Algorithms For Joint Channel Estimation and Data Recovery—Application to Equalization in Underwater Communications

Meir Feder, *Member, IEEE,* and Josko A. Catipovic, *Member, IEEE*

Abstract—One of the main obstacles to reliable underwater acoustic communications is the relatively complex and unstable behavior of the ocean channel. The channel equalization method, that can estimate and track this complex and rapidly varying ocean response, may lead to reliable data communications at high rates which utilize fully the available bandwidth. Unfortunately, standardized equalization techniques fail in this environment. In this paper we derive methods for joint ocean-channel estimation and data recovery, using optimal, Maximum Likelihood (ML) estimation criterion. The resulting ML problems may be complex; thus we will use iterative algorithms; e.g., the Expectation-Maximization (EM) algorithm. The different methods correspond to different assumptions about the ocean channel. The theoretical derivation of these methods as well as preliminary results on a simulated ocean data experiments are presented.

I. INTRODUCTION

THE underwater acoustic channel is probably one of the most complicated environments for data communications. The unique characteristics of this channel, which include fading, extended multipath and refraction, fluctuation and unstable behavior, etc., preclude direct application of standard communication techniques. Past efforts to design a reliable underwater acoustic link largely by integrating methods developed for other channels are summarized in a few review papers (e.g., [1], [2]).

In most communications channels the limitation on the rate is largely due to the bandwidth or the signal-to-noise ratio (SNR). However, if we consider, for example, the short-range acoustic channel, despite the fact that its bandwidth is about 20 kHz and the fact that it has usually a reasonable SNR, the reliable data rate achieved in state-of-the-art underwater modems is about 1 kb/s. We note that in some other communications systems, bandwidth expansion figures to 4 to 6 (i.e., reliable data rates of 4 to 6 b per available bandwidth Hz) have been achieved!

As it seems now, the available underwater communications systems operate at rates far below the capacity of that channel. The recent developments in bandwidth-efficient modulation and coding techniques (e.g., Trellis coding, [3], [4]) have led to reliable high rates in telephone and satellite channels, which approach the capacity of these channels. In the ocean channel, similar methods are just in their initial stages (e.g., [5]) and their performance has not been tested extensively yet. While better modulation techniques will certainly improve the performance of

M. Feder is with the Department of Electrical Engineering—Systems, Faculty of Engineering, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel.

J. A. Catipovic is with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

underwater acoustic communications, the crucial factor, in attempts to achieve the capacity, will be the equalization or compensation of the fluctuating ocean channel behavior. Without channel equalization, not only will the modern and efficient trellis coding techniques break down, but other signaling techniques, especially those based on phase coherent detection, cannot operate reliably.

Equalization for channel response is an extensively studied problem. Summary and review of standard equalization techniques can be found in [6]–[8, chaps. 6 and 7]. In general, equalization techniques belong to one of the following two categories: The techniques of the first category try to cancel the effects of the channel by deconvolving the received signal; i.e., by passing it through a filter whose response is the inverse of the channel impulse response. These techniques, sometimes known as inverse or zero-forcing filters, [9], will minimize the peak distortion between the transmitted and equalized signal. However, as is the case with any inverse filtering technique, in the presence of even a small amount of additive noise the performance of these equalizers is poor, since they amplify the noise considerably; when the channel has spectral nulls, the SNR after equalization will go to zero! Thus despite some recent interesting developments in blind deconvolution (e.g., [10]), we will not consider the application of techniques from this category in reliable underwater equalization and communications systems.

The techniques of the second category use as a criterion the mean square error between the transmitted signal and equalized signal. This criterion leads to a "matched filter" structure of the equalizer. The various equalization methods differ in their assumptions about the signals, the channel, and about what is known *a priori* about the data and channel. For cases where the transmitted data or input signal is known, a variety of least-squares algorithms have been derived for the channel estimation; e.g., the gradient and LMS methods [11], and the Recursive Least Squares (RLS) or Kalman method [12], [13], and their implementation via lattice filters [14], [15], etc. The complementary problem of effective equalization (i.e., data recovery from the signal that has Inter-Symbol-Interference (ISI) due to the channel when the channel structure is given) was solved in [16] by the Maximum-Likelihood-Sequence-Estimation (MLSE) approach, implemented by using the Viterbi algorithm. Attempts for joint channel and data estimation were made by using the class of Decision Feedback Equalizers (DFE), [17]–[19]; incorporating DFE with MLSE has been performed in [20]. These techniques are the closest to our approach. As will be discussed in detail in Section II, the ocean acoustic channel is more complex and changes more rapidly than virtually any other communication channel; thus the standard equalization techniques, mentioned above, are inadequate in this environment.

The approach and directions we suggest for equalization in underwater acoustic communications are based on methods for joint channel estimation and data recovery. The estimation criterion will be the optimal Maximum-Likelihood (ML) objective function. This joint estimation will be performed under a variety of assumptions or mathematical models of the communication/equalization problem and ocean channel. The statistical ML problems resulting from the mathematical modeling are often complicated and usually do not have an analytic closed-form solution. Thus in most cases we have suggested iterative algorithms.

A common feature of the suggested iterative algorithms for joint channel estimation and data recovery is an alternation between channel estimation, assuming that the data is known, and data recovery, assuming that the channel is known. For example, given the channel impulse response which is equivalent to knowing the ISI structure, we can recover the information bits using the MLSE technique, implemented via a Viterbi algorithm, mentioned above. On the other hand, given the data, the channel impulse response can be estimated—either in closed form for linear cases, or iteratively for cases where the parameters of this response enters in a nonlinear way. One of the powerful algorithms for iterative ML is the Expectation-Maximization (EM) algorithm [21], [22]. This iterative method will be incorporated in one of our equalization algorithms.

The algorithms we suggest take into account the ocean channel behavior by operating on blocks, whose size is determined by the time window in which we can assume that the channel is stable. Thus we are ready to cope with situations in which, due to fading and fluctuations, the channel has been changed completely and abruptly between blocks. In further research we plan to derive adaptive versions of our algorithms which will track the channel variations in a nonblock fashion.

The paper is organized as follows: In the next section we will discuss in detail underwater ocean acoustic channels, the special equalization problems that characterize these channels, and the inadequacy of the standard equalization methods in this environment. In Section III we present in detail the mathematical models, formulate the statistical estimation problems, and present the background for our algorithm derivation. The suggested equalization methods are given in Section IV; these methods estimate jointly the channel response and data and provide the new approach and directions. These equalization algorithms are tested using simulated data, and these experimental results are presented in Section V. A summary and suggestions for further research direction will conclude the paper.

II. THE OCEAN CHANNEL

A. Relevant Channel Characteristics

Underwater acoustic communication is limited primarily by the dispersion and rapid time-variant behavior of the ocean channel, which is a complex waveguide with a number of physical parameters causing strong fluctuation of the received acoustic waveforms. The theory of wave propagation in random media has contributed a number of insights into underwater acoustic propagation, and excellent summary articles on long- and short-range acoustic fluctuations appear; e.g., [23], [24]. A distinct characteristic of the oceanic waveguide is the long and complex multipath structure. The relatively slow speed of sound in the water results in extended multipath structures from a given waveguide geometry and long reverberation times encountered.

In the short-range channel, the multipath is largely due to reflections from large scatterers and boundary interactions.

In many cases the channel impulse response $h(t)$ can be modeled as

$$h(t, f) \sim \sum_i a_i(t - T_i, f) e^{j\omega(t - T_i, f) + \theta_i(t, f)} \qquad (1)$$

and this response may, in general, be both time and frequency varying. Each path arrival $s(t, f) \equiv a_i(t - T_i, f) e^{j\omega(t - T_i, f) + \theta_i(t, f)}$ undergoes frequency-dependent amplitude, phase, and time-delay fluctuations. If these were known, the optimal communication receiver would subtract the channel-caused effects before processing the received waveforms. In most cases the fluctuations are unknown; the system can estimate them and use the estimates to improve received signal fidelity. Alternately, the system can estimate the stochastic properties of the fluctuations and operate in the presence of random fluctuations whose moments are known. Combinations of the two approaches are implemented, for instance, in partly coherent methods where phase fluctuations are treated as random, but amplitude and delay are estimated as deterministic quantities.

A number of solutions for tracking channel behavior is available, including phase, amplitude, and time-delay tracking [25]. In this work we concentrate on estimating the impulse response, assuming a linear channel model; e.g., equalizing the time delay and amplitude distribution of the ocean multipath. While phase fluctuation can also be compensated with the proposed method, in practice, coherent signaling over the ocean channel is quite difficult and generally implemented only for the vertical path [25]–[27].

The joint channel and data estimator finds, for example, the path amplitudes α_i and relative delay times τ_i along with the ML data sequence. The path amplitude fluctuations arise from single-path effects such as turbulence, surface, and internal wave fields as well as from multipath interference [25]. Both effects are well understood and the fluctuation spectra are available for the stationary source–receiver [23], [24]. Multipath delay distribution arises from the waveguide characteristics and scatterer locations within the waveguide [28]. For the short-range channels, multipath stability is determined largely by the distribution and motion of nearby scatterers.

We now examine, in some detail, the impulse response, especially the channel amplitude and multipath delay characteristics, of three underwater acoustic channels. These channels are a deep-water vertical path and two shallow-water examples illustrating the extremes of multipath delay distribution and path amplitude fluctuation.

1) Short-Range Vertical Channel: The short-range vertical channel used for communication with bottom instruments and vehicles exhibits rather mild amplitude and phase fluctuations, and a number of modulation and equalization methods developed for telephone channels are directly applicable. The received signal typically consists of a direct arrival and surface reflection. The second path is easily eliminated with a directional receiver, whose additional benefit is the rejection of surface-generated noise. At frequencies of 10–50 kHz typically used for communication over the vertical path, ambient noise is largely generated near the surface, particularly near a ship or other marine structure. A directional receiver hydrophone can frequently yield a 40-dB SNR improvement, and angular discrimination is often implemented in practice. Under these conditions, adaptive equalization of surface multipath is not worthwhile because of additional surface noise introduced into the receiver.

