

Minimum Cross-Entropy Spectral Analysis of Multiple Signals

RODNEY W. JOHNSON AND JOHN E. SHORE, SENIOR MEMBER, IEEE

Abstract—This paper presents a new information-theoretic method for simultaneously estimating a number of power spectra when a prior estimate of each is available and new information is obtained in the form of values of the autocorrelation function of their sum. One application of this method is the separate estimation of the spectra of a signal and additive noise, based on autocorrelations of the signal plus noise. A derivation of the method from the principle of minimum cross entropy is given, and the method is compared to minimum cross-entropy spectral analysis, of which it is a generalization. Some basic mathematical properties are discussed. Three numerical examples are included, two based on synthetic spectra, and one based on actual speech data.

I. INTRODUCTION AND BACKGROUND

WE present here an information-theoretic method for simultaneously estimating a number of power spectra when a prior estimate of each is available and new information is obtained in the form of values of the autocorrelation function of their sum. The method applies, for instance, when one obtains autocorrelation measurements for a signal with independent additive interference and one has some prior knowledge concerning the signal and the noise spectra; the result is signal- and noise-spectrum estimates that take both the prior estimates and the autocorrelation information into account. One thus obtains a procedure for noise suppression that offers some advantages over more traditional procedures, such as those based on spectral subtraction.

The present method is a generalization of minimum cross-entropy spectral analysis [1], which is in turn a generalization of maximum entropy (or linear-predictive or autoregressive) spectral analysis [2], [3]. All of these methods proceed from autocorrelation values. Minimum cross-entropy spectral analysis (MCESA) differs from maximum entropy spectral analysis (MESA) in that it explicitly uses a prior estimate of the power spectrum; it reduces to MESA as a special case when the prior estimate is uniform and one of the given autocorrelation values is for zero lag. The present method, multisignal MCESA, differs from MCESA in that it treats an arbitrary number of independent spectra simultaneously; in the special case of a single spectrum, it becomes identical to MCESA.

MESA may be regarded as an application of the principle of maximum entropy [4], [5]; single- and multisignal MCESA are applications of a generalization of that principle, the principle of minimum cross entropy (also called minimum discrimination information, directed divergence, I -divergence, relative

entropy, or Kullback-Leibler number) [6]–[11]. In the remainder of this section, we describe single- and multisignal MCESA further, and include some background on the principle of minimum cross entropy. For a comparison of these spectrum-analysis methods with MESA, see [1] and [12]. Section II contains a derivation of our multisignal estimator, and Section III discusses a few of its general properties. Section IV presents three numerical examples, one of which is based on measured samples of speech signals and noise. Section V contains a few remarks on algorithms. Finally, Section VI contains a concluding discussion.

A. Single- and Multisignal MCESA

MCESA addresses the following problem: estimate the power spectrum $S(f)$ of a real, band-limited, stationary process with bandwidth W , given values of the autocorrelation function

$$R(t) = 2 \int_0^W df S(f) \cos 2\pi ft$$

for finitely many lags $t = t_r$, $r = 0, \dots, M$, and given, in addition, a *prior* estimate P of S ; P may be thought of as the best guess at S we could make in the absence of autocorrelation data. The MCESA estimator has the form [1]

$$Q(f) = \frac{1}{1/P(f) + \sum_r 2\beta_r \cos 2\pi ft_r} \quad (1)$$

where the β_r are chosen so that Q satisfies the constraint that the autocorrelation function assume the given values:

$$R(t_r) = 2 \int_0^W df Q(f) \cos 2\pi ft_r. \quad (2)$$

We call Q the *posterior* estimate of S based on the prior estimate P and constraints (2). This estimator can be obtained directly from the minimum cross-entropy principle [1]; it can also be obtained by minimizing the Itakura-Saito distortion measure [13]

$$\int df \left(\frac{Q(f)}{P(f)} - \log \frac{Q(f)}{P(f)} - 1 \right)$$

subject to (2) [1]. When $P(f)$ is uniform, and one of the autocorrelation values is at lag zero (say, $t_0 = 0$), the constant $1/P$ can be absorbed into the coefficient β_0 . Thus, in this case,

Manuscript received March 17, 1981; revised January 26, 1982 and August 16, 1982.

The authors are with the Naval Research Laboratory, Washington, DC 20375.

the MCESA estimator (1) can be written in the usual MESA form, and MCESA reduces to MESA.

