. Stat. Ass.*, vol. 75, pp. 779-788, Dec. 1980.

[7] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, 1977.

[8] J. W. Betz, "Effects of uncompensated relative time companding on a broad-band cross correlator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 505-510, June 1985.

[9] S. Pasupathy, "Equivalence of transient and narrow-band signals," *J. Franklin Inst.*, vol. 305, pp. 49-56, Jan. 1978.

[10] A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation—Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 472-486, Apr. 1983.

[11] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, part I. New York: Wiley, 1968.

[12] J. A. Stuller, "Maximum-likelihood estimation of time-varying delay—Part I," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 300-313, Mar. 1987.

[13] I. M. G. Lourtie and J. M. F. Moura, "Delay estimation with non-stationary signals and correlated observation noises," in *Adaptive Methods in Underwater Acoustics*, H. G. Urban, Ed. Dordrecht, Holland: Reidel, 1985.

[14] B. Champagne, M. Eizenman, and S. Pasupathy, "Factorization properties of optimum space-time processors in nonstationary environments," *IEEE Trans. Signal Processing*, vol. 38, pp. 1853-1869, Nov. 1990.

[15] D. C. Youla, "The solution of a homogeneous Wiener-Hopf integral equation occurring in the expansion of second-order stationary random functions," *IRE Trans. Inform. Theory*, vol. IT-3, pp. 187-193, 1957.

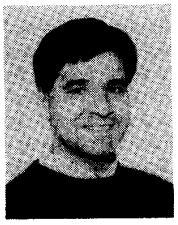
[16] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, part III. New York: Wiley, 1971.

[17] J. Mathews and R. L. Walker, *Mathematical Methods of Physics*, 2nd ed. Menlo Park, CA: Benjamin, 1970.

[18] T. Kailath, *Lectures on Wiener and Kalman Filtering*. New York: Springer, 1981.

[19] J. P. Ianniello, "Time delay estimation via cross correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 998-1003, Dec. 1982.

[20] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*. New York: Wiley-Interscience, 1971.

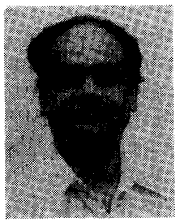


Benoit Champagne (S'87-M'89) was born in Joliette, Que., Canada, on January 13, 1961. He received the B.Ing. degree in engineering physics and the M.Sc. degree in physics from the University of Montreal in 1983 and 1985, respectively, and the Ph.D. degree in electrical engineering from the University of Toronto in 1990.

He is now an Assistant Professor at INRS-Télécommunications, Université du Québec. His current research interests are in the area of statistical signal processing, with particular emphasis on array processing and spectrum estimation.

Moshe Eizenman (M'84) was born in Tel-Aviv, Israel, on December 25, 1952. He received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, Ont., Canada, in 1978, 1980, and 1984, respectively.

Since 1982 he has developed several highly accurate eye-trackers that are currently being integrated into medical and military devices. In 1984 he joined the Institute of Biomedical Engineering at the University of Toronto as an Assistant Professor and founded El-MAR Inc. Since 1984 he has also been an Assistant Professor in the Electrical Engineering and Ophthalmology Departments at the University of Toronto. His current research interests include sonar signal processing, EEG and VEP analysis, control of eye movements, and the development of vision.



Subbarayan Pasupathy (M'73-SM'81-F'91) was born in Madras, Tamilnadu, India, on September 21, 1940. He received the B.E. degree in telecommunications from the University of Madras in 1963, the M.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 1966, and the M.Phil. and Ph.D. degrees in engineering and applied science from Yale University in 1970 and 1972, respectively.

He joined the faculty of the University of Toronto in 1973 and became a Professor of Electrical Engineering in 1983. His research interests lie in the areas of communication theory, digital communications, and statistical signal processing. He is a registered Professional Engineer in the province of Ontario. During 1982-1989 he was an Editor for *Data Communications and Modulation* for the IEEE TRANSACTIONS ON COMMUNICATIONS. He has also served as a Technical Associate Editor for the IEEE COMMUNICATIONS MAGAZINE (1979-1982) and as an Associate Editor for the *Canadian Electrical Engineering Journal* (1980-1983). Since 1984, he has been writing a regular column entitled "Light Traffic" for the IEEE COMMUNICATIONS MAGAZINE.