Detection of Random Transient Signals via Hyperparameter Estimation

Roy L. Streit and Peter K. Willett, *Senior Member, IEEE*

***Abstract*—Difficulties arise with the generalized likelihood ratio test (GLRT) in situations where one or more of the unknown signal parameters requires an enumeration that is computationally intractable. In the transient signal detection problem, the frequency characteristics of the signal are typically unknown; therefore, even if an aggregate signal bandwidth is assumed, the estimation problem intrinsic to the GLRT requires an enumeration of all possible sets of signal locations within the monitored band.**

In this paper, a prior distribution is imposed over those portions of the signal parameter space that traditionally require enumeration. By replacing intractable enumeration over possible signal characteristics with an *a priori* signal distribution and by estimating the "hyperparameters" (of the prior distribution) jointly with other signal parameters, it is possible to obtain a new formulation of the GLRT that avoids enumeration and is computationally feasible. The GLRT philosophy is not changed by this approach—what is different from the original GLRT is the underlying signal model. The performance of this new approach appears to be competitive with that of a scheme of emerging acceptance: the "power-law" detector.

***Index Terms*—Detection, EM algorithm, short-duration signals.**

I. Introduction and Context

IN UNDERWATER passive surveillance systems, many or most signals of interest are of transient nature and are otherwise of unknown structure and location. An example discussed in more depth later, but presented at this point mostly to focus the reader's attention, is given in Fig. 1. Such signals naturally pose a challenging detection problem since the hypotheses (signal-absent/signal-present) to be tested are highly composite. Essentially, the detector is given a record of observation samples and is tasked to respond whether all samples come from a stationary probability law—usually white Gaussian noise is assumed, and we do not deviate from that here—or whether they do not. Many excellent treatments of this problem are available in the literature. Whether or not there is a neat taxonomy of these is of course arguable, but for presentational purposes, it is convenient to categorize them as having come from three viewpoints.

Manuscript received October 27, 1997; revised December 14, 1998. This work was supported by the Office of Naval Research under Contracts N00014-98-1-0049 and N00014-98-WX30115. The associate editor coordinating the review of this paper and approving it for publication was Dr. Shubha Kadambe.

R. L. Streit is with Code 202, Naval Undersea Warfare Center, Newport, RI 02841 USA.

P. K. Willett is with the Department of Electrical and Systems Engineering, University of Connecticut, Storrs, CT 06269 USA.

Publisher Item Identifier S 1053-587X(99)04714-5.

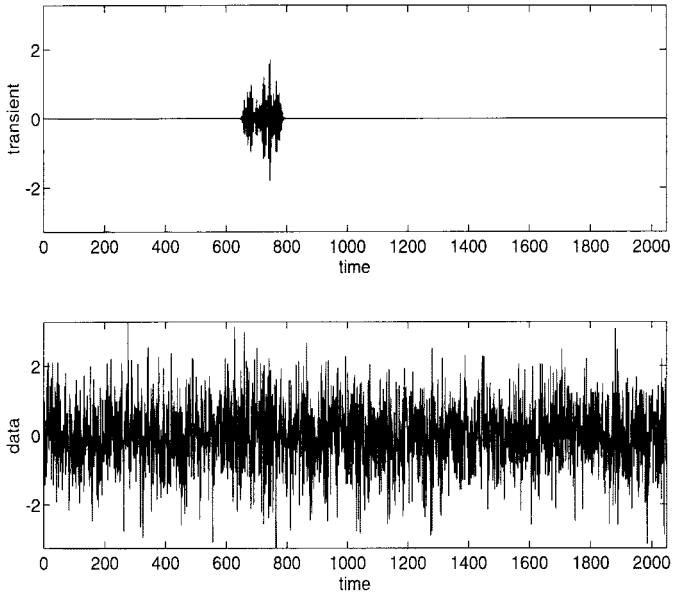

Fig. 1. Example of the sort of short-duration signal in which we are interested. The upper plot shows a narrowband transient, to which noise is added in the lower plot. The per-sample SNR is -3 dB. The signal was created by passing white Gaussian noise through an FIR filter with a passband $.4\pi \leq \omega \leq .6\pi$.

Under the first of these, which we refer to as the *change-point viewpoint*, we begin by assuming that not only is the structure of the transient signal unknown but that there is no structure save of temporal contiguity. More specifically, we have for observation sequence $\{X_n\}$

$$X_n \text{ has density } \begin{cases} f_0(x_n), & n < n_s \\ f_1(x_n), & n_s \leq n < n_s + n_d \end{cases} \tag{1}$$

where, for example, f_0 could refer to iid unit Gaussian samples and f_1 to iid Gaussian samples with variance $\sigma^2 > 1$ and in which n_s and n_d are, respectively, the transient signal's start time and duration. This probabilistic structure is very similar to that of the *permanent* change-detection problem [as in (1) with $n_d = \infty$] for which an appropriate and accepted approach is the Page or *cusum* scheme (see, among many references, [1]). In this case, the statistic

$$Z_n = \max\{0, Z_{n-1} + g(X_n)\} \tag{2}$$

is continually updated and compared with a threshold h with an exceedance indicating that a detection is declared. The nonlinearity g may be anything desired but is optimally the log-likelihood ratio $g(x) = \log(f_1(x)/f_0(x))$. The idea is that

if a transient is of sufficient length and strength that a threshold exceedance is produced before its end, then the permanent and temporary change detection problems are for the user's purposes identical; however, in any case, it can be shown that the Page scheme is a GLRT over the unknown parameters n_s and n_d [2], and in fact, it can be shown that the scheme works very well [3]. Further, although the initial assumption above was of sample-to-sample independence, variations exist based on richer dependency structures.

From the second viewpoint, which we call *signal-parametric* or *linear-subspace* based, a transient signal has not only a temporal extent but also a structure, with the task being to determine that structure and base a GLRT upon it. Borrowing from [4], the observations model is

$$\boldsymbol{x} = \boldsymbol{S}\boldsymbol{a} + \boldsymbol{e} + \boldsymbol{w} \tag{3}$$

in which $\boldsymbol{x}$ denotes the observations sequence arranged into a column vector, $\boldsymbol{S}$ and $\boldsymbol{a}$ are, respectively, a matrix and vector relating to the structure, $\boldsymbol{e}$ describes whatever modeling error remains from $\boldsymbol{S}\boldsymbol{a}$, and $\boldsymbol{w}$ is a white noise vector. The key here is that the dimension of $\boldsymbol{a}$ is small relative to that of $\boldsymbol{x}$, meaning that the transient signal—whatever it is—lies in a subspace of low dimension. If this subspace can be identified with reasonable fidelity, then a GLRT is possible.

Now, it is clear that this subspace dimension *can* be as low as one: the signal itself. This is not as trivial an observation as one may suppose since it, coupled with the knowledge of temporal contiguity, gives rise to the not-particularly-favored energy-detector GLRT structure in [5]. A more interesting class of procedures is available, however, if there is assumed to be some preprocessing linear transformation to $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}$ such that transient-signal energy is "aggregated" into relatively few components. Estimation of these components is thus facilitated, and a GLR structure is made possible. Preprocessing steps include those based on wavelets [6], [7], and on the Gabor representation [2], [8], [9].

A transient signal as observed from the third viewpoint, referred to here as *nonparametric*, is similar to that from the first: Although the signal may have structural parameters that *could* perhaps be exploited in a GLRT, these are not explicitly sought. Instead, features that a transient-signal-bearing observation record might reasonably be expected to have are identified, and testing proceeds from statistics of these. Most likely due to its relative insensitivity to starting point, the frequency domain appears to have become the home for these. For example, the effects of transients upon higher order spectra are exploited in [10]–[12], and correlation in the frequency domain is the focus of [13]. In [14], the assumed tendency of transient energy to be concentrated both in time and in frequency leads to a Gaussian-mixture time-spectrogram model. Of greatest interest here is the frequency domain detector in which aggregation of transient signal energy into a relatively few DFT bins is amplified, if the per-bin SNR is sufficiently large, by a "power-law" statistic [15].

Although this third class of algorithms may appear to be the "miscellaneous" category, this would imply an unintended lack of respect. In fact, they work very well, and our experience is that the power-law detector is tough to beat. We would prefer to think of the third category as, at least to some extent, nonparametric. We have qualified this last since the processing is based on a quantifiable aspect of a transient signal. For example, Nuttall [15][1] has shown, analogous to [18] under different assumptions, that the power-law statistic is a close approximation to the log-likelihood ratio over a wide class of transient structures. This result is analogous to that in [18], with the exception that Nuttall's detector works in the frequency domain and that of [18] is a time-domain power-law processor.

In this paper, we present a new *hyperparametric* approach. Because a GLRT is used, the idea is similar to that labeled the "second" viewpoint above. However, whereas the above techniques rely on (maximum likelihood) estimates of various unknown signal parameters—amplitude, frequency, location, etc.—here, we assume that all these are drawn from an underlying probability distribution. This prior *meta*-distribution itself has unknown parameters (we call them *hyperparameters*), and our goal is to insert estimates of these into a GLRT structure. It turns out that this estimation can be done very efficiently and neatly via the estimation-maximization (EM) algorithm [19]–[22].