For multisignal MCESA, the problem is to estimate the power spectra $S_i(f)$ of a number of independent processes, given values of the total autocorrelation

$$R(t) = 2 \sum_i \int_0^W df S_i(f) \cos 2\pi ft$$

and a prior estimate P_i for each S_i . The estimator has the form

$$Q_i(f) = \frac{1}{1/P_i(f) + \sum_r 2\beta_r \cos 2\pi f t_r} \quad (3)$$

where the β_r are chosen so that the constraint equations

$$R(t_r) = 2 \sum_i \int_0^W df Q_i(f) \cos 2\pi f t_r \quad (4)$$

are satisfied. Note that the summation term in the denominator in (3) is independent of i . In Section III, we derive the estimates (3) directly from the principle of minimum cross entropy. We also show that they can be obtained by minimizing the sum

$$\sum_i \int df \left(\frac{Q_i(f)}{P_i(f)} - \log \frac{Q_i(f)}{P_i(f)} - 1 \right)$$

of Itakura-Saito distortions subject to the constraints (4). Equations (3) and (4) reduce to (1) and (2) when there is only one spectrum S_i . Thus, multisignal MCESA reduces to ordinary MCESA in case there is only one signal.

B. Cross-Entropy Minimization

The principle of minimum cross entropy is a general method for inference about probability distributions when information is available in the form of expectation values of known functions.

Let q^\dagger be a probability density on a space of states \mathbf{x} of some system. Suppose that q^\dagger is not known, but there is some *prior* density p (on the same space) that is our current estimate of q^\dagger . Now suppose we gain new information about q^\dagger in the form of expectation values

$$\int d\mathbf{x} q^\dagger(\mathbf{x}) g_r(\mathbf{x}) = \bar{g}_r \quad (5)$$

of known functions g_r . In general, these *constraints* do not determine q^\dagger uniquely: the equations (5) are satisfied by other densities q than q^\dagger (but not necessarily by p). The problem to be solved is, given p and the constraints (5), to make the best possible choice of a new (or *posterior*) estimate q of q^\dagger . The principle of minimum cross entropy states that one should choose that density q , among all the densities that satisfy the constraints, that has the least *cross entropy*

$$H(q, p) = \int d\mathbf{x} q(\mathbf{x}) \log (q(\mathbf{x})/p(\mathbf{x})) \quad (6)$$

with respect to p . For justifications of the principle, see [7], [14], and [12].

Given a positive prior probability density p , if there exists a

posterior q that minimizes the cross entropy and satisfies the constraints (5), it has the form

$$q(\mathbf{x}) = p(\mathbf{x}) \exp \left(-\lambda - \sum_r \beta_r g_r(\mathbf{x}) \right) \quad (7)$$

with the possible exception of a set of states on which the constraints imply that q vanishes [6, p. 38], [10]. In (7), λ and β_r are Lagrange multipliers whose values are determined by the normalization constraint

$$\int d\mathbf{x} q(\mathbf{x}) = 1 \quad (8)$$

and by the constraints (5), respectively. Conversely, if there are values for λ and β_r for which the constraints are satisfied, then the solution exists and is given by (7) [10]. Conditions for existence of solutions are given by Csiszár [10].

II. DERIVATION

We assume that the time-domain signal is a sum of stationary random processes $g_i(t)$, $i = 1, \dots, K$. In many applications, K will be 2—one signal process and one noise process—but the case of arbitrary K is no harder than $K = 2$, so we do the derivation in that generality. It is convenient to work with discrete-spectrum approximations to the g_i [1], [15]:

$$s_i(t) = \sum_{k=1}^N (a_{ik} \cos 2\pi f_k t + b_{ik} \sin 2\pi f_k t)$$

where the a_{ik} and the b_{ik} are random variables and the f_k are nonzero frequencies, not necessarily uniformly spaced. We write x_{ik} for the power of the process s_i at frequency $\pm f_k$,

$$x_{ik} = \frac{1}{4} (a_{ik}^2 + b_{ik}^2),$$

and will describe the process in terms of a joint probability density $q^\dagger(\mathbf{x}) = q^\dagger(x_1, \dots, x_K)$ where \mathbf{x}_i stands for (x_{i1}, \dots, x_{iN}) . The marginal density for each \mathbf{x}_i is defined by

$$q_i^\dagger(\mathbf{x}_i) = \int q^\dagger(\mathbf{x}) \prod_{j \neq i} d\mathbf{x}_j$$

where each component x_{jk} of the variables of integration ranges from 0 to ∞ .