In this channel, sometimes, a bottom-induced multipath can arise from bottom-mounted or near-bottom transmitters. It is not separable on the surface with angle of arrival processing and can dominate the signal quality at times; for example, when exploring and photographing an underwater wreck. Again, multipath can be avoided with directional transducers, but this surface-station positioning can become a problem. Fortunately, the multipath in this case is limited to a few specular arrivals and the multipath dynamics are governed by geometric positioning between the bottom transmitter and reflecting object.

2) Short-Range Shallow Water Horizontal Channel: The 1–10-km shallow water waveguides are among the most dynamic multipath channels in use. From channel measurements made in relatively calm water, collected over a 3-km channel approximately 15-m deep, the following features were observed:

- A multipath channel in which the first arrivals show evidence of through-bottom propagation.
- Several independent amplitude and delay fluctuation mechanisms are seen in the data. Path appearance (emergence from a fade), discrete time-delay shift, path splitting, and interference of bottom and water-propagating arrivals are evident.
- This channel is not severely Doppler spread, but the fluctuation behavior of individual arrivals causes tracking difficulties for classical equalization methods.

We believe that these features primarily show the effects of large stationary scatterers and bottom/surface interaction in that channel.

The short-range shallow-water channel can show a totally different behavior. In measurements made over a 700-m range in Woods Hole Harbor in a highly turbulent and high current environment, we have seen the effects of mid-column turbulence and diffuse moving scatterers. In this case the validity of the discrete multipath model is questionable, as no discrete paths are evident in the data. Sometimes a low-frequency fluctuation of the primary arrivals can be caused by the dominant turbulence scale in the propagation path. Strong spatial dependence of the received acoustic field is observed under these circumstances [29]. The acoustic intensity distribution resembles the distribution of light intensity at the bottom of a disturbed swimming pool, and strong temporal behavior observed at a point is a result of the time-variant refraction properties of the medium. Under strong sound-focusing conditions, spatial diversity processing is a promising telemetry method.

B. Equalizer Performance Requirements for Ocean Acoustic Channels

A number of adaptive channel estimators and equalizers were successfully implemented on the ocean channel [30], [27], [31], [32]. However, the fluctuation rate of many ocean channels, particularly the shallow-water horizontal waveguide, precludes equalization and multipath processing with algorithms developed for more benign channels. Classical methods such as LMS and RLS adaptive equalizers typically require an initial training sequence and either periodic updates or decision feedback [7], [33]. An additional complication is the difficulty of synchronizing to a channel without a well-defined first or principal arrival. The resultant synchronizer jitter is exhibited as additional multipath fluctuation, further degrading equalizer performance.

The focusing behavior of ocean channels often requires spatial diversity receiver implementations [29]. The channel impulse response of each diversity path differs markedly, and a given

channel may have high data quality for only a few seconds. Independent path equalizers are required which can operate without training sequences in the presence of rapidly time-variant channels and relatively poor data quality. This work discusses an equalizer formulation for the ocean acoustic channels of this type. The algorithm formulation was driven by the following constraints not commonly found on other channels:

- The discrete multipath model is applicable in many situations of interest, but a significant percentage of the channels of interest cannot be modeled by discrete multipath arrivals.
- Channel fluctuation rates may approach the system baud rate; i.e., the channel can change significantly during a few frames.
- Channel fluctuations arising from energy focusing and scattering cause frequent signal fades and data degradations. The system must recover from fades without recourse to training sequences.
- The ocean acoustic channel bandwidth is severely constrained, and operation without dedicated channel probes is preferable.
- Computational complexity is not a severe constraint. The desired data rates over the 10-km shallow-water channel are on the order of 10 kb/s, and computational engines approaching 1 gigaflop are realizable.

These requirements and the shortcomings of present equalization methods for time-variant channels have motivated our work on joint data and channel estimators.

III. MATHEMATICAL MODELS AND BACKGROUND

In this section we present the common mathematical models for communication in ISI channels and formulate the problem of joint channel estimation and data recovery as a statistical composite hypothesis problem. We then describe how to solve each part of the problem separately; i.e., methods for data recovery when the channel ISI structure is known, and methods for channel estimation when the data is known. The suggested algorithms, which will be described in the following section, will iterate between the partial solutions of the original problem. Thus we will also describe briefly some aspects of iterative algorithms for Maximum Likelihood (ML) estimation, and, more specifically, the Expectation-Maximization (EM) algorithm.

A. Problem Description

The communication problem in any channel characterized by inter-symbol interference (ISI) is characterized, mathematically, as follows: Let the transmitted symbols be denoted $\underline{a}$; for the case where n symbols were transmitted, $\underline{a} = a_0 \cdots a_{n-1}$. The modulation procedure generates a signal, $s(t; \underline{a})$:

$$s(t; \underline{a}) = \text{Re}\left\{ \sum_i s_p(t - iT; a_i) e^{j\omega_c t} \right\} \quad (2)$$

where ω_c is the carrier frequency, and $s_p(t; a)$ is a complex pulse signal; i.e., it is equal to zero outside the interval $[0, T]$, whose shape depends on the modulation. For signaling techniques such as PSK, QAM, etc., this pulse signal has a constant complex value (that depends on a_i) throughout the pulse period, while it is $e^{ja_i \cdot \Delta \omega t}$ for FSK modulation. In the analysis, throughout the rest of the paper we will consider only the complex demodulated signals; for example, the transmitted signal will be $\tilde{s}(t; \underline{a}) = \sum_i s_p(t - iT; a_i)$.

The modulated signals enter a channel whose impulse response is $h(t)$. This is the effective "base-band" channel response. Note that we assume that the channel is linear and time-invariant. As discussed above, this assumption can be valid in the ocean only for some observation window of about half a second; this observation window will be, say, $[0, nT]$ if n symbols can be transmitted while the channel is stable. We also assume that nonlinear effects due to demodulation, Doppler shift, etc., have been compensated. The impulse response $h(t)$ is assumed to be causal and of finite duration; i.e., $h(t) = 0$ for $t > D$, where D is usually of several $(2 \div 5)$ symbol periods. Now the observed signal will be

$$r(t) = \tilde{s}(t; \underline{a}) * h(t) + n(t) \qquad (3)$$

where $n(t)$ is assumed to be a white (in the channel bandwidth) Gaussian signal. Clearly, due to the channel, the observed signal at each time point is composed of contributions from the previous symbols—the ISI problem.

The task we encounter is to recover the information bits from the observed signal despite the channel ISI effects. It is well known, and will be further described below, that if we know the channel impulse response, we can find the best (in terms of minimum probability of error) information bit sequence. On the other hand, if a sequence of information bits is known, we face a channel-estimation problem, where the input to the channel is given. The solution for this problem, under the various assumptions, will also be further described below. Sometimes known preamble sequences are used, especially for this task; however, in the ocean environment where the channel is varying rapidly, we may waste our available bandwidth if we transmit known information bits for estimating the channel in the rate that the channel is varying. Thus the algorithms we propose in this paper will jointly estimate the channel response and information bits.

In a more formal setting the problem we encounter is as follows: Assuming that the *a priori* probability of the information bits is uniform, and due to the fact that the noise is white and Gaussian, we are looking for the symbol sequence and the channel that solve:

$$\min_{\underline{a}, h(\cdot)} \int |r(t) - \tilde{s}(t; \underline{a}) * h(t)|^2 \, dt \qquad (4)$$

where the integration is over the time window for which the channel is stable, and $\tilde{s}(t; \underline{a}) * h(t) = \int_0^D h(\tau)\tilde{s}(t - \tau; \underline{a}) \, d\tau$. Note that for a fixed $h(\cdot)$ the minimization above is equivalent to the Likelihood Ration Test (LRT) for the information bits, and since we minimize over $h(\cdot)$, we perform the Generalized LRT.

B. Extraction of the Information Symbols in a Known ISI Channel

We will now briefly present the solution to the information bits extraction problem, given the channel impulse response; i.e., the LRT solution. Within this solution we will find out, unsurprisingly, that there are sufficient statistics extracted at the rate of the information symbols; i.e., every T seconds; thus we do not need to process the entire continuous-time received signal. This set of measurements is the output of a whitened matched filter, sampled at the symbol rate. Unfortunately, for estimating both the channel and data this set of measurements is not sufficient, and we have to process and keep the entire observed signal $r(t)$.

The goal function $M(\underline{a})$ for estimating the information bits is

(4), which is a function only of the information symbols by the assumption that the channel is known. Now, since $\tilde{s}(t; \underline{a}) = \sum_i s_p(t - iT; a_i)$, we get:

$$M(\underline{a}) = \int \left| r(t) - \sum_i \int_{iT}^{(i+1)T} s_p(\tau - iT; a_i) h(t - \tau) \, d\tau \right|^2 dt$$

$$= \int |r(t)|^2 \, dt - 2 \operatorname{Re} \left\{ \sum_i \int_{iT}^{(i+1)T} s_p^*(\tau - iT; a_i) \right.$$

$$\cdot \left(\int r(t) h^*(t - \tau) \, dt \right) d\tau \right\}$$

$$+ \sum_i \sum_j \int_{iT}^{(i+1)T} s_p^*(\tau - iT; a_i)$$

$$\cdot \int_{jT}^{(j+1)T} s_p(\tau' - jT; a_j)$$

$$\cdot \left(\int h^*(t - \tau) h(t - \tau') \, dt \right) d\tau \, d\tau' \qquad (5)$$

and we have to minimize (5) with respect to $\underline{a}$.

Looking on (5), the integral of $|r(t)|^2$ is independent of the information sequence and the third term does not include the observed signal. From the second term above we notice that we have to pass the observed signal $r(t)$ through a filter whose impulse response is $h^*(t)$ (i.e., a filter matched to the channel impulse response) and then to integrate it independently, every symbol period, against the various possible pulse signals. Thus if the symbol alpha-bet size is A, we get A numbers for each symbol duration, and these numbers form the set of sufficient statistics for the extraction of the information sequence. Note that for PAM, PSK, or QAM modulation where $s_p(t; a)$ has a constant value a throughout the symbol duration, the sufficient statistics are composed of a single value for each symbol period.