This work has all along been motivated by the transient detection problem and, in particular, by the success of the power-law detector in that regard. Hence, we also use its preprocessing step of the short-time (discrete) Fourier transform, and indeed, our comparisons are to the power-law detector. However, we wish to note that the recasting of many composite detection problems as involving hyperparameters is possible, and the success of the approach is largely dictated by the ease by which the EM algorithm proceeds. For example, the basic assumptions are that the DFT magnitudes are conditionally independent and distributed according to exponential laws: homogeneously under the no-transient hypothesis H_0 and as two populations under the transient-present alternative H_1. There is thus but one parameter to estimate under H_0 (the mean level), whereas under H_1, there are three parameters (two mean levels and one mixing proportion).

In the following sections, we discuss the hyperparameter abstraction and derive the GLRT and estimation procedures. We then compare the performance of our scheme to the power-law detection; comparison is on the basis of simulation since here, as in many GLRT structures, direct analytical evaluation appears to be impossible. In this basic case, our scheme performs comparably to the power-law detector. However, a feature of our formalism is its extensibility to more complex meta-distributional assumptions, and we explore a number of these. Specifically, we relax our assumption of binary to multiple levels under H_1, we explore constant false-alarm rate processing, we investigate the applicability of a hidden Markov model (HMM) meta-distribution to model contiguity of transient-energy-bearing DFT bins. Most of the above results begin with DFT data as simulated in the frequency domain; we also look at time-domain transient signals (i.e., what we are really dealing with) and demonstrate that the

[1] There are a number of excellent technical reports on this subject, among them [15]–[17].

hyperparameter approach shows greater promise even than indicated by frequency domain simulation.

II. Theory of Hyperparameter-Based GLR Detection

A. Abstraction of the Detection Problem

The focus of this paper is signal detection. Posed in a hypothesis-testing framework, we have an observation X (generally not scalar) with probability law $dF_\theta(X)$, where θ is a parameter. We assume two mutually exclusive and exhaustive sets Θ_0 and Θ_1, and the test is to decide between

$$\begin{aligned} \boldsymbol{H_0}\ &\theta \in \Theta_0 \\ \boldsymbol{H_1}\ &\theta \in \Theta_1 \end{aligned} \tag{4}$$

based on the observation X.

The application of greatest interest to us here is transient detection. Here, the hypotheses are *composite* (many-valued), as the nature, duration, and location are unknown. As we shall discuss, our approach is to pose a physically meaningful and tractable prior distribution on a subset of θ and to estimate the remaining parameters and hyperparameters (the parameters of the posed prior) jointly.

The problem in which we are interested here can most conveniently be expressed as

$$X \sim dF_\theta(X) = \int_Z dF_\theta(X|Z)\, dG_\theta(Z) \tag{5}$$

coupled with the usual hypothesis-test of (4). What is notable here is that the problem is *hybrid*; it is composite through the unknown and multivalued parameter θ, and a prior distribution (conditioned on θ) over some *intermediate* and *hidden* random quantity Z is available. The fact that Z is hidden suggests that the expectation-maximization (EM) algorithm may be of interest, and the hybrid nature suggests a hybrid approach.

Our approach, which was first propounded in [23],[2] is to develop and use the test statistic

$$T(X) = \frac{\max_{\theta\in\Theta_1}\left\{\int_Z dF_\theta(X|Z)\, dG_\theta(Z)\right\}}{\max_{\theta\in\Theta_0}\left\{\int_Z dF_\theta(X|Z)\, dG_\theta(Z)\right\}} \tag{6}$$

to be compared to a threshold. In most cases, the hidden process Z will be relevant only under the signal-present alternative H_1, and hence, we may write (6) as

$$T(X) = \frac{\max_{\theta\in\Theta_1}\left\{\int_Z dF_\theta(X|Z)\, dG_\theta(Z)\right\}}{\max_{\theta\in\Theta_0}\{dF_\theta(X)\}}. \tag{7}$$

In many cases, the problem remains that the integration—which is generalized and may be a summation—is very complicated. Our solution is to use the EM algorithm [19] to perform the maximization in a very convenient way; other numerical methods may be considered but are often less tractable.

[2]These ideas were first developed in the presentation cited; this paper has much in common with [24], and some extensions to this work are in [25].

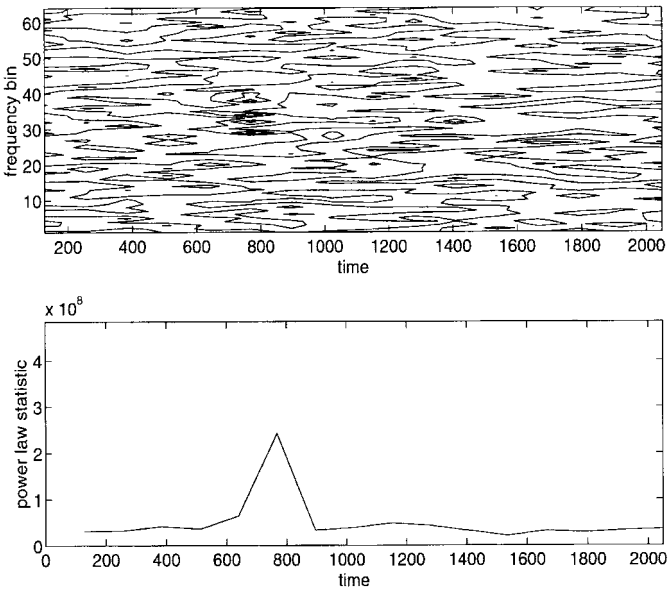

Fig. 2. Upper plot shows contours of the magnitude of the block-FFT (length 128) processed noisy data in Fig. 1. The lower plot is the power-law statistic (with power $\nu = 2.5$) as operated on the data in the upper plot. Notice that the transient is difficult to observe in both the upper plot and the lower plot of Fig. 1 but that the power-law detector is effective in its enhancement.

B. The Model and Detector

Let us consider the detection of a short-duration signal. A popular preprocessing step is the magnitude-squared FFT. For signals of practical interest, provided the duration is short compared with the FFT length N, it is expected that energy be concentrated in M $(<N)$ bins. Statistically, this amounts to N unit-exponential random variables under H_0 or $N-M$ unit-exponentials and M exponentials with increased scale parameter under H_1.[3] All random variables are assumed independent, and the possibly thorny issue of normalization is, for now, ignored.

A detector attracting considerable interest of late uses the "power-law" statistic [15]

$$T_{PL}(\boldsymbol{X}) = \sum_{i=1}^{N} X_i^\nu \tag{8}$$

in which X_i denotes the ith magnitude-squared FFT bin output. Some thought reveals that $\nu = 1$ is optimal for $M = N$ and that $\nu = \infty$ corresponds to the GLR for $M = 1$. It has been discovered computationally that $2 \le \nu \le 3$ is a good choice over a wide range of M. An example of this detector, and of the situation in which we are interested, is given in Figs. 1 and 2.

The detector of [15] is not optimal; however, it is analytically derived as an approximation to the detector optimal under the assumption that all possible combinations of M signal-containing bins—with M fixed—are equally likely under H_1.

[3]If the original observation sequence is real and of length $2N$, then refer to its DFT as $\{Y_k\}_{k=0}^{2N-1}$. In this case, it is necessary to form N ersatz "FFT bins" as $\{X_k\}_{k=0}^{N-1} = \{Y_0 + jY_N, Y_1, \cdots, Y_{N-1}\}$.

The optimal detector under this assumption is

$$T_{\text{opt}}(\boldsymbol{X}) = \frac{1}{\binom{N}{M}} \sum_{\mathcal{S}} \frac{1}{\mu_1^M} e^{\Sigma_{i\in\mathcal{S}}(1-1/\mu_1)X_i} \tag{9}$$

in which $\mathcal{S}$ denotes any of the possible ways to select M of the integers $0, 1, \cdots, N-1$ (i.e., the signal-containing bins), and μ_1 is the exponential scale parameter for such bins. Nuttall's derivation begins with the writing of the exponentials in (9) as a power series and proceeds with appropriate collection of terms. It is clear, since $\mathcal{S}$ has cardinality $\binom{N}{M}$ (typical values might be $N = 128$ and $M = 50$, meaning that there are about 10^{36} $\mathcal{S}$'s), that exact calculation of T_{opt} is not generally feasible. A further complication is that M is not known, and hence, if optimality is desired, some distribution for it must be posited and averaged over. It turns out that the approximated form depends neither on M nor on μ_1, which, as noted in [15], is fortunate.

The power-law statistic has been found to work very well, but it is a (clever) tractable approximation to the intractable one used in the optimal test. It is worth asking whether by some intelligent grouping of terms either full or at least "more" optimality could be preserved. An interesting answer emerges from the additional specification that the number of signal-containing bins M itself be uniformly distributed; we get

$$T_{\text{opt}}(\boldsymbol{X}) = \prod_{i=0}^{N-1} \left(1 + \frac{1}{\mu_1} e^{(1-1/\mu_1)X_i}\right) \tag{10}$$

by averaging (9) over M. The statistic of (10) is certainly tractable, and we further note that it is optimal under the signal model

$$\begin{aligned} \boldsymbol{H_0}\text{:}\quad & \boldsymbol{X} \sim \prod_{i=0}^{N-1} e^{-X_i} \\ \boldsymbol{H_1}\text{:}\quad & \boldsymbol{X} \sim \prod_{i=0}^{N-1} \left(\frac{1}{2} e^{-X_i} + \frac{1}{2\mu_1} e^{-X_i/\mu_1}\right) \end{aligned} \tag{11}$$

meaning that if a transient signal is present, each DFT bin is equally likely: either exponential of unit scale or of increased scale.