Let $P_{ik} = P_i(f_k)$ be prior estimates of the power spectra of the s_i . Then we may take

$$p_i(\mathbf{x}_i) = \prod_{k=1}^N (1/P_{ik}) \exp(-x_{ik}/P_{ik}) \quad (9)$$

as prior estimates of $q_i^\dagger(\mathbf{x}_i)$. The assumed exponential form is equivalent to a Gaussian distribution in the amplitude variables a_{ik} and b_{ik} ; for justification of its use, see [1] and [16]. Note that the coefficients are chosen so that the expected value of the power x_{ik} of the process s_i at frequency f_k is equal to the prior estimate P_{ik} . Since we assume independence of \mathbf{x}_i and \mathbf{x}_j ($i \neq j$), our prior estimate of q^\dagger becomes

$$p(\mathbf{x}) = \prod_{i=1}^K \prod_{k=1}^N (1/P_{ik}) \exp(-x_{ik}/P_{ik}). \quad (10)$$

The autocorrelation function of each process s_i is given by

$$R_{ir} = \sum_{k=1}^N c_{rk} \int d\mathbf{x} q^\dagger(\mathbf{x}) x_{ik} \quad (11)$$

where $c_{rk} = 2 \cos 2\pi t_r f_k$. Suppose we obtain information about q^\dagger in the form of autocorrelation values for the sum of the s_i ,

$$R_r = \sum_{i=1}^K R_{ir}, \quad (12)$$

$r = 0, \dots, M$ where $t_0 = 0$. In view of (11), this has the form of linear constraints on expectation values of q^\dagger . We apply the principle of minimum cross entropy to these constraints and the prior equation (10). Following the steps that led to (9) in [1] yields a posterior estimate q of q^\dagger given by

$$q(\mathbf{x}) = \prod_{i=1}^K \prod_{k=1}^N A_{ik} \exp(-A_{ik} x_{ik}) \quad (13)$$

where

$$A_{ik} = 1/P_{ik} + \sum_r \beta_r c_{rk}$$

and the β_r are Lagrange multipliers corresponding to the constraints. The posterior estimate of the power spectrum of s_i is

$$Q_{ik} = \int d\mathbf{x} q(\mathbf{x}) x_{ik} = \frac{1}{A_{ik}};$$

thus,

$$Q_{ik} = \frac{1}{1/P_{ik} + \sum_r \beta_r c_{rk}} \quad (14)$$

where the β_r must be chosen so that the constraints

$$R_r = \sum_{i=1}^K \sum_{k=1}^N c_{rk} Q_{ik} \quad (15)$$

are satisfied. Equations (14) and (15) are simply discrete analogs of (3) and (4).

When p and q are given by (10) and

$$q(\mathbf{x}) = \prod_{i=1}^K \prod_{k=1}^N (1/Q_{ik}) \exp(-x_{ik}/Q_{ik})$$

[cf. (13)], the cross entropy (6) can be calculated explicitly:

$$H(q, p) = \sum_{i=1}^K \left[\sum_{k=1}^N \left(\frac{Q_{ik}}{P_{ik}} - \log \frac{Q_{ik}}{P_{ik}} - 1 \right) \right]. \quad (16)$$

The quantity in brackets is a discrete analog of the Itakura-Saito distortion measure [13], [17] of P_i with respect to Q_i ; cross-entropy minimization is thus equivalent to choosing the Q_i so as to minimize the sum of Itakura-Saito distortions. We obtain an alternative derivation of (14) by minimizing the right side of (16) directly, subject to the constraints (15). Namely, we form the expression

$$\sum_{i=1}^K \sum_{k=1}^N \left(\frac{Q_{ik}}{P_{ik}} - \log \frac{Q_{ik}}{P_{ik}} - 1 \right) + \sum_{r=0}^M \beta_r \sum_{i=1}^K \sum_{k=1}^N c_{rk} Q_{ik}$$

involving Lagrange multipliers β_r , and we set the partial derivative with respect to each Q_{ik} equal to zero:

$$1/P_{ik} - 1/Q_{ik} + \sum_{r=0}^M \beta_r c_{rk} = 0.$$

This yields (14).

III. PROPERTIES

In this section, we discuss three miscellaneous properties of our multisignal method. We call the first "order preservation"; briefly, it states that the method preserves the relative magnitudes of the priors. The second, "preservation of independence," is related to the assumption of statistical independence of the processes s_i ; it follows from a generalization of the property of cross-entropy minimization that was called "system independence" in [7] and [14]. The third is related to a phenomenon that we call "prior washout," and that occurs when a posterior resulting from one analysis is used as a prior for a subsequent analysis; we compare and contrast the behavior of the single- and multisignal methods in this situation.