It is easy to see that the third term, which depends on the information sequence, although it is independent of the observations and thus can be calculated in advance, depends only on the autocorrelation of the channel impulse response; i.e., on:

$$R(\sigma) = \int h(t) h(t - \sigma) \, dt$$

and by our assumptions $R(\sigma) = 0$ for $|\sigma| > D$. The minimization of (5), which in principle depends on the entire data sequence, would have required an exponentially complex exhaustive search over all possible sequences. The fact that the channel impulse response autocorrelation function is nonzero only over the interval $[-D, D]$, which implies that locally the goal function depends only on a few symbols, has led to a more efficient computational method based on the dynamic-programming Viterbi algorithm. This algorithm, proposed in [16], is described in the version we use later in the paper, in Appendix A.

The analysis of the performance in terms of bit error probability when the channel is known and the Viterbi algorithm for equalization is used can be found in [8, chap. 6], following [16]. This analysis is similar to the performance analysis done for convolutional codes; indeed, the ISI effect is sometimes described as a channel-induced convolutional coding whose rate is 1. We note, however, that despite this "coding," the bit error performance when ISI exists is usually poorer then when no ISI exists, even when the ISI structure is known, since the channel

destroys the optimal design characteristics such as orthogonality that the original transmitted pulse signals may have had.

C. Maximum-Likelihood Channel Estimation – Known Data

In several realistic communication situations the transmitted information bits may be available to the receiver, at least for some portion of the communication session. For example, the transmitter and receiver may agree on a known bit sequence to be transmitted every predefined time period, or they may set a special probe channel in which known data sequences will be transmitted. In both cases, the goal is to use the known information symbols which define the input to the channel to estimate the channel characteristics. In the channel equalizer context, often the known data sequence can be the information symbols just extracted, as in decision-directed and decision-feedback equalization, or, as is the case in our alternating algorithm, the known sequence is the data extracted in the previous iteration and used for better estimating the channel response.

Under our assumption that the additive noise is white and Gaussian, the ML estimation of the impulse response minimizes the goal function given by (4), which is a function of the channel-impulse response, since the data sequence $\underline{a}$ is assumed to be known. Explicitly, for known data we have to find the minimum of

$$L(h(\cdot)) = \int_t \left| r(t) - \int_{\tau=0}^{D} h(\tau)\tilde{s}(t - \tau; \underline{a}) \, d\tau \right|^2 dt. \quad (6)$$

1) *No Constraints Case:* If no constraints are imposed on $h(\cdot)$, the minimization problem above is a linear least-squares problem. The solution $\hat{h}(\cdot)$ must satisfy the equation:

$$\int_{\tau=0}^{D} \hat{h}(\tau) \left(\int_t \tilde{s}(t - \tau)\tilde{s}^*(t - \sigma) \, dt \right) d\tau$$
$$= \int_t r(t)\tilde{s}^*(t - \sigma) \, dt \quad (7)$$

where we have omitted the dependency on $\underline{a}$, since it is assumed to be fixed and known. The explicit solution of (6) for $\hat{h}(\cdot)$ can be given using the reciprocal Kernel formulation (see [34, chap. 4]).

For clarity and for further derivations, we will present the approximate solution in which we assume that the observed signals are discrete, but sampled fine enough (in the channel bandwidth) (i.e., the observed signals are given at time points $\{0, \pm\Delta, \pm 2\Delta, \cdots\}$), and that $\Delta \approx 1/2W$ is much finer than, for example, the symbol period T. With this approximation the integration above becomes a summation, the observed signals can be represented as vectors, e.g., $\underline{r} = [\cdots, r(-n\Delta), \cdots r(0), \cdots, r(n\Delta), \cdots]^T$, and the convolution operator is approximated by

$$s(t)*h(t) = \int_{\tau} h(\tau)s(t - \tau) \, d\tau \approx S \cdot h$$

where S is a Toeplitz matrix, $S_{ij} = s(i\Delta - j\Delta)$, and h is a vector that represents the impulse response. The channel estimation problem becomes the least-squares problem:

$$\min_h \| r - S(\underline{a}) \cdot h \|^2$$

whose (unconstrained) solution is

$$\hat{\underline{h}} = \left(S(\underline{a})^\dagger S(\underline{a}) \right)^{-1} S(\underline{a})^\dagger \cdot \underline{r} = \text{PINV}\left(S(\underline{a}) \right) \cdot \underline{r} \quad (8)$$

where "$\dagger$" denotes the conjugate-transpose operation, and PINV(A) denotes the pseudo-inverse of A. The formula with the pseudo-inverse can be used even when $S^\dagger S$ is singular.

2) *Parametric Channel Models:* In many situations of interest we have some knowledge about the channel to be estimated, in terms of a parametric modeling of the channel impulse response. In this channel we will denote the channel impulse response $h(t; \underline{\theta})$, and the channel estimation problem is reduced to the problem of estimating the parameters $\underline{\theta}$. One example of such parametric modeling is the multipath channel, in which the impulse response can be written as

$$h(t) = \sum_{k=1}^{p} \alpha_k \delta(t - \tau_k). \quad (9)$$

This model has been justified above for the ocean channel. Assuming that the number of paths is known, say, from classical acoustic theory in the ocean channel, the unknown parameters are the path delays $\{\tau_k\}$ and the path attenuation $\{\alpha_k\}$, which can be complex to model phase shifts.

Incorporating this parametric knowledge in the channel estimation procedure can be very important. Not only will it improve the channel estimate for cases when the model indeed holds, but it will also lead to a smaller error probability of the extracted information bits when we jointly estimate the channel and the data. Intuitively, when there are no constraints on the channel impulse response, we increase the chance to find a wrong data sequence, which together with an erroneous channel response will lead to a small total square error.

The channel estimation problem can be written as the following nonlinear least-squares problem:

$$\min_{\underline{\theta}} \int_t \left| r(t) - \int_{\tau=0}^{D} h(\tau; \underline{\theta})\tilde{s}(t - \tau; \underline{a}) \, d\tau \right|^2 dt. \quad (10)$$

Unfortunately, this minimization can be highly complicated, even for a known data sequence. Nonlinear least-squares problems can be solved by standard iterative algorithms, e.g., the Gauss–Newton algorithm, or other nonlinear optimization techniques (see [35], [36]). However, due to the fact that we maximize a likelihood function, we can utilize the iterative EM algorithm which exploits the stochastic system under consideration. This algorithm has been suggested in [21] and has been applied to signal-processing problems in [37], [22], and elsewhere. The multipath channel model and other composite channels models can be considered as applications of the superimposed signal problem in [38]. The EM algorithm and its applications to cases like the multipath channel model is presented in more detail in Appendix B.

The performance of the channel estimation can be measured in terms of the mean square error between the true channel and the estimated one. Thus we can consider $E\{|h(t) - \hat{h}(t)|^2\}$ or its trace $\int_t E\{|h(t) - \hat{h}(t)|^2\} \, dt$, and if the channel is modeled parameterically, we can consider $E\{(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T\}$ or its trace, $tr(E\{(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T\})$. A lower bound can be easily found by calculating the Cramer–Rao lower bound for this Gaussian case. The lower bound for the channel parameter estimation error is

$$E\left\{(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^T\right\} \geq I^{-1} \quad (11)$$

where I is Fisher's information matrix, which for our problem its $\{ij\}$th element is

$$I_{ij} = \frac{1}{\sigma^2} \int_t \frac{\partial y(t; \underline{\theta})}{\partial \theta_i} \frac{\partial y(t; \underline{\theta})}{\partial \theta_j} \, dt \quad (12)$$

and $y(t) = h(t) * \tilde{s}(t; \underline{a})$ is the channel response to the input $\tilde{s}(t; \underline{a})$. Substituting the specific parametric model will lead to an explicit bound.

We note that the Cramer–Rao bound represents local effects on the estimation error. In our case we expect a large ambiguity error, especially since the symbols are modulated and we get signals whose structure is periodic; thus the bound is not tight and the square error is larger. The observation above explains the possible difficulties due to ambiguity, etc., which arise for different choices of the signaling techniques. As we have encountered in our experiments, these difficulties indeed exist in our equalization techniques that are based on joint data and channel estimation. However, this is an inherent problem of the channel estimation problem and not necessarily a fault of our suggested equalization algorithms; the ambiguity can only be fixed by using different signaling.

IV. ALGORITHMS FOR JOINT CHANNEL ESTIMATION AND DATA RECOVERY

In this section we present methods for jointly finding the information bit sequence and channel impulse response; i.e., a solution to the problem presented formally by (4). As was implied above, a common "theme" of our suggested algorithms is the alternation between two simpler optimization problems; namely, extracting the data assuming that the channel is given, and estimating the channel assuming that the data is known. More specifically, this "coordinate-search" algorithm is:

1) Set $n = 0$ and make an initial estimate of the channel $\hat{h}^{(n)}(t)$, (or its parameters $\hat{\underline{\theta}}^{(n)}$),
2) Based upon $\hat{h}^{(n)}(t)$ (or $\hat{\underline{\theta}}^{(n)}$), find the MLSE of the data sequence $\hat{\underline{a}}^{(n)}$,
3) Based upon $\hat{\underline{a}}^{(n)}$ (or $\tilde{s}(t; \hat{\underline{a}}^{(n)})$), update the channel estimate $\hat{h}^{(n+1)}(t)$ (or its parameters $\hat{\underline{\theta}}^{(n+1)}$),
4) Set $n = n + 1$ and return to step 2. Continue until the algorithm converges by some criterion.

By its nature the algorithm increases the likelihood, or decreases the square error, in each iteration. Under mild conditions it also converges to a stationary point of the goal function (which may unfortunately be a local minimum).

We will present below specific algorithms of the structure above which were designed by having the specific ocean-channel characteristics in mind. The data is processed in blocks, whose size fits our assumptions for the time period in which the channel is stable; i.e., no more than half a second. Thus we are ready to cope with situations in which the channel has been changed completely and abruptly between blocks. Block "tailoring" methods such as block overlapping and using previous block estimation as the next block initial condition will also be mentioned. Although other variations and scenarios can be thought of, we present the following specific algorithms:

- In Section IV-A we deal with parametric channel models, and more specifically, the multipath channel model. The nonlinear least-squares problem needed for step 3 above is implemented using the EM algorithm.
- In Section IV-B we deal with the case where the channel's impulse response is finite but otherwise unconstrained, and thus it can be found as a solution of a linear least-squares problem. As will be seen, we will be able to combine steps 2 and 3 above and get a closed-form solution for the information bit-extraction problem.