Weaknesses of the model (11) include both the need to specify μ_1 and the fixed "fair coin flip" bin-occupancy. Our model, then, is a generalization of (11); we do not fix M but instead allow it to vary according to a binomial distribution, where the hyperparameter q of the binomial distribution is to be estimated. We use the same "uniform" formulation as above, but due to this binomial distribution, or, perhaps more accurately, due to the underlying Bernoulli random variates, we are able to derive the optimal processor in a tractable form. More specifically, we assume the model

$$\begin{aligned} \theta\text{:}\quad & \begin{cases} \{q, \mu_{0,0}\}, & H_0 \\ \{q, \mu_{1,0}, \mu_{1,1}\}, & H_1 \end{cases} \\ \{Z_i\}\text{:}\quad & \begin{cases} 0, & H_0 \\ iid\ \mathcal{B}(Z_i; q), & H_1 \end{cases} \\ \{X_i\}\text{:}\quad & \begin{cases} \text{independent } \mathcal{E}(X_i; \mu_{0,0}), & H_0 \\ \text{independent } \mathcal{E}(X_i; \mu_{1,Z_i}), & H_1 \end{cases} \end{aligned} \tag{12}$$

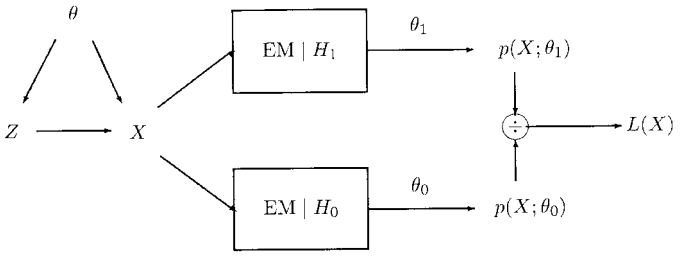

Fig. 3. Illustration of the procedure. The (hyper-)parameters θ control the generation of the "hidden" random variables Z and the observation X. This observation is used in a pair of "EM" estimation algorithms (under H_0 and H_1) to generate separate estimates of θ. These estimates are used in a GLR procedure.

in which $\mathcal{E}(X; \mu)$ means that X is exponential with scale parameter μ and in which $\mathcal{B}$ similarly refers to a Bernoulli random variable. Note that under the model as given, $\mu_{0,0}$ and $\mu_{1,0}$ are to be estimated; an assumed perfect normalization fixes these values to unity (as for the power-law detector) and removes the estimation over these parameters.

In the Appendix, we show how to estimate θ from X; in fact, assuming that convergence of the EM algorithm to the global maximum has occurred, we obtain the ML estimate

$$\theta_i^* = \arg\max_{\theta\in\Theta_i} \{dF_\theta(X)\} \tag{13}$$

for $i = 0, 1$. We are thus able to determine

$$\begin{aligned} \max_{\theta\in\Theta_1} \{dF_\theta(X)\} &= \max_{\theta\in\Theta_1} \left\{ \prod_{i=1}^{N} \left[(1-q) \frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} \right.\right. \\ &\qquad \left.\left. + q \frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}} \right] \right\} \\ &= \prod_{i=1}^{N} \left[(1-q^*) \frac{1}{\mu_{1,0}^*} e^{-X_i/\mu_{1,0}^*} \right. \\ &\qquad \left. + q^* \frac{1}{\mu_{1,1}^*} e^{-X_i/\mu_{1,1}^*} \right]. \end{aligned} \tag{14}$$

For the denominator, we have

$$\begin{aligned} &\max_{\theta\in\Theta_0} \{dF_\theta(X)\} \\ &\quad = \max_{\theta\in\Theta_0} \left\{ \prod_{i=1}^{N} \left[\frac{1}{\mu_{0,0}} e^{-X_i/\mu_{0,0}} \right] \right\} \\ &\quad = \begin{cases} \left(\dfrac{N}{\sum_{i=1}^{N} X_i} \right)^N e^{-N} & \mu_{0,0} \text{ estimated} \\ \dfrac{1}{\mu_{0,0}^N} e^{-(\Sigma_{i=1}^N X_i)/\mu_{0,0}} & \mu_{0,0} \text{ known} \end{cases} \end{aligned} \tag{15}$$

depending on whether or not the data has been normalized. The procedure is illustrated in Fig. 3.

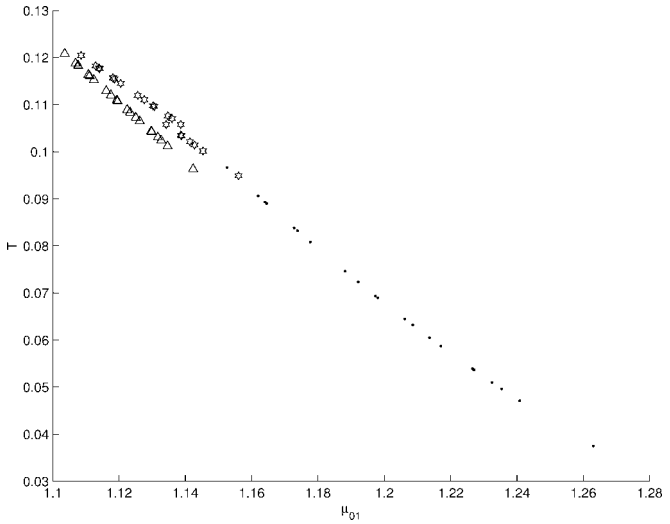

Fig. 4. Scatter plot of EM values returned for $\hat{\mu}_{1,1}$ versus GLRT statistic T in the case that $M = 50, N = 128$, and the situation where H_0. These are the results of 50 different EM procedures, each on the same data with different initializations. The points are for tolerance $\epsilon_{\rm tol} = 10\%$, the hexagons for $\epsilon_{\rm tol} = 1\%$, and the triangles for $\epsilon_{\rm tol} = 0.01\%$.

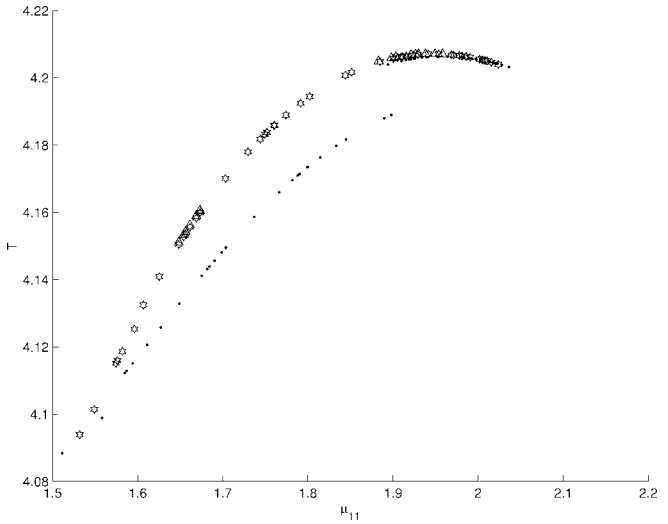

Fig. 5. Scatter plot of EM values returned for $\hat{\mu}_{1,1}$ versus GLRT statistic T in the case that $M = 50, N = 128$, and the aggregate SNR $= 60$ (i.e., the situation is H_1). These are the results of 50 different EM procedures each on the same data with different initializations. The points are for tolerance $\epsilon_{\rm tol} = 10\%$, the hexagons for $\epsilon_{\rm tol} = 1\%$, and the triangles for $\epsilon_{\rm tol} = 0.01\%$.

C. EM Initialization and Convergence

The EM algorithm is, of course, just a good computational procedure for finding maximum-likelihood estimators of parameters. It is an "ascent" method, meaning that each iteration finds new parameters that increase the likelihood function; hence, the geography of the likelihood function landscape is of major importance since only in the case of convexity is convergence to a true maximum assured, and indeed, there is no general guarantee of convergence at all.

Regarding the latter issue, it is known [19] that if the likelihood function is bounded, then the EM algorithm will converge to at least a local maximum; in the present case, the likelihood function is indeed bounded above, as can be seen from the fact that we have

$$\lim_{\mu \to 0} \frac{1}{\mu} e^{-X/\mu} = \lim_{\mu \to \infty} \frac{1}{\mu} e^{-X/\mu} = 0 \tag{16}$$

for any positive X.

As regards convergence to a local maximum, it is our experience that this can, in fact, occur. In Figs. 4 and 5, we show the results of 50 different EM procedures, each on the same data but with different initializations, for $N = 128, M = 50$, and, in the latter figure, SNR $= 60$ (the situation of Fig. 7 to follow), with prior knowledge that $\mu_{1,0} = \mu_{0,0} = 1$. The initializations used were so that $\hat{q}(0) = 0.5$, and $\hat{\mu}_{1,1}$ is drawn uniformly in (1,2). We define the "relative error" stopping criterion

$$\text{Stop when } \frac{\hat{\mu}_{1,1}(n)}{\hat{\mu}_{1,1}(n-1)} + \frac{\hat{q}(n)}{\hat{q}(n-1)} \leq \epsilon_{\rm tol} \tag{17}$$

in which $\hat{\mu}_{1,1}(n)$ and $\hat{q}(n)$ are the values of the two parameters being estimated at EM iteration n; the three sets of points in the figure refer to $\epsilon_{\rm tol} = 0.1$, 0.01, and 0.0001. Based on our computational experience, as exemplified by this figure, we note the following.