A. Order Preservation

Let P_i and P_j be two prior spectra and let Q_i and Q_j be corresponding posterior spectra resulting from a multisignal MCESA analysis. The order-preservation property is the observation that for each frequency f_k we have $Q_i < Q_j$, $Q_i = Q_j$, or $Q_i > Q_j$ if and only if $P_i < P_j$, $P_i = P_j$, or $P_i > P_j$, respectively. This follows from the form of the representation of the Q_i in (3). The property accords well with intuition: if we expect *a priori* that s_i has greater power than s_j at frequency f_k , that expectation should not be altered by new information that concerns only the sum of the two powers.

B. Preservation of Independence

In (10), we wrote the prior probability density p in the form

$$p(\mathbf{x}) = \prod_{i=1}^K p_i(\mathbf{x}_i)$$

[cf. (9)] to reflect the initial assumption that the \mathbf{x}_i are independent. Preservation of independence is the property that the posterior density q has the same form,

$$q(\mathbf{x}) = \prod_{i=1}^K q_i(\mathbf{x}_i)$$

[cf. (13)], so that the \mathbf{x}_i remain independent after the prior density is replaced by the posterior. This posterior independence would be a simple consequence of the system independence property of [7] and [14] if the constraints were of the form

$$R_r = \int d\mathbf{x} q^\dagger(\mathbf{x}) g_r(\mathbf{x}_{i(r)}),$$

that is, if each constraint involved only one of the sets \mathbf{x}_i of variables (where which set was involved might depend on the constraint). System independence was one of the consistency axioms in [7]; informally, it states that it does not matter whether independent constraint information about separate systems with independent priors is accounted for separately, for each system, or jointly, by treating the system as one com-

posite system. In the present case, the constraints have the more general form [cf. (11) and (12)]

$$R_r = \int d\mathbf{x} q^\dagger(\mathbf{x}) \sum_{i=1}^K g_{ri}(\mathbf{x}_i);$$

each constraint involves a linear combination of functions, each involving one of the \mathbf{x}_i . Nevertheless, posterior independence still follows from prior independence in this more general case.

C. Prior Washout

The phenomenon we are here calling "prior washout" was mentioned in [14] in connection with "Property 14." Property 14, in slightly specialized form, states the following. Let p be a prior probability density. Let $I^{(1)}$ and $I^{(2)}$ be sets of constraints of the form (5), but with the right side replaced by $\bar{g}_r^{(1)}$ for $I^{(1)}$ and by $\bar{g}_r^{(2)}$ for $I^{(2)}$; that is $I^{(1)}$ and $I^{(2)}$ both constrain the expectations of the same set of functions g_r , but the expected values may differ. Let $q^{(1)}$ be the posterior density obtained from the prior density p by cross-entropy minimization subject to the constraints $I^{(1)}$, and let $q^{(2)}$ be the posterior density obtained when $q^{(1)}$ is taken as a new prior and cross entropy is minimized subject to the constraints $I^{(2)}$. Then the same posterior density $q^{(2)}$ is obtained when cross entropy is minimized subject to $I^{(2)}$, but p rather than $q^{(1)}$ is used as the prior. The effects of taking the constraint information $I^{(1)}$ into account are thus completely washed out when $I^{(2)}$ is taken into account.

One consequence of prior washout is a similar property of single-signal MCESA. For definiteness, consider a speech-processing system; say we wish to estimate the speech spectra $S^{(1)}, S^{(2)}, \dots$ in a succession of analysis frames, and we can measure the speech autocorrelations $R_r^{(1)}, R_r^{(2)}, \dots$ in these frames at a fixed set of lags r . Starting with a prior spectral estimate P , suppose we form a posterior estimate $Q^{(i)}$ for a frame i by taking the autocorrelation information $R^{(i)}$ into account. Suppose we then use this posterior $Q^{(i)}$ as a prior estimate for a later frame j and obtain a posterior estimate for that frame by taking $R^{(j)}$ into account. Prior washout implies that the result $Q^{(j)}$ is the same that we would have gotten if we had used P instead of $Q^{(i)}$ as the prior estimate for frame j ; taking $R^{(i)}$ into account completely washes out the effects of having taken $R^{(i)}$ into account.