A. The Parametric Multipath Channel Model

Joint channel and data estimation, when a parametric multipath model for the channel is assumed, is now considered. Under this modeling the channel impulse response is given by

$$h(t) = \sum_{k=1}^{p} \alpha_k \delta(t - \tau_k). \qquad (13)$$

The motivation and validity of this model have been presented above. With this channel model, the transmitted complex signal $\tilde{s}(t; \underline{a})$ is observed at the receiver as

$$r(t) = \sum_{k=1}^{p} \alpha_k \tilde{s}(t - \tau_k; \underline{a}) + n(t). \qquad (14)$$

We assume that the multipath order is known. A typical number for the ocean channel is 4.

Following the generic algorithm described above, we start with an estimate of the parameters $\{\alpha_k^{(0)}\}$ and $\{\tau_k^{(0)}\}$. At each iteration having the current estimate, we can find the data sequence:

$$\underline{a}^{(n)} = \arg \min_{\underline{a}} \sum_{i} \left| r(t) - \sum_{k=1}^{p} \alpha_k^{(n)} \tilde{s}(t - \tau_k^{(n)}; \underline{a}) \right|^2 \qquad (15)$$

using the Viterbi algorithm. Note that we have replaced the integral by a summation over t, assuming that the observed signal is discrete and finely sampled. In the implementation of the Viterbi algorithm each state s_i in the trellis is defined by the value of the current symbol a_i and the values of the previous q symbols, where $q = \lfloor D/T \rfloor$, D is the longest delay, and T is the symbol period. The knowledge of the state allows us to calculate the metric,

$$\sum_{t=(i-1)T}^{iT-1} \left| r(t) - \sum_{k=1}^{p} \alpha_k^{(n)} \tilde{s}(t - \tau_k^{(n)}; a_i, \cdots, a_{i-q}) \right|^2$$

for the ith symbol period, needed for the channel estimation.

The next step, having the data sequence estimate, is to update the channel parameters. We use the EM algorithm described in Appendix B to solve the nonlinear least-squares problem needed for this update. In a specific implementation, say, only one iteration of the EM algorithm can be performed. This iteration leads to the following procedure: Define

$$e(t) = r(t) - \sum_{k=1}^{p} \alpha_k^{(n)} \tilde{s}(t - \tau_k^{(n)}; \underline{a}^{(n)}) \qquad (16)$$

to be the error signal associated with the optimal path of the Viterbi algorithm. Generate, then, the p signals:

$$x_k(t) = \alpha_k^{(n)} \tilde{s}(t - \tau_k^{(n)}; \underline{a}^{(n)}) + \beta_k e(t), \qquad k = 1, \cdots, p \qquad (17)$$

where $\{\beta_k\}$ are non-negative numbers whose sum is 1. The new estimate of the delays is given by

$$\tau_k^{(n+1)} = \arg \min_{\tau} \sum_{t} |x(t) - \tilde{s}(t - \tau; \underline{a}^{(n)})|^2$$

$$\approx \arg \max_{\tau} \left| \sum_{t} x^*(t) \tilde{s}(t - \tau; \underline{a}^{(n)}) \right| \qquad (18)$$

and the approximation is valid when the allowable delay is small compared to the entire observation window or when the energy $\sum_t \tilde{s}^2(t - \tau; \underline{a})$ is independent of τ.

When we use the EM procedure in a strict way, the kth amplitude $\alpha_k^{(n+1)}$ is the value of the cross correlation in (18) at the estimated delay $\tau_k^{(n+1)}$. However, we observe that the non-linear least-squares problem for estimating the channel parameters, i.e.,

$$\sum_t \left| r(t) - \sum_{k=1}^{p} \alpha_k \tilde{s}\left(t - \tau_k^{(n)}; \underline{a}^{(n)}\right) \right|^2 \qquad (19)$$

becomes a linear least-squares problem with respect to the amplitudes, and thus having an estimate $\tau_k^{(n+1)}$, the amplitude update will be:

$$\alpha^{(n+1)} = \mathscr{R}_s^{-1} r_{rs} \qquad (20)$$

where $\alpha = [\alpha_1, \cdots, \alpha_p]^T$ is the vector of the amplitudes, $R_s = R_s(\underline{\tau}; \underline{a})$ is a $p \times p$ correlation matrix such that $(R_s)_{ij} = \sum_t \tilde{s}^*(t - \tau_i)\tilde{s}(t - \tau_j)$, the vector $r_{rs}(\underline{\tau}; \underline{a})$ is the p dimensional cross-correlation vector whose ith component is $\sum_t r(t)\tilde{s}^*(t - \tau_i)$, and τ is the vector of delays.

The algorithm is now completely specified. At each iteration, start with an estimate of the paths delays and amplitudes and find a data sequence by minimizing (15) using the Viterbi algorithm of Appendix A. Then find the updated delays according to (18) and the updated amplitudes according to (20). Initial conditions for delays and amplitudes can be achieved using any prior knowledge or other common technique; e.g., looking at the peaks of the signal autocorrelation.

We note that this algorithm iterates on blocks of the observed data. In each such set of iterations the observed signal depends on the previous $q = \lfloor D/T \rfloor$ symbols in addition to the n symbols transmitted at the observation window. In the practical implementation of the algorithm, the previous D samples of the signals, together with an estimate of the previous q symbols, are used in (18) and (20) for updating the delay and amplitude estimates. These q symbols have already been recovered while the previous block has been processed, and we suggest using these previous recovered symbols while processing the current block. Other "block overlapping" methods can also be considered and the various methods should be analyzed further.

The algorithms presented in this section are also valid for other models, which can be described as "superimposed channel" models. For example, consider the channel in which each different path also undergoes a different Doppler frequency shift. The observed signal in this case will be

$$r(t) = \sum_{k=1}^{p} \alpha_k \tilde{s}(t - \tau_k; \underline{a}) e^{j\omega_k t} + n(t). \qquad (21)$$

The suggested algorithm is analogous to the algorithm presented above, where we have to use $\tilde{s}(t - \tau_k^{(n)})e^{j\omega_k t}$ instead of $\tilde{s}(t - \tau_k)$ in (16)–(18), and the optimization in (18) should be with respect to τ and ω.

B. Closed-Form Solution for the Unconstrained Channel Case

The algorithm below considers the case where no assumptions on the channel impulse response have been made, besides the fact that it is finite and its length D is known. This is the most general case we consider and it will lead to the most simply expressed solution (which may be computationally complex, however) for the joint channel and data estimation problem. This algorithm will be very powerful for cases of fast, arbitrarily varying channels where indeed no assumption can be made about

the channel and where all the other model-dependent methods fail. However, the available degrees of freedom in the choice of the channel may result in a higher probability of error compared to the case where the impulse response is known.

In deriving the method, we use, for clarity, the approximation in which the observations are discrete-time, sampled finely enough. The goal function and the joint channel and data estimation are given by

$$\min_{\underline{a}, \underline{h}} M(\underline{a}, \underline{h}) = \min_{\underline{a}, \underline{h}} \| \underline{r} - S(\underline{a}) \cdot \underline{h} \|^2. \qquad (22)$$

As mentioned in Section III-C above, when the information bits are known and thus $S(\underline{a})$ is given, the channel estimation is given by (see (8)):

$$\tilde{\underline{h}} = \hat{\underline{h}}(\underline{a}) = \left(S(\underline{a})^\dagger S(\underline{a})\right)^{-1} S(\underline{a})^\dagger \underline{r} \qquad (23)$$

which is the best channel for that data sequence $\underline{a}$. We can now substitute $\hat{\underline{h}}(\underline{a})$ in (22) and the optimal data sequence will be found by minimizing the resulted goal function,

$$\min_{\underline{a}} M\left(\underline{a}, \hat{\underline{h}}(\underline{a})\right) = \min_{\underline{a}} \| \underline{r} - S(\underline{a})\left(S(\underline{a})^\dagger S(\underline{a})\right)^{-1} S(\underline{a})^\dagger \underline{r} \|^2$$

$$= \min_{\underline{a}} \| \left(I - S(\underline{a})\left(S(\underline{a})^\dagger S(\underline{a})\right)^{-1} S(\underline{a})^\dagger\right) \underline{r} \|^2. \qquad (24)$$

Now, since $I - S(\underline{a})(S(\underline{a})^\dagger S(\underline{a}))^{-1}S(\underline{a})^\dagger$ is a projection operator, it is idempotent and the optimal data sequence is found as

$$\hat{\underline{a}} = \arg \min_{\underline{a}} \underline{r}^\dagger \left(I - S(\underline{a})\left(S(\underline{a})^\dagger S(\underline{a})\right)^{-1} S(\underline{a})^\dagger \right)$$

$$\cdot \underline{r} = \arg \max_{\underline{a}} \underline{r}^\dagger S(\underline{a})\left(S(\underline{a})^\dagger S(\underline{a})\right)^{-1} S(\underline{a})^\dagger \underline{r}. \qquad (25)$$

In a detailed implementation of the algorithm, suppose that T samples are available for each symbol period, and assume that the observation window is n-symbol long. In this case the observed signal is the vector:

$$\underline{r} = \left[r(0), \cdots, r(n-1), r(n), \cdots, \right.$$

$$\left. \cdots, r(n(T-1)), \cdots, r(nT-1) \right]^T$$

and $S(\underline{a})$ is a $nT \times D$ Toeplitz matrix, where D is the channel impulse response length or effective ISI length, such that $S_{t,\sigma}(\underline{a}) = \tilde{s}(t - \sigma; \underline{a})$. Note also that the multiplication $S(\underline{a})^\dagger \underline{r} = \underline{y}(\underline{a})$ represents the cross correlation between the transmitted and observed signal, and $S(\underline{a})^\dagger S(\underline{a}) = R_s(\underline{a})$ is the autocorrelation matrix of the transmitted signal. Note that although the observation window is n-symbol long, the expression (25) depends on the $n + q$ symbols $[a_{-q} \cdots a_0 \cdots a_{n-1}]$, where q is the smallest integer greater than D/T.