- Tightening the tolerance improves the reliability of the test statistic.
- There can be more than one point to which eventual convergence is drawn.
- While the effect on the test statistic of a local maximum is not negligible, the "spread" of T is not large.

TABLE I
ILLUSTRATION OF CONVERGENCE OF EM ALGORITHM, FROM SIMULATION. THE FIRST COLUMN IS THE TRUE HYPOTHESIS, AND SECOND IS THE VALUE OF $\epsilon_{\rm tol}$ USED [SEE (17)]. THE THIRD THROUGH SIXTH COLUMNS SHOW THE PERCENTAGE OF EM TRAJECTORIES, WITH RANDOM INITIALIZATIONS, WHICH LIE WITHIN 1, 5, 10, AND 25% OF THE MEDIAN VALUE, WHILE THE SEVENTH THROUGH NINTH COLUMNS DENOTE THE MEDIAN, MEAN, AND MAXIMUM OF THE NUMBER OF ITERATIONS REQUIRED TO MEET THE STATED CONVERGENCE CRITERION $\epsilon_{\rm tol}$

truth	ϵ_{tol}	1 %	5 %	10 %	25%	med.	mean	max
H_0	.1	73	60	50	36	4	4	6
H_0	.01	58	43	32	23	6	6	40
H_0	.0001	37	28	22	18	10	67	620
H_1	.1	47	5	.2	0	3	3	4
H_1	.01	34	3	.08	0	5	5	39
H_1	.0001	.5	.1	.01	0	106	144	539

The last point, in particular, is amplified in Table I, which shows Monte Carlo results in the above situations and further gives the number of runs needed to achieve the various values of $\epsilon_{\rm tol}$. Convergence of the EM algorithm to proper parameter estimates is important, and perhaps improvements could be made in terms of intelligent initializations, but based on our simulation experience, it appears that the detector is not overly sensitive.

III. RESULTS

A. The Basic Assumption Case

Here, we compare the detectors developed in the previous section to the "power-law" detector. The situation is of $N = 128$ FFT bins; under H_0, each is distributed as unit-

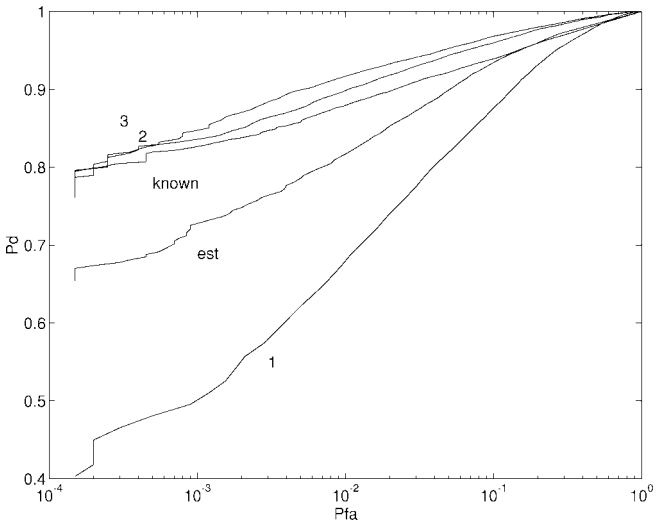

Fig. 6. Receiver operating characteristic for aggregate SNR $\Sigma_{i=1}^{N}\ [\mu_{1,Z_i} - \mu_{0,0}] = 40$, $N = 128$ FFT bins, and $M = 5$ signal-containing bins. The numbers 1, 2, and 3 refer to the power ν in (8), whereas "known" and "est" correspond to the new detector with $\mu_{0,0}$ fixed at its correct value of unity and estimated, respectively. This is the "basic assumption" case, in which random variates are exponential with means $\mu_{0,0} = 1$ under H_0 and either $\mu_{1,0} = 1$ ($N - M = 123$ of them) or $\mu_{1,1} = 1 + \text{SNR}/M = 9$ ($M = 5$ of them) under H_1.

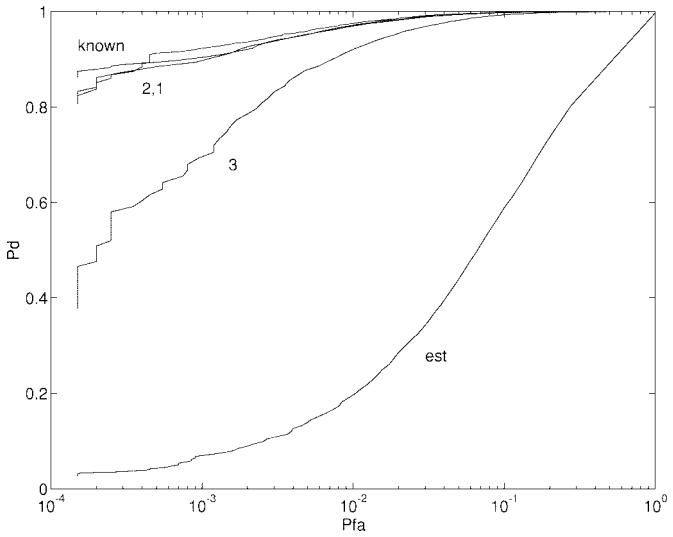

Fig. 7. Receiver operating characteristic for aggregate SNR of 60, $N = 128$ FFT bins, and $M = 50$ signal-containing bins. The labels are as in Fig. 6. Like that figure, this is the "basic assumption" case, in which random variates are exponential with means $\mu_{0,0} = 1$ under H_0 and either $\mu_{1,0} = 1$ ($N - M = 78$ of them) or $\mu_{1,1} = 1 + \text{SNR}/M = 2.2$ ($M = 50$ of them) under H_1.

exponential, whereas under H_1, $(N-M)$ are unit-exponential, and M are exponential with mean $(1 + \text{SNR}/M)$—note that this fixed-M situation is that under which (8) was developed and does not match our Bernoulli model of (12). Comparison is on the basis of 100 000 simulations, and the ordering of the M signal-containing bins under H_1 is relevant to none of the detectors implemented; hence, we arbitrarily choose these as $\{X_i\}_{i=0}^{M-1}$. It will be obvious that the proposed detector with unknown (i.e., estimated) $\mu_{0,0}$ is not particularly good, but to a large extent, this is an unfair comparison since it must estimate a parameter known by all other schemes.

In Fig. 6, we see the results for the case $M = 5$. In this "narrowband" case, it is not surprising that the power law detectors with amplifying exponents (i.e., $\nu = 2, 3$) do quite well, nor is it surprising that the energy-detector $(\nu = 1)$ is poor. However, it is notable that the proposed detector offers only a slight degradation relative to the power-law detector, that is, it is essentially as good as is currently available.

In Fig. 7, we see the same situation, except now in the relatively broad-band case $M = 50$. Here the less-amplifying exponents ($\nu = 1$, 2) are good and $\nu = 3$ is poor; remarkably, we observe that the new detector (with known $\mu_{0,0}$ and $\mu_{1,0}$) is actually *superior* to the power-law detector.

B. Extension to Multiple Level Transients

One possible criticism of the model (12) is its assertion that a signal, if one exists, is either on or off. The power-law detector of (8) does not require this, and indeed, a realistic transient event is unlikely to be so ideally tuned that all in-band energy is evenly distributed.

There is no particular difficulty, however, in modifying our detector to accommodate multiple signal levels. We rewrite (12) as

$$\begin{aligned}
\theta\text{:}\quad & \begin{cases} \{\boldsymbol{q}, \mu_{0,0}\}, & H_0 \\ \{\boldsymbol{q}, \mu_{1,0}, \mu_{1,1}\}, & H_1 \end{cases} \\
\{Z_i\}\text{:}\quad & \begin{cases} 0, & H_0 \\ iid\, \mathcal{C}(Z_i; \boldsymbol{q}), & H_1 \end{cases} \\
\{X_i\}\text{:}\quad & \begin{cases} \text{independent } \mathcal{E}(X_i; \mu_{0,0}), & H_0 \\ \text{independent } \mathcal{E}(X_i; \mu_{1,Z_i}), & H_1 \end{cases}
\end{aligned} \tag{18}$$

in which $\mathcal{C}$ now refers to probability mass function over the integers $\{0, \cdots, P-1\}$ such that $\Pr(Z_i = j) = q_j$ and in which P is the number of distinguishable levels. Extension of our hybrid procedure to this case is straightforward. Equation (14) is replaced by

$$\max_{\theta \in \Theta_1} \{dF_\theta(X)\} = \prod_{i=1}^{N} \left[\sum_{j=0}^{P-1} q_j^* \frac{1}{\mu_{1,j}^*} e^{-X_i/\mu_{1,j}^*} \right] \tag{19}$$

in which the EM estimates for the parameters, corresponding to (33) in the Appendix, become

$$\begin{aligned}
\mu_{1,j}^* &= \frac{\sum_{i=1}^{N} w_i(j) X_i}{\sum_{i=1}^{N} w_i(j)} \\
q_j^* &= \frac{1}{N} \sum_{i=1}^{N} w_i(j)
\end{aligned} \tag{20}$$

and the posterior probabilities $\{w_i(j)\}$ are defined in a manner analogous to (31).