This property has implications for certain noise-suppression schemes in which one might envision using MCESA. Suppose that additive noise is present in a speech-analysis system. It is often possible to detect whether or not speech is present in an analysis frame. If frame i is such a frame, then $Q^{(i)}$ is an estimate of the noise spectrum. Since the noise spectrum contains information about part of what is likely to be present in a later frame j that contains speech plus noise, it follows that using $Q^{(i)}$ as a prior for frame j might result in more accurate estimation of the total spectrum in that frame, thus allowing more accurate compensation for the noise, say by subtraction of the noise spectrum. (On the other hand, we might worry that this procedure would unduly enhance the noise component of the later estimate, thus further degrading the speech.) However, if the analyses of frames i and j are based on the same set of autocorrelation lags, prior washout occurs, and the use of $Q^{(i)}$

as a prior for frame j has no effect whatever on the result $Q^{(j)}$ of the analysis in frame j .

Although the same property holds for multisignal MCESA, a combination of single- and multisignal MCESA can be used to avoid prior washout and exploit the results of analyzing frames containing noise only. In particular, during a frame when speech is absent, obtain an estimated noise spectrum by a single-signal analysis. Use this spectrum as a prior noise estimate, together with some appropriate spectrum as a prior speech estimate, for a multisignal analysis in later frames. A procedure of this sort is illustrated in Section IV.

The reason that prior washout does not occur in this case is that the initial computation of the estimated noise spectrum uses constraints on noise autocorrelations values, while the subsequent computations use constraints on total autocorrelations; thus, different sets of functions are being constrained. In fact, let P_N be the prior used in obtaining the initial estimated noise spectrum $Q_N^{(1)}$ by single-signal MCESA. Then $Q_N^{(1)}$ has components at frequency f_k of the form

$$Q_{Nk}^{(1)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r c_{rk}}.$$

If $Q_N^{(1)}$ is used as a noise prior in later computations, and a spectrum P_S is used as a speech prior, the resulting noise and speech posteriors $Q_N^{(2)}$ and $Q_S^{(2)}$ have the form

$$Q_{Nk}^{(2)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r c_{rk} + \sum_r \beta'_r c_{rk}} \quad (17)$$

$$Q_{Sk}^{(2)} = \frac{1}{1/P_{Sk} + \sum_r \beta'_r c_{rk}}. \quad (18)$$

If P_N were used in place of $Q_N^{(1)}$ in the later computations, the resulting posteriors would have the form

$$Q_{Nk}^{(2)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r^* c_{rk}} \quad (19)$$

$$Q_{Sk}^{(2)} = \frac{1}{1/P_{Sk} + \sum_r \beta_r^* c_{rk}}. \quad (20)$$

Now, for linearly independent constraints, (17) and (19) are compatible only if $\beta_r^* = \beta_r + \beta'_r$ holds, and (18) and (20) are compatible only if $\beta_r^* = \beta'_r$ holds. Thus, the analog of prior washout will not, in general, occur here unless $\beta_r = 0$ holds, that is, unless $Q_N^{(1)} = P_N$.

IV. EXAMPLES

In this section, we present three numerical examples; in each, a given set of data is analyzed both by multisignal MCESA and by either single-signal MCESA or a conventional MESA method. In the first example, autocorrelations at a number of equally spaced lags are computed from the sum of a pair of assumed "true" spectra, and single- and multisignal MCESA estimates are obtained from them. The second example is based on the same set of assumed spectra as the first, but instead of exact autocorrelations computed from the spectra, we use autocorrelations estimated from samples of a process, gen-

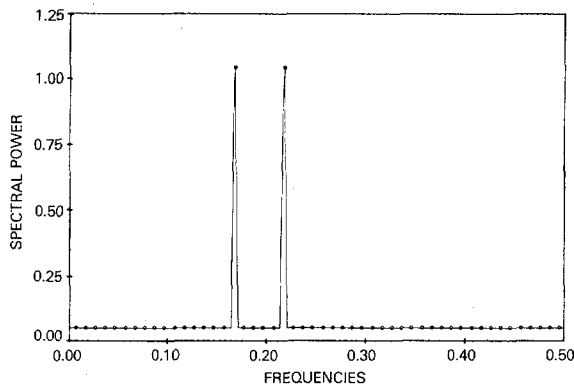


Fig. 1. Total assumed original spectrum (first two examples).

erated with the help of a random-number generator, whose theoretical power spectrum is the sum of those assumed. In the third, autocorrelations are estimated from sums of speech-signal and noise samples, and spectral estimates are obtained by MESA and multisignal MCESA.