Computational Aspects and Implementation Details: The complexity of the computation in (25) is due to two factors: First, the goal function depends on the entire data sequence and thus its minimization requires an exhaustive search, whose complexity is exponential in the sequence length. Secondly, for each candidate sequence we have to calculate (25), which can be computationally complex. We will deal with the complications resulted from the first factor later.

The computation of the goal function (25) for each data sequence can be calculated efficiently, using the following recur-

sive method. As noted above, the goal function can be written as

$$M(\underline{a}) = M(\cdots, a_n)$$

$$= \underline{y}(\cdots, a_n)^\dagger R_s^{-1}(\cdots, a_n) \underline{y}(\cdots, a_n) \quad (26)$$

where we recall that $\underline{y}$ is of length D, and R_s is a $D \times D$ matrix. Now, given a new set of measurements for the observation window $[nT, (n + 1)T - 1]$ which corresponds to the new symbol a_{n+1}, the new cross-correlation vector $\underline{y}(\cdots, a_n, a_{n+1})$ and the new autocorrelation matrix $R_s(\cdots, a_n, a_{n+1})$ can be calculated from the previous cross correlation and previous autocorrelation as follows: By definition, the cross-correlation vector satisfies

$$\underline{y}(\cdots, a_n, a_{n+1}) = \underline{y}(\cdots, a_n) + \sum_{t=nT}^{(n+1)T-1} r(t)\underline{\tilde{s}}^*(t) \quad (27)$$

where $\underline{\tilde{s}}(t) = [\tilde{s}(t), \cdots, \tilde{s}(t - D + 1)]^T$. The autocorrelation matrix satisfies

$$R_s(\cdots, a_n, a_{n+1}) = R_s(\cdots, a_n) - \sum_{t=nT}^{(n+1)T-1} \underline{\tilde{s}}(t) * \underline{\tilde{s}}(t)^T. \quad (28)$$

Thus, using the matrix inversion formula,

$$(A + \underline{v}\underline{v}^T)^{-1} = A^{-1} - \frac{1}{1 + \underline{v}^T A^{-1} \underline{v}} \underline{v}\underline{v}^T \quad (29)$$

we can perform, for the $(n + 1)^{st}$ symbol duration, the loop:

- Set $A^{-1}(0) = R_s^{-1}(\cdots, a_n)$
- For $i = 0, \cdots, T - 1$

$$A^{-1}(i + 1) = A^{-1}(i)$$

$$- \frac{1}{1 + \underline{\tilde{s}}(nT + i)^T A^{-1}(i)\underline{\tilde{s}}(nT + i)^*}$$

$$\cdot \underline{\tilde{s}}(nT + i)^* \underline{s}_D(nT + i)^T \quad (30)$$

and get $R_s^{-1}(\cdots, a_n, a_{n+1}) = A^{-1}(T)$, recursively, without having to do any explicit matrix inversion.

The major complexity factor of the closed-form solution is the requirement for an exhaustive search over all possible data sequences. However, in the ocean environment where the channel may change rapidly, only a few symbols are transmitted while the channel is stable. Thus this search may be tractable; ironically, the same aspects of the problem that make it hard and lead to poor error-probability performance, make the closed-form solution a valid answer to the equalization problem.

An alternative to the exhaustive search will be a suboptimal tree search. Tree search is used in several noncoherent detection problems. Under this procedure we will first fix a small m (but $m > q$) and calculate the score of (26) for all A^m paths, where A is the symbol alphabet size of all possible data sequences of length m. Then as the tree is extended by another symbol, the score is calculated recursively for all the extensions using (27) and (28). Now we keep only B paths, where B is defined by the computational complexity and available memory. Only these B paths are extended later, and we keep at each stage of the tree only the best B paths. Clearly, this procedure is suboptimal, but

in a good SNR the loss with this procedure is expected to be minimal.

The incorporation of the q symbols from the previous block that affect the current block can be done in a fashion similar to what was suggested for the multipath model in the previous section. Other alternatives can also be considered and are now under investigation. We also mention that a modification of the algorithm to a nonblock form in which for tracking purposes we use a weighted least-squares goal function that weights past samples in an exponential manner can be considered. Under this modification we get recursive formulas similar to (27) and (28) for calculating the goal function, since the weighting is exponential. Analysis and experiments with this modified algorithm are also under current investigation.

V. EXPERIMENTAL RESULTS

A. The Multipath Model

We have tested the algorithm described in Section IV-A by simulations using modulation and impulse response parameters which characterize underwater modems and the ocean channel. Specifically, a common modulation technique is multiple tones FSK, (MFSK), which makes the communications more robust to underwater phase and fading instability; for example, in an experimental modem developed in the Woods Hole Oceanographic Institution [39] this MFSK modulation is used. The frequencies of the signaling pulses are chosen in a way that they are orthogonal; i.e., they are $\Delta f = 1/T$ apart, where T is the pulse length. In this experimental modem the symbol duration is 12.5 ms; this pulse width is long enough so that the ISI effects will spread out to only a few past symbols, but it is small enough so that the pulse frequencies are far enough apart to allow compensation for Doppler shifts.

We have generated in a simulation an FSK signal with orthogonal signaling, as shown in Fig. 1(a), where this signal modulates 32 b chosen at random. We assume that along this 32 symbol duration, i.e., with T as defined above, along 409.6 ms the channel is fixed. A typical ISI effect is shown in Fig. 1(b), where we passed the modulated FSK signal through a multipath channel having four paths, whose parameters are

$$\tau_1 = 0.4 \quad \tau_2 = 10.8 \quad \tau_3 = 14 \quad \tau_4 = 27.4$$

$$\alpha_1 = 1 \quad \alpha_2 = 0.5 \quad \alpha_3 = 0.35 \quad \alpha_4 = -0.7$$

we see that the longest delay is a little longer than a two-pulse duration.

The signals are observed with noise. In Fig. 2(a) and (b) we can see the simulated observed signals, where the SNR is 25 dB in Fig. 2(a), and 8 dB in Fig. 2(b). The SNR is the post-integration signal-to-noise ratio, defined as

$$10 \log \frac{\int_0^T s_p^2(t; a) \, dt}{\sigma^2} \quad \text{or} \quad 10 \log \frac{\sum_0^{T-1} s_p^2(t; a) \, dt}{\sigma^2}$$

where T is the symbol duration (either in time or samples).

The algorithm suggested in Section IV-A has been tested on three observed signals. In a specific example in the high SNR case, we have observed the following performance: The channel parameters' behavior as a function of the iteration index is summarized in Fig. 3, where in Fig. 3(a) we see the delay estimates, and in Fig. 3(b) the amplitude estimates. The true delays and amplitudes are shown in dotted lines. Note that after 20 iterations we are very close to the true parameters. Since the
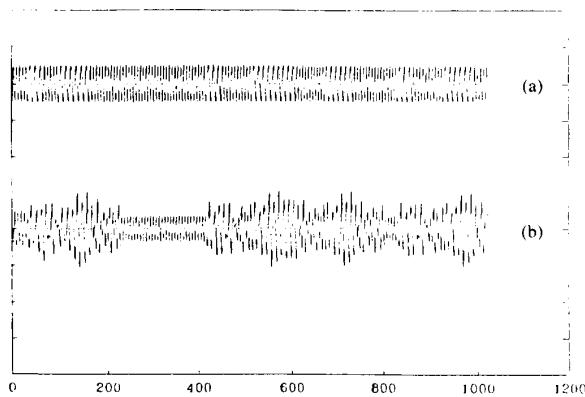
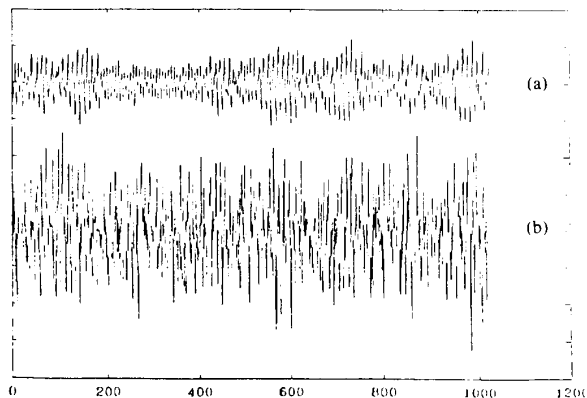Fig. 1. (a) The modulated signal. (b) After channel.

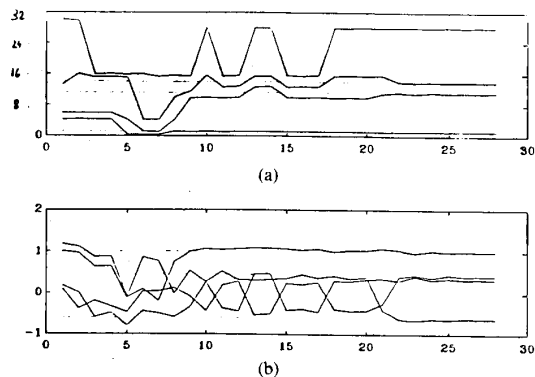Fig. 2. Signal with noise. SNR: (a) 25 dB, and (b) 8 dB.