In Fig. 8, we compare the performance of the hybrid detector with that of the power law detector for the case that $P = 3$, that is, there are indeed two different power levels among the signal-containing bins: one twice the other. More specifically,

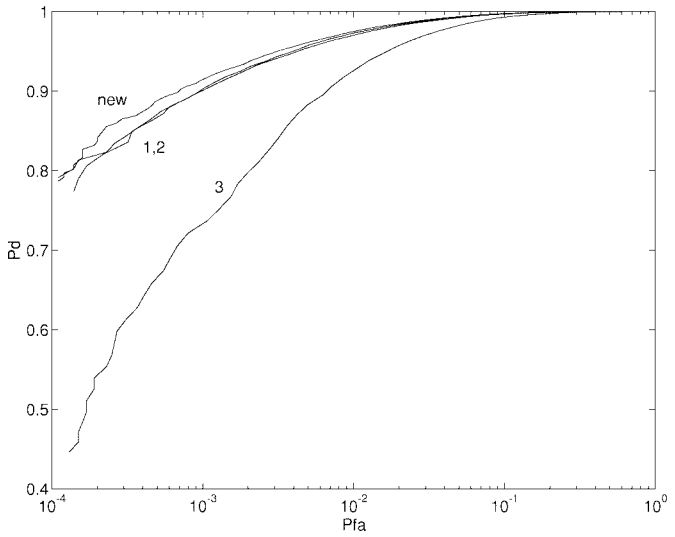

Fig. 8. Receiver operating characteristic for aggregate SNR of 60, $N = 128$ FFT bins, and $M = 50$ signal-containing bins. This is the "multiple-level" case, corresponding to (18). Random variates are exponential with means $\mu_{0,0} = 1$ under H_0; and either $\mu_{1,0} = 1$ ($N - M = 78$ of them) or $\mu_{1,2} = 1 + 2\text{SNR}/3M = 1.8$ ($M/2 = 25$ of them) or $\mu_{1,2} = 1 + 2\text{SNR}/3M = 1.8$ ($M/2 = 25$ of them). The hybrid ("new") detector assumes $P = 3$ levels under $\boldsymbol{H}_1$.

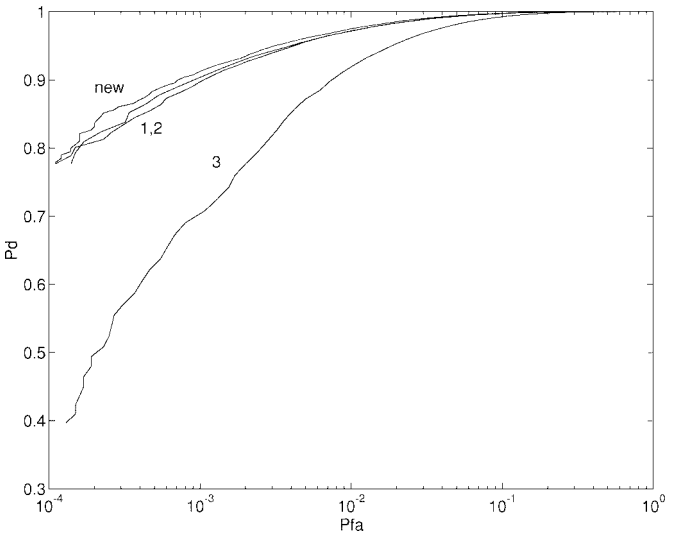

Fig. 9. Receiver operating characteristic for aggregate SNR of 60, $N = 128$ FFT bins, and $M = 50$ signal-containing bins. All signal containing bins have equal means. The hybrid detector assumes $P = 3$ levels under $\boldsymbol{H}_1$, that is, the detector is the same as for Fig. 8, whereas the underlying situation is the same as for Fig. 7.

the 100 000 simulations are carried out using $N = 128$ exponentially-distributed random variables $\{X_i\}_{i=0}^{127}$, which under H_0 have unit mean. Under H_1, $N-M$ ($128-50=78$, in this case) of the X_i's have mean unity, whereas 25 have mean 1.8, and 25 have mean 2.6—the aggregate SNR is $60 = 25 \times (1.8-1) + 25 \times (2.6-1)$. The flexibility of the hybrid detector allows it, in this case, a performance advantage over all power-law detectors, albeit a small one.

This added flexibility might be thought a burden when unnecessary. However, Fig. 9, in which the ground truth is that only one signal-containing population exists ($P = 2$), which is the same situation as that of Fig. 7, shows a similar picture. Both of these figures show the case of $M = 50$ signal-containing bins; other plots that are similar to Fig. 6 show that the hybrid detector with $P = 3$ levels has performance comparable with the power-law detector with $\nu = 2$ when there are $M = 5$ signal-containing bins.

C. Extension to CFAR Operation

Both the power-law detector and our new "known-$\mu_{0,0}$" hybrid detector are open to criticism in their avoidance of normalization issues. The new "estimated-$\mu_{0,0}$" detector is naturally constant false-alarm rate (CFAR); in this subsection, we compare this to the power-law detector with a cell-averaging (CA) CFAR preprocessor. (There are other proposed power-law processors with CFAR capability [17].)

With Y_{in} denoting the actual (exponentially distributed) magnitude-square ith bin value of the nth FFT, the CA-CFAR approach is to preprocess to

$$X_i = \frac{Y_{\text{in}}}{\dfrac{1}{n_w} \displaystyle\sum_{m=n-n_w}^{n-1} Y_{im}} \tag{21}$$

such that each bin, regardless of its actual signal-absent power level, is normalized to a unity parameter. Specifically, we modify the last element of (12) to

$$\{X_i\}: \quad \begin{cases} \text{independent } \mathcal{CA}(X_i; \mu_{0,0}), & H_0 \\ \text{independent } \mathcal{CA}(X_i; \mu_{1,Z_i}), & H_1 \end{cases} \tag{22}$$

in which $\mathcal{CA}(X, \mu)$ means

$$\Pr(X > x) = \frac{1}{(1 + x/(n_w \mu))^{n_w}}. \tag{23}$$

This is the input to the power-law detector. It would be interesting to compare this with the "correct" hybrid detector—that which estimates $\mu_{1,1}$ according to (22) and (23). However, (23) does not result in an explicit estimator for μ and since it is practically unappealing to nest an iterative zero-finding routine within another iterative EM procedure, we discuss only the new detectors developed in Section II-B.

For simulation, we use, as in Fig. 6, $M = 5$ (out of $N = 128$) signal-containing bins under H_1, an aggregate SNR $= 40$, and 100 000 runs. Under both hypotheses, the variates $\{X_i\}_{i=0}^{127}$ are independent and distributed according to (23) with $n_w = 8$; under H_0, all have $\mu = 1$, whereas under H_1, $M = 5$ have $\mu = 1 + \text{SNR}/M = 9$, with the remaining $N - M = 123$ having $\mu = 1$. In Fig. 10, we show the results. It is clear that the power-law detector, regardless of exponent ν, suffers a considerable performance degradation due to the heavy tailed nature of the CA-CFAR distribution of (23). The new detector, which must estimate $\mu_{0,0}$ ("est"), shows a significant performance advantage, this being somewhat surprising as its data window is of unity length rather than $n_w + 1$ as for the power-law detector. The new "known" detector has performance similar to the power-law detector with $\nu = 2$, but considering that no attempt has here been made to account for the CFAR statistic situation, this is not surprising.

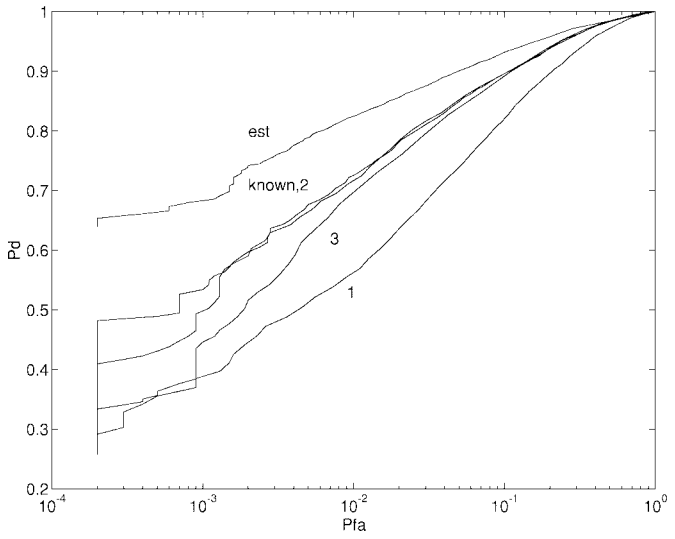

Fig. 10. Receiver operating characteristic for aggregate SNR of 40, $N = 128$ FFT bins, and $M = 5$ signal-containing bins. The situation is CFAR, with a window of size $n_w = 8$ for the power-law detector. The new detector "est" is that which estimates $\mu_{0,0}$ based only on the current FFT snapshot, and the "known" detector uses the same data as the power-law with an (incorrect) assumption of an exponential distribution.