The assumed original spectra for the first two examples are a pair S_B and S_S , which we think of as a known "background" component and an unknown "signal" component of the total spectrum. For numerical purposes, we use the spectral powers S_{Bk} and S_{Sk} at 100 equally spaced frequencies $f_k = \pm 0.005, \pm 0.015, \dots, \pm 0.495$ between -0.5 and $+0.5$ (the Nyquist band: we take the spacing between autocorrelation lags to be unity). The background consists of an approximation to white noise plus a peak corresponding to a sinusoid at frequency 0.215:

$$S_{Bk} = \begin{cases} 1.05, & f_k = \pm 0.215 \\ 0.05, & \text{otherwise.} \end{cases}$$

The signal term consists of a nearby, similar peak at frequency 0.165:

$$S_{Sk} = \begin{cases} 1, & f_k = \pm 0.165 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the total assumed spectrum $S_B + S_S$ is as shown (for positive frequencies) in Fig. 1. Here are the corresponding autocorrelations R_r at six lags $t_r = 0, 1, \dots, 5$:

| | | | | | | |
|-------|--------|--------|---------|---------|--------|---------|
| t_r | 0 | 1 | 2 | 3 | 4 | 5 |
| R_r | 9.0000 | 1.4544 | -2.7732 | -3.2248 | 0.2032 | 2.6900. |

For the multisignal calculation, we use a pair of prior spectral estimates P_B and P_S . Since we are assuming prior knowledge of the background spectral component S_B , we simply take $P_B = S_B$ as shown in Fig. 2. To reflect prior ignorance of the signal component S_S , we take P_S to be uniform as in Fig. 3; for this example, we have somewhat arbitrarily normalized P_S to have the same total power as P_B . For the single-signal calculation, we use $P = P_B + P_S$ as the prior spectral estimate.

Fig. 4 shows the result of the single-signal analysis—the

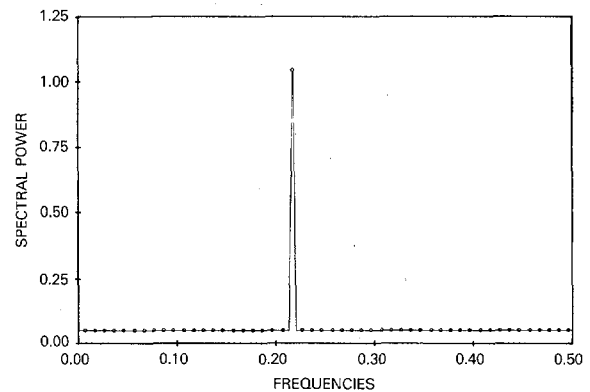


Fig. 2. Prior estimate of background spectrum: assumed original background spectrum (first two examples).

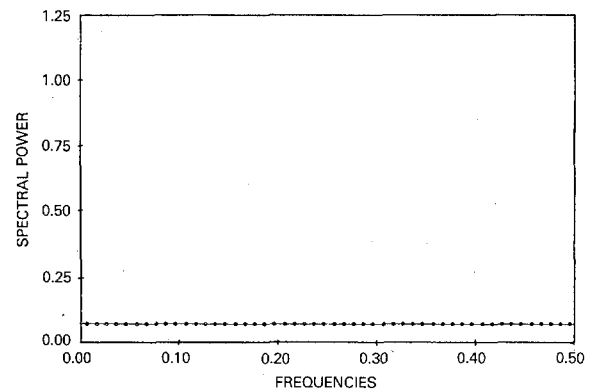


Fig. 3. Prior estimate of signal spectrum: uniform (first two examples).

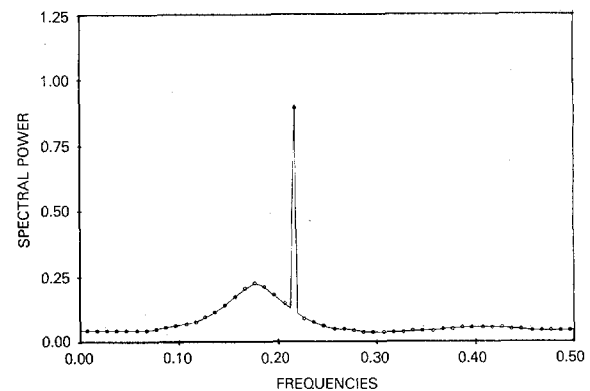


Fig. 4. Single-signal MCESA posterior estimate of total spectrum (first example).

MCESA posterior estimate Q obtained from the prior estimate P and autocorrelations R_r . Corresponding to the "known" peak at frequency 0.215 (which was included in the prior), there is a sharp peak in the posterior at that frequency; corresponding to the "unknown" peak at frequency 0.165, there is a maximum at approximately that frequency that is broader than the first, but resolvable from it.

The same original spectra S_B and S_S and the same autocorrelations R_r were used in an example in [1]. There a MESA and an MCESA spectral estimate were compared (see [1, Figs.