Fig. 3. Channel parameter estimates as a function of iteration index, SNR = 25 dB. (a) Delays. (b) Amplitudes.
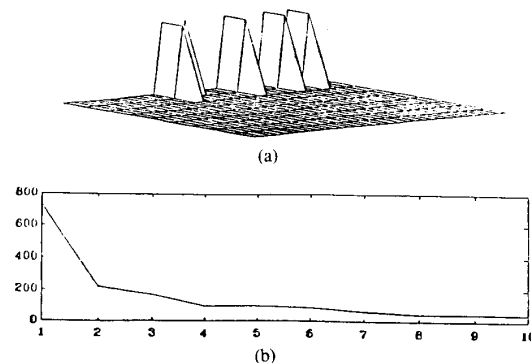

Fig. 4. The extracted bits as a function of iteration index. (a) Bit errors. (b) Logarithm of the signal square error.
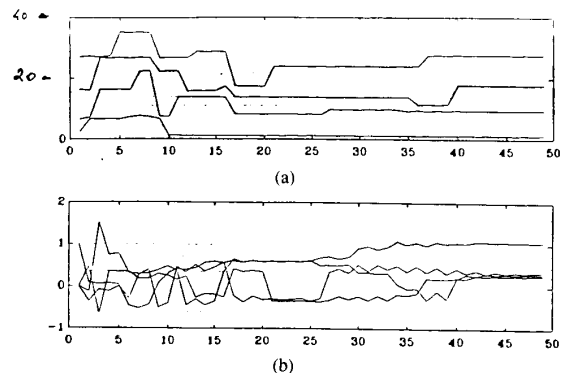

Fig. 5. Channel parameter estimates as a function of iteration index, SNR = 8 dB. (a) Delays. (b) Amplitudes.

Viterbi algorithm is the computationally heaviest part of the equalization procedure, we have performed four EM iterations for updating the channel parameters for each Viterbi iteration for data extraction. The bit errors as a function of the iteration index are shown by the difference between the true bit sequence and the extracted data sequence, plotted as a mesh graph in Fig. 4(a); thus we can see which of the 32 bits were erroneous. We see that in this high SNR case after the third Viterbi iteration, the data sequence is extracted without any error. In Fig. 4(b) we see the logarithm of the square error (which is proportional to

the minus likelihood) in the output of the Viterbi iteration, decreasing as a function of the iteration index.

The low SNR case has also been tested. The channel parameter behavior as a function of the iteration index is given in Fig. 5. Here we see that the first and fourth delay are estimated correctly, while the two which are closer together, the second and third delays, are estimated with larger errors. Nevertheless, when we observe Fig. 6, in which Fig. 6(a) described the number of bit errors as a function of the iteration index, and Fig. 6(b) is the mesh plot as described above, we see that after 11 Viterbi iterations the data is recovered without any error. The logarithm of the square error in the output of the Viterbi iteration is shown in Fig. 6(c) as a function of the iteration index.

As a comparison, we have checked the bit error-rate performance in these examples if no equalization was made. The procedure is a simple coherent detection (we assume that the initial delay is given), and we get 2-b errors in the high SNR case and 6-b errors in the low SNR case. This fact implies that the interference effects indeed make our example a nontrivial one.

The channel parameter estimation with FSK signaling may be an ambiguity-prone problem, as noted above. We can appreciate this problem by observing in a specific EM iteration the output of the cross-correlation function for each channel, calculated in the M step and shown in Fig. 7. Despite the fact that in this example the maximum for each path is achieved near the true delay, the existence of other peaks may lead in previous iterations and in other noise levels to large delay errors.
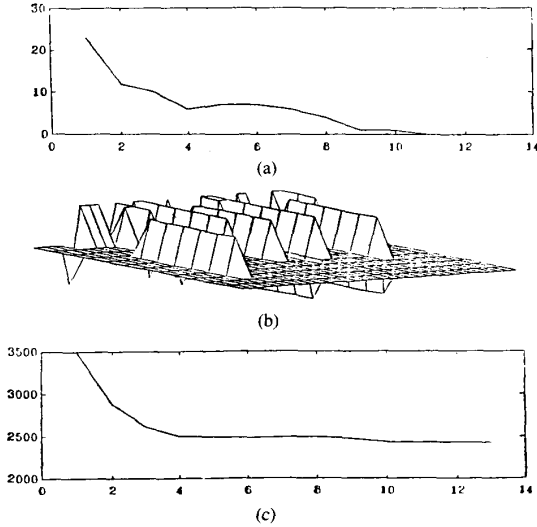
Fig. 6. Bit error, SNR = 8 dB: (a) Number of bit errors. (b) Their location. (c) Logarithm of the signal square error as a function of iteration index.
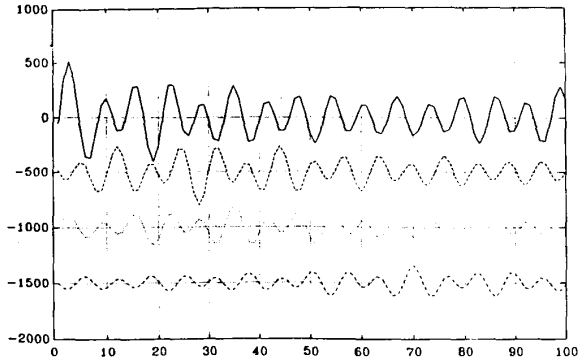

Fig. 7. The cross correlation in each channel of the EM iteration.

To get more reliable statistics on the performance of the suggested equalization procedure, we have run the procedure on 1000 blocks, where in each block the channel parameters are reestimated independently. These experiments have been conducted for SNR of 5 and 8 dB. The channel in this example was composed of two paths: The first path was at a delay of 0.4 ms and amplitude 1, and the second at a delay of 28 ms and amplitude 0.7. We have compared the following three procedures: The first is a known channel case in which we assume that the channel is given and we just extract the data using the Viterbi algorithm. The second procedure is our equalization algorithm; and the third is the "no equalization" case, in which we just estimate the first path delay. The average bit-error percentage for these three procedures were: (i) Known channel (MLSE algorithm): 10%, at 5 dB, 4% at 8 dB; (ii) equalization using our algorithm: 20% at 5 dB, 11% at 8 dB; and (iii) no equalization (delay estimate): 23% at 5 dB, 14% at 8 dB.

Thus there is some improvement in this example over the no-equalization case. The fact that the MLSE performance is superior implies that in our experiments the channel estimate was occasionally wrong, probably due to convergence to the local maximum of the likelihood function. This issue, and the

related issue of choosing an initial guess for the algorithm, affect the practical implementation of our algorithm and are now under investigation.

B. Closed-Form Solution

The closed-form solution has been examined on the modulation technique described above; i.e., FSK modulation. We have examined the version that requires an exhaustive search over all possible bit sequences. Due to computational constraints we have searched only over 11 b; i.e., over 2^{11} possibilities. The generated signals are as above. We assume that the length of the impulse response is 2.5 symbol durations. The assumed length D of the impulse response in samples affects the computational complexity, since it defines the size $D \times D$ of the matrices to be inverted.

In our experiments at high SNR the true bit sequence has always maximized the score, given by (25). To illustrate the performance of this method we look on this score as a function of the 2048 candidate bit-strings (see Fig. 8). In Fig. 8(a) we see the score as a function of the bit sequences given in lexicographic order and in Fig. 8(b) we have sorted the score according to its values, and thus we can see the "sharpness" of the correct solution with respect to the other candidate solutions. This sorted score is shown also in Fig. 8(c) for eight different transmitted bit-sequences, chosen at random.

VI. SUMMARY AND CONCLUSIONS

In this paper we have examined the equalization problem in underwater acoustic channels. We have pointed out the uniqueness and special difficulties which arise in these channels and the inadequacy of applying standard equalization techniques in the ocean. These special features have led us to consider joint estimation of the channel parameters and information symbols. This idea has been studied for other communication channels; however, in many of these channels it was ruled out, since the rate of information transmission was much higher than the rate of channel variations. In these channels, which may include even the fading HF channel, it was unnecessary to re-estimate the channel very often, especially since the joint estimation may be more complicated and less robust than in other techniques; e.g., using an adaptive linear equalizer or the MLSE with channel parameters that are updated slowly. The ocean channel is, unfortunately, more fluctuating, especially when we compare its instability rate to the relative slow symbol rate.

The solutions we have suggested (i.e., the closed-form solution in the unconstrained case and the iterative method based on the MLSE and EM algorithms for the parametric case) assume a linear channel model which in some observation window is time invariant. This observation window must be small, following the discussion above. Although this fact limits the possible performance, it may allow relatively complex processing techniques for extracting the small number of symbols in the processed window. The performance of our suggested method is now studied and we examine ways to make our algorithms more robust.

As mentioned above, one possible drawback of our iterative algorithm based on the EM and MLSE algorithms is the sensitivity to the initial guess of the channel parameters; this sensitivity is due to the multimodal structure of the likelihood function and the existence of local maxima. This problem, however, is primarily a result of the modulation method used. Any equalization technique will fail to cope with multipath when the modulated
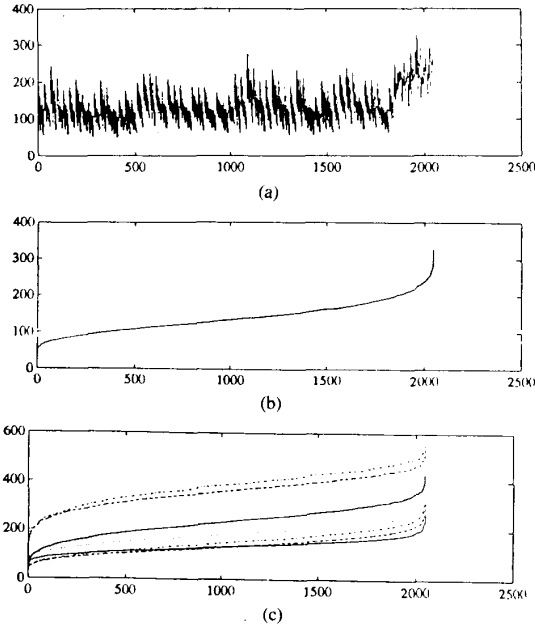
Fig. 8.   The closed-form solution, score for each candidate sequence. (a) In lexiographic order; (b) sorted; and (c) eight different experiments, sorted.

signal correlation function tends to be periodic. Thus our research may suggest criteria for choosing modulation waveforms that have a better correlation behavior; taking in mind, say, the way whales communicate underwater, one such signal may be a chirp signal.

Another research direction we pursue is the study of sequential algorithms for joint channel and data estimation. The structure of the solution for the closed-form algorithm will be based on the properties of the matrices used in the derivation of Section IV-B. A sequential version of the suggested iterative algorithm will be based on the general structure of the sequential EM-type algorithms derived in [22], [40], [41]. Having a sequential algorithm, the equalizer structure is such that it uses an initial estimate of the channel and then tracks its variations. This may improve the bit error performance for cases where the channel is more stable without losing the ability to track the channel variations.