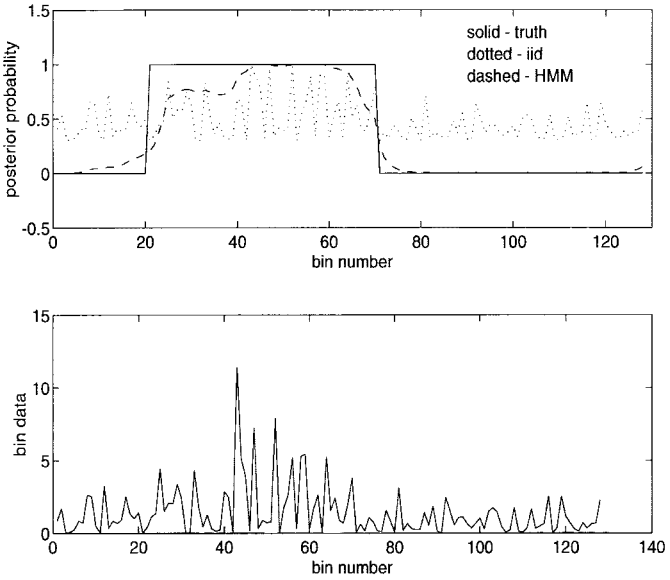

Fig. 11. Lower plot shows a transient with $M = 50$ beginning at bin 21 (see "truth" in the upper plot) having aggregate SNR $= 60$. The upper plot compares the posterior probabilities of a bin being signal-containing under the HMM and iid assumptions.

D. Extension to Exploit Signal Contiguity

Neither the power law nor the new hybrid detector, as described in Section II-B, assume any structure to the signal-containing bins. This is particularly clear from the model of (12), in which under H_1, the on/off signal process is allotted a Bernoulli character. In certain applications, however, due to the tonal—or at least bandpass—nature of many short-duration signals, it is reasonable to expect signal containing bins to appear in simply connected "clumps." Our goal in this section is to determine whether advantage can be taken.

It seems most appropriate therefore to model the binary process $\{Z_i\}$ of model (12) as (hidden) Markov. That is, (12) is replaced by

$$\begin{aligned} p(\{X_i\}_{i=1}^N, \{Z_i\}_{i=1}^N) &= q(Z_1)\frac{1}{\mu_{1,Z_1}} e^{-X_1/\mu_{1,Z_1}} \prod_{i=2}^{N} \\ &\quad \cdot \left[A(Z_i|Z_{i-1}) \frac{1}{\mu_{1,Z_i}} e^{-X_i/\mu_{1,Z_i}} \right] \end{aligned} \tag{24}$$

under H_1 and remains unchanged under H_0. The associated parameter set θ is thus augmented under H_1 to $q(0)(= 1 - q(1))$, which is the *a priori* probability that $Z_1 = 0$; the 2×2 transition matrix A, in which the (i, j)th element is the conditional probability that $Z_i = i$ given $Z_{i-1} = j$; and $\{\mu_{1,0}, \mu_{1,1}\}$, as before.

Model (24) appears complicated, and it is certainly more numerically challenging than that of Section II-B. However, in the HMM case, the EM procedure devolves to the well-known forward-backward (Baum–Welch) algorithm [26]. This elegant scheme solves simultaneously for all elements of θ; we do not give its details here. The idea is illustrated in Fig. 11, in which, for a particular data sequence, the posterior probabilities $\Pr(Z_i = 1|\{X_j\}_{j=1}^N)$ are compared under the current HMM and under that of Section II-B (iid). It is clear that the "sticky" HMM emphasizes the transient more effectively.

The simulations are essentially the same as those used for Figs. 6 and 7, except that the number of runs is reduced to 10 000; the Baum–Welch routine is more computationally intense than the original EM. (In fact, the complexity of the sum in the "optimal" test is considerably reduced here versus the unstructured situation of (9) since only those sets $\mathcal{S}$ containing contiguous subsets of length M need to be considered. In principle, all possible $M \in \{1, N\}$ must be used but, depending on N the numerical load, may be competitive with our Baum–Welch approach. A further difference is that since the absolute location of the signal-containing bins *does* matter here, we have under H_1

$$\begin{aligned} dF_\theta(X) = &\left[\prod_{i=0}^{m-1} \frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} \right] \\ &\cdot \left[\prod_{i=m}^{m+M-1} \frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}} \right] \\ &\cdot \left[\prod_{i=M+m}^{N} \frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} \right] \end{aligned} \tag{25}$$

in which m is drawn uniformly from $\{1, N-M\}$. The results are shown in Figs. 12 and 13 for numbers of (contiguous) signal-containing bins $M = 5$ and $M = 50$, respectively. The forward–backward routines in both cases use the knowledge that $\mu_{1,0} = 1$ and are, hence, comparable with the "known" characteristics in previous plots. It is apparent that use of signal contiguity enhances performance.

E. Time-Domain Transients

The previous results have dealt solely with transient changes, as observed in the frequency domain, that is, the

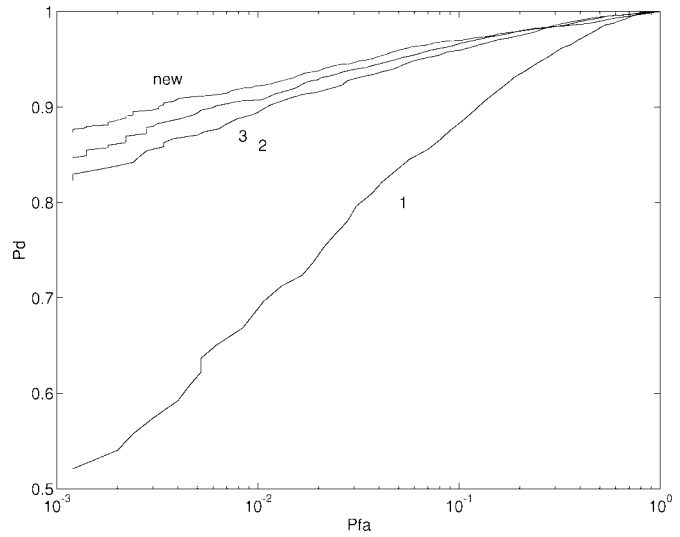

Fig. 12. Receiver operating characteristic for $M = 5$ contiguous signal-containing bins and aggregate SNR $= 40$. The "new" detector employs HMM estimation via the Baum–Welch procedure. The situation is similar to that of Fig. 6, except that the location of the first of the $M = 5$ contiguous signal-containing bins is drawn uniformly each run from $\{0, 122\}$.

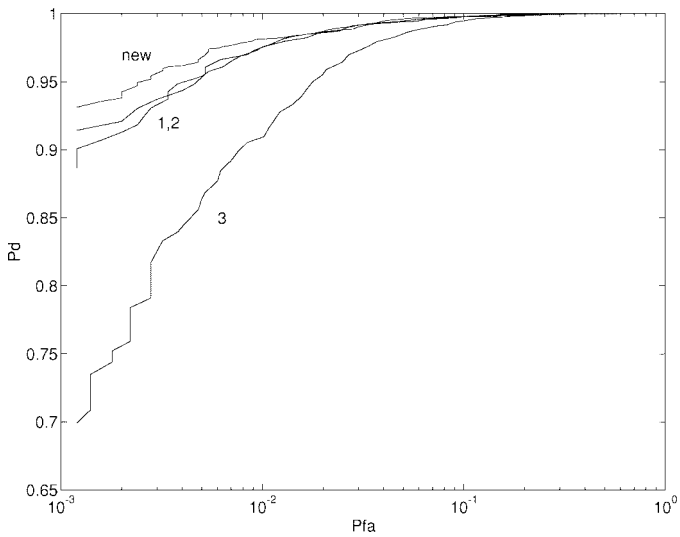

Fig. 13. Receiver operating characteristic for $M = 50$ contiguous signal-containing bins and aggregate SNR $= 60$. The "new" detector employs HMM estimation via the Baum–Welch procedure. The situation is similar to that of Fig. 6, except that the location of the first of the $M = 50$ contiguous signal-containing bins is drawn uniformly each run from $\{0, 77\}$.

random variables generated for the simulations were in fact exponentially distributed, according to the model. In this subsection, we generate transient signals in the time domain and use the FFT to transform these to the frequency domain. In all cases, the data record (and FFT) length is $N = 128$, the signals are complex, and 10^4 simulation runs are used. Signal-to-noise ratios are adjusted to give receiver operating characteristics that are nontrivial.

We first investigate the relationship between the length of the transient signal and the effective "bandwidth" observed in the frequency domain. In Fig. 14, we show these results for a randomly located white complex Gaussian transient with length n_d varying from 1–100, with pointwise signal-to-noise

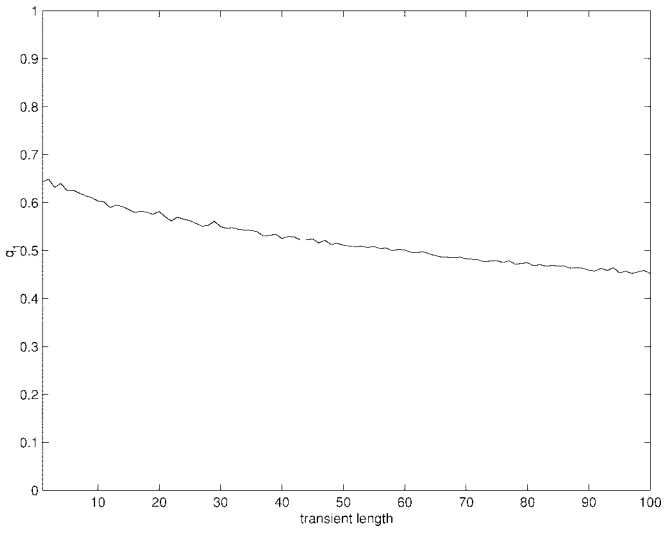

Fig. 14. EM-estimated probability q_1 that a frequency-domain sample is from the larger-amplitude population, from simulation, plotted as a function of transient length in the *time* domain. The time domain transient signal is complex, white, and Gaussian.