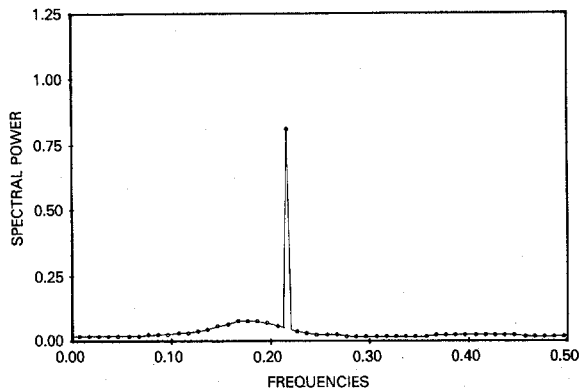


Fig. 5. Multisignal MCESA posterior estimate of background spectrum (first example).

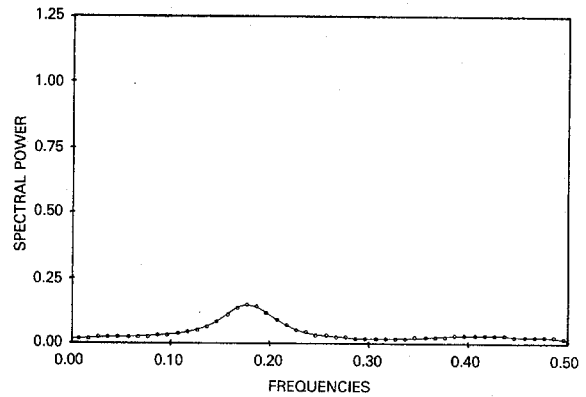


Fig. 6. Multisignal MCESA posterior estimate of signal spectrum (first example).

5 and 6]). The MESA estimate failed to resolve the two peaks, and showed a single maximum at about the midfrequency. The MCESA estimate was based on P_B instead of $P_B + P_S$ as a prior; the result differed from Fig. 4, but was qualitatively similar. In both cases, the MCESA estimate implies the presence of the signal at frequency 0.165, but does not provide a numerical estimate of the signal. Such an estimate is provided by multisignal MCESA.

The two individual posterior estimates Q_B and Q_S from the multisignal analysis are shown in Figs. 5 and 6. The sharp peak at frequency 0.215 is seen to be correctly assigned entirely to the background posterior Q_B —unsurprisingly, since it was present in the background prior, but not the signal prior. The broader maximum corresponding to the original peak at frequency 0.165 is present in the signal posterior Q_S and is also present, although less prominent, in Q_B . To understand why, qualitatively, consider that the autocorrelations depend only on the total spectrum; the autocorrelation constraints can be equally well satisfied by allocating spectral power near frequency 0.165 to Q_B or to Q_S . By the discussion in Section III, the relative magnitudes of the posteriors at each frequency depend on the relative magnitudes of the priors. Both P_B and P_S are flat near frequency 0.165, and because of the normalization chosen, P_S is somewhat greater there. Consequently, the broad maximum in Q_S is somewhat greater than that in Q_B .

For the second example, we generated a sum of pseudorandom processes whose theoretical power spectrum is $S_B + S_S$ (in the notation of the previous example). The white part of S_B was modeled by independent, zero-mean Gaussian pseudorandom variables with variance 5. The two peaks were modeled by sinusoids with fixed amplitude, but randomly chosen phase. For each of ten repetitions of the experiment, we generated 180 samples s_t , ($t = 1, \dots, 180$) and computed autocorrelation estimates R_r , ($r = 0, 1, \dots, 5$) by the formula

$$R_r = \frac{1}{180} \sum_{j=1}^{180-r} s_j s_{j+r}. \quad (21)$$

This is a biased estimate, but guarantees positive-definiteness. No additional windowing or filtering was used. Then, using

these estimated autocorrelations and the same prior spectra P, P_B, P_S as in the first example, we computed a single-signal MCESA posterior estimate Q for the total spectrum and multisignal MCESA posterior estimates Q_B, Q_S for the background and signal spectra.

The results of the single-signal analysis are shown in Fig. 7, in which the posterior spectra from the ten repetitions of the experiment are overlaid. Overlay plots of Q_B and Q_S from the same ten repetitions are shown in Figs. 8 and 9. Thus, Figs. 7, 8, and 9 are to be compared, respectively, to Figs. 4, 5, and 6 from the first example. The qualitative similarities are apparent.