APPENDIX A
MLSE VIA THE VITERBI ALGORITHM

The General Bellman Algorithm

The Viterbi algorithm is a special case of the more general dynamic programming approach developed originally by Bellman [42]. The problem solved by this algorithm can be described as follows.

Consider the directed layered graph composed of N layers, such that each layer contains K nodes, and each node in the ith layer can be connected only to nodes in the next $(i + 1)^{st}$ layer. In general, a weight can be associated with each node and each edge of this graph. An example of this graph is given in Fig. 9. The nodes at the ith layer are sometimes called the possible states at time i. Now suppose that we want to find the path in the graph that has, say, the minimal weight. The weight of a path is by definition the sum of the weights associated with the nodes and edges along the path. The general exhaustive search ap-

proach would require passing over all K^N paths; i.e., the complexity of the search is exponential.

Bellman's dynamic programming approach is based on the following observation. Suppose we know that the optimal path will pass through the state $j \in \{1, \cdots K\}$ at time i. Clearly, the accumulated weight of this path up to time i must be smaller than that of any other path passing through that state j at time i. Otherwise the optimal path would include the part of this other path. Thus at each time i we only have to find the K paths from the beginning to time i ending in each of the states, whose accumulated weight is minimal. These K paths can be found recursively. Suppose we have the K optimal paths up to time i. When we go to the next time, $i + 1$, we only extend these optimal paths (in general there are K^2 possible extensions), and for each state at time $i + 1$ we find the new optimal path. When we reach the final time N, we have the K paths, each ending at a different state such that the weight of each path is the minimal out of all the paths ending at the same state. We can now search over these K paths to find the minimal path through the layered graph, which is the solution to our problem. This algorithm is recursive and its complexity is linear in N.

The Viterbi Algorithm — Application to MLSE

The application of Bellman's algorithm to decoding convolutional codes was originally termed the "Viterbi algorithm," and it is described in detail in [43]. In this algorithm the time index represents the data symbol index. The state at each time i is composed of all possible A^{q+1} values of the symbols $a_i, a_{i-1}, \cdots, a_{i-q}$, where q is the constraints length, in the convolutional codes terminology, and the length of the impulse response for our equalization purposes.

Now as we go to time $i + 1$ each state can lead to one of A states, depending on the values of a_{i+1} and, vice versa, each new state can be generated from one of A previous states, as shown in Fig. 9. Thus at time $i + 1$, we find for each state the minimum out of the A paths passing through its state at time i.

The only distinction between our implementation and any other implementation of the Viterbi algorithm is the weight associated with the nodes and edges. For our equalization purposes, we associate weights only to nodes. Specifically, for the state defined by $a_i, a_{i-1}, \cdots, a_{i-q}$ at time i, the weight given by

$$W_i(j) = W_i(a_{i-q}, \cdots, a_i)$$

$$= \int_{(i-1)T}^{iT} |r(t) - \tilde{s}(t; \underline{a}) * h(t)|^2 dt \quad (A1)$$

where

$$\tilde{s}(t; \underline{a}) * h(t) = \int_0^D h(\tau) \sum_k s_p(t - kT - \tau; a_k) d\tau \quad (A2)$$

and since $s_p(t)$ is different from zero only at $[0, T]$ and $q = \lfloor D/T \rfloor$, this convolution, calculated for the time period $(i - 1)T \leq t < iT$, depends only on symbols $a_{i-q}, \cdots, a_i$. The total score for each path is $\sum_i W_i(j)$; i.e., the sum of partial weights above, which is also:

$$\int_0^{NT} |r(t) - \tilde{s}(t; \underline{a}) * h(t)|^2 dt \quad (A3)$$

where $\underline{a}$ is the entire data sequence, and its minimization provides the MLSE.

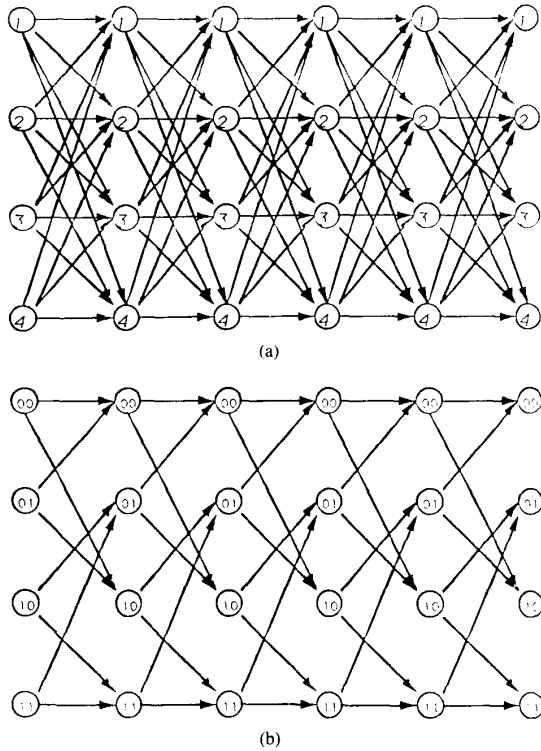The MLSE can be extracted only after processing the entire

(a)

(b)

Fig. 9. Graphs of Bellman's and Viterbi's algorithm. (a) General state diagram—layered graph. (b) State diagram for the Viterbi algorithm (binary case).

observation signal. Then when we reach the final time, the optimal path is available and we can reconstruct, backwards, the data sequence. In practical implementation of the algorithm the data is recovered in blocks. In our algorithm, where we also estimate the channel, block processing enters in a natural way.

We also note that while calculating the score of the optimal path, we calculate the difference $r(t) - h(t)*s(t; \underline{a})$ needed later for the channel estimation part of the algorithm. This motivated the minor variation from Forney's original presentation [16], in which the Viterbi algorithm operates on the sufficient statistics, extracted at the symbol rate, as discussed in Section III-B. Regarding this point, we emphasize again that this statistic is not sufficient for our case where the channel is also estimated, and if we want to use the original form of Forney's algorithm the statistics should have be extracted again for every new estimate of the channel; thus we may as well use the weights as given in (A1) above.

APPENDIX B

THE EM ALGORITHM

In our channel equalization algorithm the EM algorithm is applied to the estimation of the channel parameters in each iteration. The EM algorithm will treat the estimate of the data sequence $\underline{\hat{a}}$ as fixed, and update the estimate of the channel parameters to yield $\underline{\hat{\theta}}^{(n+1)}$. This appendix will start with a description of the generic EM algorithm, and then describe its application to the multipath channel-estimation problem. As noted in Section IV-A, during each iteration of the equalization algorithm only one iteration of the generic EM algorithm is made.

The Generic EM Algorithm

The EM algorithm is an iterative procedure suggested originally in [21], which at each step of its iteration updates its estimate of the desired parameters in such a way as to increase the log-likelihood function based upon the observed data y. The algorithm is motivated by the observation that there may be a data set, called the "complete" data set, denoted $\underline{x}$, with which it would be easier to determine an ML estimate of the parameters of interest than it is with the actual observation data. The "complete" data $\underline{x}$ is related to the observed data by the non-invertible transformation $y = H(\underline{x})$ and thus contains more information on the parameters than y. The principle of the algorithm is that given the observed data and an estimate of the parameters of interest $\hat{\theta}$; one can estimate the sufficient statistics $\underline{x}$ by its conditional expectation and thus to have the conditional expectation, given y and $\underline{\theta}^{(n)}$, of the log-likelihood of the "complete" data (the E step). Then one can make an updated estimate of the parameters of interest $\hat{\theta}^{(n+1)}$ by maximizing this estimate of the log-likelihood of the "complete" data (the M step). It can be shown that this procedure guarantees that the log-likelihood function based on the observed data will be greater for $\hat{\theta}^{(n+1)}$ than for $\hat{\theta}^{(n)}$ (see, for example, [38, pp. 477 and 478]). Given a new estimate of the parameters of interest, a new set of sufficient statistics of the "complete" data is estimated and the procedure iterates until convergence. This procedure can be summed up as:

1) Make an initial estimate of the parameters of interest ($\underline{\hat{\theta}}^{(0)}$) and set $n = 0$.
2) E Step: Estimate the sufficient statistics of the complete data, using the observed data, y, and $\underline{\hat{\theta}}^{(n)}$. Substitute it in the expression for the log-likelihood of the complete data to get its conditional expectation.
3) M Step: Solve $\underline{\hat{\theta}}^{(n+1)} = \arg\max_\theta E\{\log f(\underline{x}; \underline{\theta}) \mid y, \underline{\hat{\theta}}^{(n)}\}$.
4) Let $n = n + 1$ and go to step 2. Iterate until convergence.

Note that the term to be maximized in step 2 is denoted by $U(\underline{\theta}, \underline{\hat{\theta}}^{(n)})$ in [38]. If the choice of $\underline{x}$ is made intelligently, the maximization of $U(\underline{\theta}, \underline{\hat{\theta}}^{(n)})$ will be much easier to perform than the explicit maximization of the log-likelihood function based on the observed data $\mathscr{L}(\theta)$. The disadvantage of this approach to ML estimation is that the algorithm, as with any other "hill climbing" algorithm, is not guaranteed to converge to the true ML estimate of the parameters of interest, but only to converge to an estimate which is a stationary point of $\mathscr{L}(\theta)$.