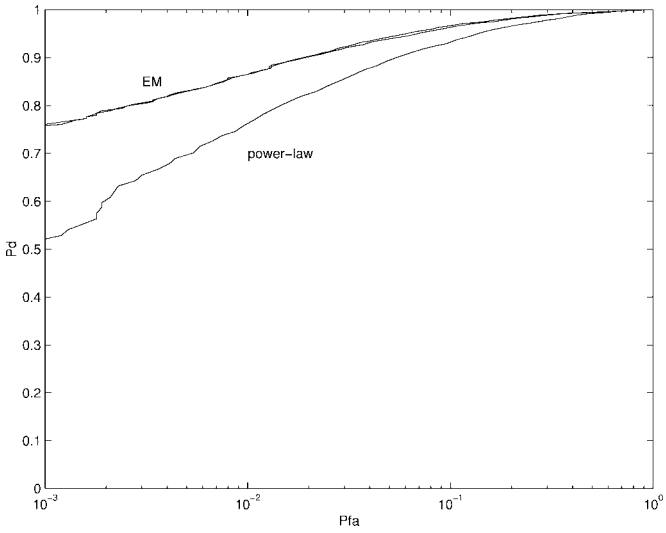

Fig. 15. Receiver operating characteristic for a complex white-Gaussian time-domain transient signal of length 5. We show the EM procedure with $P = 4$ frequency-domain levels and "basic-assumption" EM procedure with $P = 2$ frequency-domain levels and power-law processor with $\nu = 2.5$.

ratio (meaning the ratio of signal and noise variances for a given sample in which the transient is "on") of $16/n_d$. In accordance with our model, we have chosen to represent bandwidth by the estimated q_1, where the binomial probability of a frequency-domain sample is in the larger amplitude population. What is particularly notable is that this measure of bandwidth indicates that a transient signal, regardless of its length, appears to split the frequency-domain data into two populations of comparable size. Since the EM procedure tends to work best for populations that are not too asymmetric, this is promising.

In Fig. 15, we show the results of applying the detectors for a time-domain transient signal of length 5; the pointwise SNR is 8 dB. Two hyperparameter-estimation procedures are used: one is that corresponding to the "basic assumptions" (i.e., two

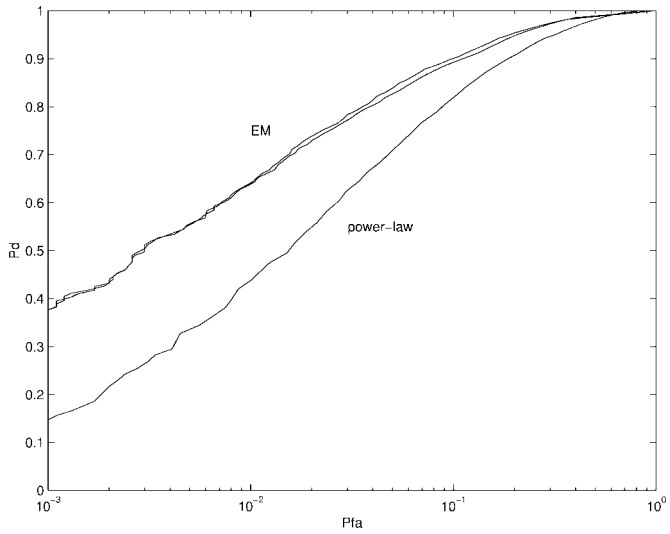

Fig. 16. Receiver operating characteristic for two complex white-Gaussian time-domain transient signals of length 20. An example is plotted in Fig. 17. We show the EM procedure with $P = 4$ frequency-domain levels, the "basic-assumption" EM procedure with $P = 2$ frequency-domain levels and power-law processor with $\nu = 2.5$.

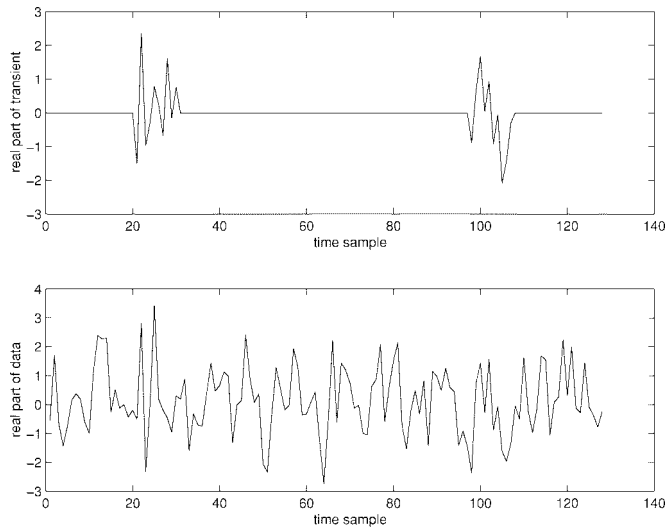

Fig. 17. Real part of time-domain transient signal consisting of two length-20 parts. Top: transient without noise; Bottom: transient with additive noise. The imaginary data is not shown.

levels) and one with $P = 4$ levels. These are compared with the power-law detector with exponent $\nu = 2.5$. It is clear that there is little difference between the two hyperparameter-estimation detectors but that there is a significant improvement in these with respect to the power-law detector. As noted above, the effective "bandwidth" of a time-domain transient is particularly favorable to the hyperparameter-estimation approach.

The experiment is repeated in Fig. 16 for a time domain signal consisting of two transient bursts; an example time-domain signal is pictured in Fig. 17. Again, the hyperparameter-estimation detector shows significant improvement with respect to the power-law detector. A further simulation involving a decaying complex-exponential transient in shown in Fig. 18; in this case, there is little difference between the schemes, most likely due to the aggregation of transient-signal energy into a relatively few frequency bins; as indicated above, the EM procedure works best when there is less asymmetry between the population sizes.

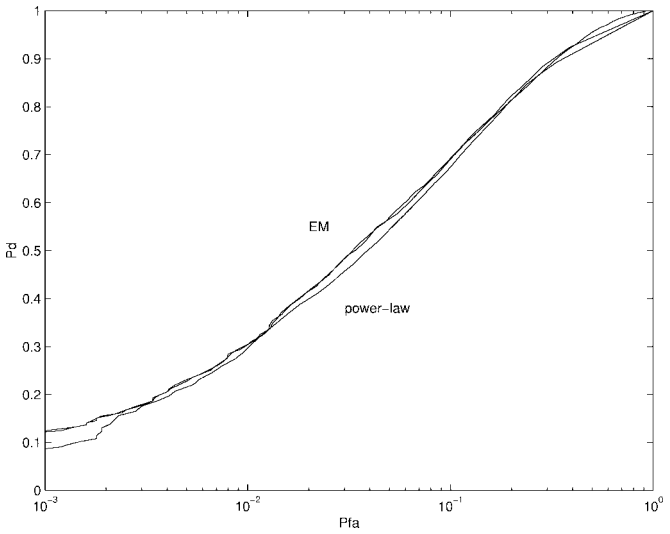

Fig. 18. Receiver operating characteristic for a decaying complex-exponential transient signal. We show the EM procedure with $P = 4$ frequency-domain levels and "basic-assumption" EM procedure with $P = 2$ frequency-domain levels and power-law processor with $\nu = 2.5$.

IV. Summary

The discrete Fourier transform of a block of data containing only white Gaussian noise gives rise to a uniform set of exponential variates, whereas an additive short-duration signal has the effect of destroying this uniformity. One of the best known ways to recognize such a deviation from uniformity is the power-law detector, particularly if the power ν is chosen with care to match an assumed transient signal bandwidth. In this paper, we have pursued a different approach: a GLRT whose estimation is over a set of *hyper*parameters describing the nonuniformity and whose computation relies on the EM algorithm. In its original formulation these hyperparameters dealt with an assumed underlying Bernoulli signal/no-signal model, but we have extended it to incorporate multiple signal levels, CFAR operation, and a hidden-Markov signal-contiguity model.

We have demonstrated that our approach has performance comparable with and often exceeding that of the best power-law detector.

Appendix
Development of the EM Procedure

According to the EM framework [19]–[22], locally maximum-likelihood estimation of a parameter θ based on observations X and with "hidden" random variates Z can be accomplished as the iteration of the following two steps.

E-Step: Form the ***e***xpectation

$$Q(\theta;\theta') = \int_Z \log\left(p_\theta(X,Z)\right) p_{\theta'}(Z|X). \tag{26}$$

M-Step: Then, *maximize*

$$\theta^* = \arg\max_{\theta} \{Q(\theta;\theta')\}. \tag{27}$$

Between each iteration, θ' is replaced by θ^*. The algorithm is guaranteed to converge at least to a *local* maximum of the likelihood surface $p_\theta(X)$, provided $p_\theta(X)$ is bounded above [22]. Its initial-convergence properties are good, although refinement to high accuracy can be slow; fortunately, in the current application, precise estimation accuracy is seldom important.