Multipath Estimation with the EM Algorithm

Let the observed signal be:

$$y(t) = \sum_{k=1}^{p} \alpha_k s(t - \tau_k) + n(t) \qquad (B1)$$

where $\{\tau_k, \alpha_k\}$ are the delays and amplitudes associated with the paths, $s(t)$ is some deterministic known signal, and $n(t)$ is a sample function of a white Gaussian noise process. In our case, the known signal will be the modulated signal $\tilde{s}(t; \underline{a})$. With this signal model, the log-likelihood function of the observed signal is

$$\mathscr{L}(\underline{\hat{\theta}}) = -\int_t \left| y(t) - \sum_{i=1}^{p} \alpha_i s(t - \tau_i) \right|^2 \approx$$

$$-\left\| \underline{y} - \sum_{i=1}^{p} \underline{s}(\tilde{\theta}_i) \right\|^2 \qquad (B2)$$

where $\underline{\theta} = [\tau_1, \cdots, \tau_p, \alpha_1, \cdots, \alpha_p]^T$ is the parameter vector, and $\underline{s}$, $\underline{y}$ are vectors that represent the signals $y(t)$ and $s(t)$, sampled finely enough (e.g., at their Nyquist rate). The maximization of this function is a coupled multiparameter optimization problem with $2p$ parameters. In [38], $\underline{x}$ is defined in such a way that the maximization of $U(\underline{\theta}, \hat{\underline{\theta}}^{(n)})$ becomes p decoupled optimization problems each with two parameters. This "complete" data is the set of p signals $\underline{x}(t) = [x_1(t), \cdots, x_p(t)]^T$ which are independent observations of the signal that is received via each of the p propagation paths, each contaminated by white Gaussian noise; i.e.,

$$x_k(t) = \alpha_k s(t - \tau_k) + n_k(t). \qquad (B3)$$

Given the current estimates of the channel parameters $\hat{\underline{\theta}}^{(n)}$, the conditional expectation of the log-likelihood of the complete data is given by

$$U(\underline{\theta}, \hat{\underline{\theta}}^{(n)}) = -\sum_{i=1}^{p} \left(\int_t | \hat{x}_i^{(n)}(t) - \alpha_i s(t - \tau_i; \hat{\underline{a}}^{(n+1)}) |^2 \, dt \right)$$

$$\approx -\sum_{i=1}^{p} \| \hat{\underline{x}}_i^{(n)} - \underline{s}(\theta_i, \hat{\underline{a}}^{(n+1)}) \|^2 \qquad (B4)$$

where the estimate of the complete data, in the vector notation, is

$$\hat{\underline{x}}_i^{(n)} = E\{ \underline{x}_i | \underline{y}, \hat{\theta}_i^{(n)} \} = \underline{s}(\hat{\theta}_i^{(n)}) + \beta_i \hat{\underline{n}}^{(n)}$$

and $\beta_i \hat{\underline{n}}^{(n)}$ is the estimate of the noise in the ith propagation path. $\hat{\underline{n}}^{(n)}$ is the difference between the received signal and predicted received signal based upon the current channel parameter estimate and is given by

$$\hat{\underline{n}}^{(n)} = \underline{y} - \sum_{i=1}^{p} \underline{s}(\hat{\theta}_i^{(n)}).$$

The betas are arbitrary non-negative scalars chosen such that $\sum_{i=1}^{p} \beta_i = 1$; they represent the ratio $\sigma_{n_i}^2 / \sigma_n^2$ between the variance of the noise in the ith path to the variance of the total noise. Similar formulas can be written for the continuous time-signal notation.

The maximization of (B4) can be seen to be p decoupled matched filtering problems where the estimated signal for the ith propagation path is passed through a matched filter to estimate the attenuation and delay for that path. Therefore to find the set of attenuations and delays to maximize $U(\underline{\theta}, \hat{\underline{\theta}}^{(n)})$, we simply have to find the delays which independently maximize the output of each of the matched filters and then calculate the associated attenuation. This maximization procedure guarantees that $U(\hat{\underline{\theta}}^{(n+1)}, \hat{\underline{\theta}}^{(n)}) > U(\hat{\underline{\theta}}^{(n)}, \hat{\underline{\theta}}^{(n)})$, and so by the properties of the EM algorithm, $\mathcal{L}(\hat{\underline{\theta}}^{(n+1)}) > \mathcal{L}(\hat{\underline{\theta}}^{(n)})$; i.e., the log-likelihood function based on the observed data is increased in each iteration until convergence.

We recall that in our equalization algorithm, after one step through this EM algorithm, the new estimate of the channel parameters is fed back to the Viterbi algorithm that finds the MLSE to re-estimate the transmitted data.

ACKNOWLEDGMENT

REFERENCES

[1] A. B. Baggeroer, "Acoustic telemetry—an overview," *IEEE J. Oceanic Eng.*, vol. OE-9, pp. 229–235, Oct. 1983.

[2] R. Coates and P. Williston, "Underwater acoustic communications: A review and bibliography," in *Proc. Instit. Acoust.*, Dec. 1987.

[3] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, 1982.

[4] L. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 483–501, 1987.

[5] J. A. Catipovic, "Bandwidth-efficient trellis-coded modulation for phase random Rayleigh fading channels," *IEEE J. Oceanic Eng.*, to be published.

[6] J. G. Proakis, "Advances in equalizations for intersymbol interference," in *Advances in Communications Systems*, vol. 4, A. J. Viterbi, Ed. New York: Academic, 1975.

[7] S. U. H. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, pp.1348–1388, Sept. 1985.

[8] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.

[9] R. W. Lucky, "Automatic equalization for digital communications," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, 1965.

[10] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.

[11] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp.1692–1716, 1975.

[12] D. Godard, "Channel equalization using a Kalman filter for fast data transmission," *IBM J. Res. Develop.*, pp. 267–273, May 1974.

[13] D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-26, pp. 1439–1446, 1978.

[14] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Commun.*, vol. COM-27, pp. 899–905, 1979.

[15] E. H. Satorius and J. D. Packs, "Application of least-squares lattice algorithms to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-29, pp. 136–142, Feb. 1981.

[16] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of inter-symbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, 1972.

[17] M. A. Austin, "Decision-feedback equalization for digital communications over dispersive channels," MIT Lincoln Lab., Lexington, MA, Tech. Rep. 437, Aug. 1967.

[18] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 56–64, Jan. 1971.

[19] D. A. George, R. R. Bowen, and J. R. Storey, "An adaptive decision-feedback equalizer," *IEEE Trans. Commun.*, vol. COM-19, pp. 281–293, June 1971.

[20] W. U. Lee and F. S. Hill, "A maximum likelihood sequence estimator with decision feedback equalization," *IEEE Trans. Commun.*, vol. COM-25, pp. 971–979, Sept. 1977.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, Ser. 39, pp. 1–38, 1977.

[22] M. Feder, "Iterative algorithms for parameter estimation with applications to signal processing," Ph.D. thesis, MIT, Cambridge, MA, 1987.

[23] S. M. Flatte, "Wave propagation through a random media," *Proc. IEEE*, vol. 71, pp. 1267–1294, 1983.

[24] T. F. Duda, S. M. Flatte, and D. B. Cramer, "Modeling meter-scale acoustic intensity fluctuations from oceanic fine structure and microstructure," *J. Geophys. Res.*, vol. 93, pp. 5130–5142, 1988.

[25] J. A. Catipovic, "Performance limitations in underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, vol. 15, pp. 205–216, July 1990.

[26] J. A. Catipovic and A. B. Baggeroer, "Analysis of high-frequency multitone transmission propagation in the marginal ice zone," *J. Acoust. Soc. Amer.*, vol. 88, no. 1, pp. 185–190, July 1990.

[27] M. Suzuki, K. Nemoto, T. Tsuchiya, and T. Nakanishi, "Digital acoustic telemetry of color video information," in *Proc. Oceans'89*, Sept. 1989, pp. 893–896.

[28] L. Berkhovski and Y. Lysanov, *Fundamentals of Ocean Acoustics*. New York: Springer–Verlag, 1982.

[29] J. A. Catipovic and L. Freitag, "Spatial diversity processing for underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, this issue, pp. 86–97.

[30] J. L. Spiesberger, P. J. Bushong, K. Metzger, Jr., and T. G. Birdsall, "Ocean acoustic tomography: Estimating the acoustic travel time with phase," *IEEE J. Oceanic Eng.*, vol. 14, pp. 108–119, Jan. 1989.

[31] S. J. Roberts, "An echo cancelling technique applied to underwater acoustic data link," Ph.D. thesis, Herriot–Watt Univ., Edinburgh, Scotland, 1983.

[32] D. R. Carmichael and R. M. Dunbar, "Adaptive estimation of underwater acoustic channel using real test data," in *Proc. 6th Int. Symp. Unmanned, Untethered Submersible Techn.* (Baltimore, MD), June 1989, pp. 893–896.

[33] F. Ling and J. G. Proakis, "Adaptive lattice decision-feedback equalizers," *IEEE Trans. Commun.*, vol. COM-33, pp. 348–356, 1985.

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

[35] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, NJ: Prentice–Hall, 1969.

[36] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison–Wesley, 1984.

[37] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," Ph.D. thesis, MIT, Cambridge, MA, Sept. 1982.

[38] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 477–489, Apr. 1988.

[39] J. A. Catipovic and L. E. Freitag, personal communications.

[40] M. Feder, E. Weinstein, and A. V. Oppenheim, "A new class of sequential and adaptive algorithms with application to noise cancellation," in *Proc. 1988 Int. Conf. Acoust., Speech, Signal Processing*, 1988.

[41] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback–Leibler information measure," *IEEE Trans. Acoust. Speech, Signal Processing*, to be published.

[42] R. E. Bellman and S. E. Dreyfus, *Applied Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1962.

[43] G.D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, 1973.

Meir Feder received the B.Sc. and M.Sc. degrees (summa cum laude) from Tel-Aviv University, Tel-Aviv, Israel, and the Sc.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, and the Woods Hole Oceanographic Institution, Woods Hole, MA, all in electrical engineering, in 1980, 1984, and 1987, respectively.

During 1987–1988 he was a Research Associate in the Department of Electrical Engineering and Computer Science at MIT. He then worked in Elbit Computers, Haifa, Israel. In October 1989 he joined the faculty of the Department of Electrical Engineering—Systems, Tel-Aviv University. He was also a Visiting and Guest Investigator at the Woods Hole Oceanographic Institution in the summers of 1983 and 1988–1990. His research interests include signal and image processing, sonar signal processing, data and signal compression, and topics in information theory.

Josko A. Catipovic (S'81–M'85) received the B.S. degree in electrical engineering and the B.S. degree in ocean engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1981, where he also received the Sc.D. degree in oceanographic engineering in 1987.

He is currently an Assistant Scientist in the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA. His research interests include underwater data transmission, acoustic channel modeling, acoustic imaging, and application of estimation and information theory to ocean engineering problems.