In our case, under H_1 and where the parameters are $\theta = \{q, \mu_{1,0}, \mu_{1,1}\}$, we have

$$\begin{aligned} p_\theta(X,Z) &= \prod_{i=1}^{N} dF_\theta(X_i|Z_i)\, dG_\theta(Z_i) \\ &= \prod_{i=1}^{N} \left(\frac{1}{\mu_{1,Z_i}} e^{-X_i/\mu_{1,Z_i}}\right)(1-q+(2q-1)Z_i) \end{aligned} \tag{28}$$

where $Z_i \in \{0,1\}$. From this, we can derive

$$\begin{aligned} p_\theta(X) &= \int_Z p_\theta(X,Z) \\ &= \prod_{i=1}^{N} \left((1-q)\frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} + q\frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}}\right). \end{aligned} \tag{29}$$

As a consequence, we can write

$$p_\theta(Z|X) = \prod_{i=1}^{N} w_i(Z_i) \tag{30}$$

in which

$$\begin{aligned} w_i(0) &= \frac{(1-q)\frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}}}{(1-q)\frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} + q\frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}}} \\ w_i(1) &= \frac{q\frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}}}{(1-q)\frac{1}{\mu_{1,0}} e^{-X_i/\mu_{1,0}} + q\frac{1}{\mu_{1,1}} e^{-X_i/\mu_{1,1}}} \end{aligned} \tag{31}$$

describe the posterior probabilities of Z_i. We thus write

$$\begin{aligned} Q(\theta;\theta') &= \sum_{Z} \left(\sum_{i=1}^{N} \left[-\frac{X_i}{\mu_{1,Z_i}} - \log(\mu_{1,Z_i}) \right.\right. \\ &\qquad \left.\left. + \log((1-q)+(2q-1)Z_i) \right] \prod_{j=1}^{N} w_j(Z_j) \right) \\ &= \sum_{i=1}^{N} \left(\left[-\frac{X_i}{\mu_{1,0}} - \log(\mu_{1,0}) + \log(1-q) \right] w_i(0) \right. \\ &\qquad \left. + \left[-\frac{X_i}{\mu_{1,1}} - \log(\mu_{1,1}) + \log(q) \right] w_i(1) \right) \end{aligned} \tag{32}$$

where $w_i(0)$ and $w_i(1)$ are calculated using θ', that is, using the previous parameter estimates. What is fortunate, and is usually the case when the EM algorithm is to be applied successfully, is that $Q(\theta,\theta')$ can be maximized (over θ) without necessity of its formation. Specifically, we have

$$\begin{aligned} \mu_{1,0}^* &= \frac{\sum_{i=1}^{N} w_i(0) X_i}{\sum_{i=1}^{N} w_i(0)} \qquad \mu_{1,1}^* = \frac{\sum_{i=1}^{N} w_i(1) X_i}{\sum_{i=1}^{N} w_i(1)} \\ q^* &= \frac{1}{N} \sum_{i=1}^{N} w_i(1). \end{aligned} \tag{33}$$

Thus, the iteration is straightforward and simple.

1) Initialize $\{\mu'_{1,0}, \mu'_{1,1}, q'\}$ to some nonextreme values.
2) Compute $\{w_i(0)\}$ from (31) and the current parameter estimates $\{\mu'_{1,0}, \mu'_{1,1}, q'\}$; note that $w_i(1) = 1 - w_i(0)$.
3) Update $\{\mu^*_{1,0}, \mu^*_{1,1}, q^*\}$ according to (33).
4) Substitute the new values $\{\mu^*_{1,0}, \mu^*_{1,1}, q^*\}$ into $\{\mu'_{1,0}, \mu'_{1,1}, q'\}$.
5) Check a stopping criterion; if not met, return to 2).

Under H_0, due to the structure of Z, there is no need for the EM procedure, and straightforward estimates of $\theta = \{q, \mu_{0,0}\}$ are available.

REFERENCES

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] B. Broder and S. Schwartz, *Quickest Detection Procedures and Transient Signal Detection*, ONR Tech. Rep. 21, Nov. 1990.

[3] C. Han, P. Willett, and D. Abraham, "Some methods to evaluate the performance of Page's test as used to detect transient signals," *IEEE Trans. Signal Processing*, to be published.

[4] B. Porat and B. Friedlander, "Performance analysis of a class of transient detection algorithms—A unified framework," *IEEE Trans. Signal Processing*, vol. 40, pp. 2536–2546, Oct. 1992.

[5] M. Frisch and H. Messer, "Transient signal detection using prior information in the likelihood ratio test," *IEEE Trans. Signal Processing*, vol. 41, pp. 2177–2192, June 1993.

[6] S. Demarco and J. Weiss, "Improved transient signal detection using a wavepacket-based detector with an extended translation-invariant wavelet transform," *IEEE Trans. Signal Processing*, vol. 45, pp. 841–850, Apr. 1997.

[7] M. Frisch and H. Messer, "The use of the wavelet transform in the detection of an unknown transient signal," *IEEE Trans. Inform. Theory*, vol. 38, pp. 892–897, 1992.

[8] N. Lee and S. Schwartz, "Robust transient signal detection using the oversampled Gabor representation," *IEEE Trans. Signal Processing*, vol. 43, pp. 1498–1502, June 1995.

[9] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 169–180, Feb. 1989.

[10] J. Fonollas and C. Nikias, "Wigner higher-moment spectra—Definition, properties, computation and application to transient signal analysis," *IEEE Trans. Signal Processing*, vol. 41, pp. 245–266, Jan. 1993.

[11] M. Hinich, "Detection of a transient signal by bispectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1277–1283, July 1990.

[12] M. Bartlett, K. Baugh, and G. Wilson, "Transient detection using the nonstationary bispectrum," *J. Acoust. Soc. Amer.*, vol. 99, no. 5, pp. 3018–3028, 1996.

[13] K. Hardwicke, G. Wilson, and K. Baugh, "Characterization of spectral correlation detector statistics useful in transient detection," *Circuits, Syst., Signal Process.*, vol. 13, no. 4, pp. 497–511, 1994.
[14] L. Perlovsky, "A model-based neural network for transient signal processing," *Neural Networks*, vol. 7, no. 3, pp. 565–572, 1994.
[15] A. Nuttall, "Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength," NUWC-NPT Tech. Rep. 10 751, Sept. 1994.
[16] ———, "Near-optimum detection performance of power-law processors for random signals of unknown location, structure, extent, and arbitrary strengths," NUWC-NPT Tech. Rep. 11 123, Apr. 1996.
[17] ———, "Performance of power-law processor with normalization for random signals of unknown structure," NUWC-NPT Tech. Rep. 10 760, May 1997.
[18] J. Fawcett and B. Maranda, "The optimal power law for the detection of a Gaussian burst in a background of Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 37, pp. 209–214, 1991.
[19] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, pp. 1–88, 1977.
[20] R. Redner and H. Walker, "Mixture densities, maximum likelihood, and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, 1984.
[21] R. Redner, R. Hathaway and J. Bezdek, "Estimating the parameters of mixture models with modal estimators," *Commun. Stat., Part A: Theory Methods*, vol. 16, no. 9, pp. 2639–2660, 1987.
[22] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, pp. 95–103, 1983.
[23] R. Streit, "Combined Bayes/GLRT detection," seminar presented at NUWC, New London, CT, June 1996.
[24] R. Streit and P. Willett, "Hybrid Bayes/GLRT detection of transient signals," in *Proc. CISS*, Baltimore, MD, Mar. 1997.
[25] B. Chen, P. Willett, and R. Streit, "Improved Bayes/GLRT transient detection," in *Proc. CISS*, Princeton, NJ, Mar. 1998.
[26] B. Juang and L. Rabiner, "An introduction to hidden Markov models," *IEEE Acoust., Speech, Signal Processing Mag.*, Jan. 1986.

Roy L. Streit received the B.S. degree with honors in physics and mathematics from East Texas State University, Commerce, the M.A. degree in mathematics in 1970 from the University of Missouri, Columbia, and the Ph.D. degree in mathematics from the University of Rhode Island, Kingston.

From 1981 to 1982, he was a Visiting Scholar with the Department of Operations Research, Stanford University, Stanford, CA, and from 1987 to 1989, he was an Exchange Scientist with the Defence Science and Technology Organization, Canberra, Australia. He joined the Navy Underwater Sound Laboratory, New London, CT, in 1970 and is now with the Naval Undersea Warfare Center, Newport, RI. His research has included the application of hidden Markov models to frequency line detection and tracking, diverse investigations in support of several NUWC towed array R&D programs, the development of maximum likelihood training algorithms for probabilistic neural networks, and the design of algorithms for complex FIR filter design using linear programming methods. His current interests include passive-sonar localization using low-fidelity ray trace models and probabilistic multihypothesis tracking (PMHT).

Peter K. Willett (SM'97) received the B.A.Sc. degree from the University of Toronto, Toronto, Ont., Canada, in 1982 and the Ph.D. degree from Princeton University, Princeton, NJ, in 1986.

He is a Professor at the University of Connecticut, Storrs, where he has worked since 1986. His interests are generally in the areas of detection theory and signal processing.