The third example is based on time-domain samples of voiced speech and noise. The speech comprises a portion of an English sentence spoken by a male speaker and includes the first word, "Sue," of the sentence, together with silent segments before and after it. The noise consists of a segment of helicopter noise equal in duration to the speech. These were separately filtered, sampled, and digitized at 8000 samples/s. The speech and noise data were then added sample by sample, resulting in samples of noisy speech. These samples were segmented into analysis frames of 180 samples, and 11 autocorrelations R_r , $r = 0, 1, \dots, 10$ were estimated for each frame by (21) where s_j is the j th sample in the frame.

The last frame before the actual beginning of the word was selected; this frame of "noisy speech" thus consisted entirely of noise. From the autocorrelation estimates for this frame, a uniform-prior, single-signal MCESA spectral estimate was computed for use as a prior estimate of the noise spectrum in subsequent frames. This should be essentially equivalent to a conventional MESA estimate of the noise spectrum as obtained by the Yule-Walker method. A uniform spectrum was used as a prior estimate for the speech spectrum in the subsequent frames. These two priors are shown in Fig. 10. Much of the noise power is concentrated in a peak near 2780 Hz.

From the two priors and the autocorrelation estimates, multisignal MCESA estimates of the speech and noise spectra were computed for later frames. From the autocorrelation estimates, MESA (LPC) spectral estimates were computed for the noisy speech. We present the results for a selected frame of

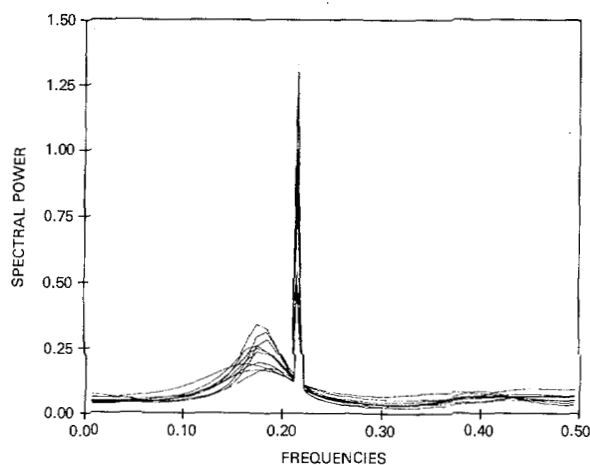


Fig. 7. Single-signal MCESA posterior estimates of total spectrum (second example).

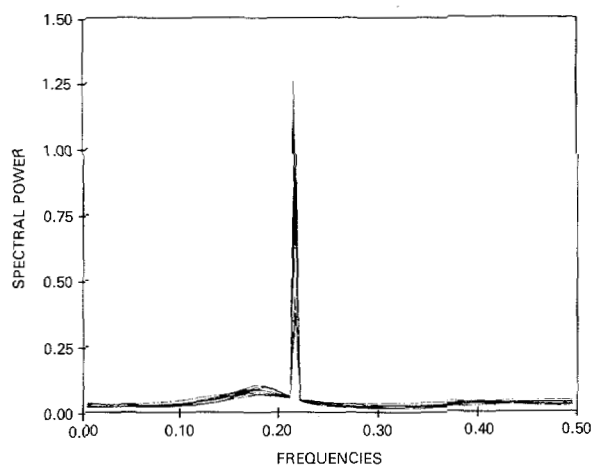


Fig. 8. Multisignal MCESA posterior estimates of background spectrum (second example).

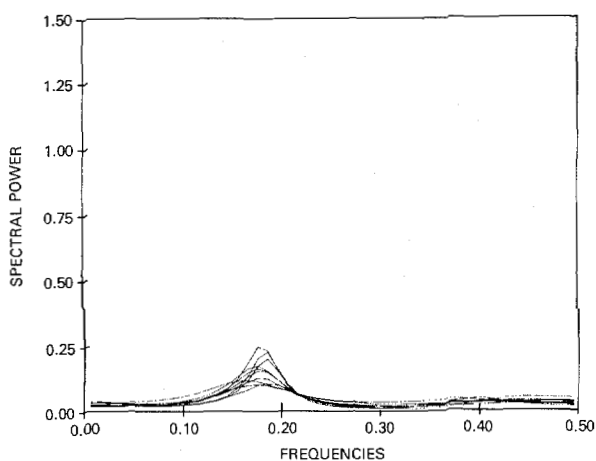


Fig. 9. Multisignal MCESA posterior estimates of signal spectrum (second example).

voiced speech—the second of seven that span the vowel /u/. For comparison to these results, we present in Fig. 11 a MESA estimate of the uncorrupted speech. This was computed exactly like the MESA estimate for the noisy speech, except that the R_T were estimated from the speech samples only, not

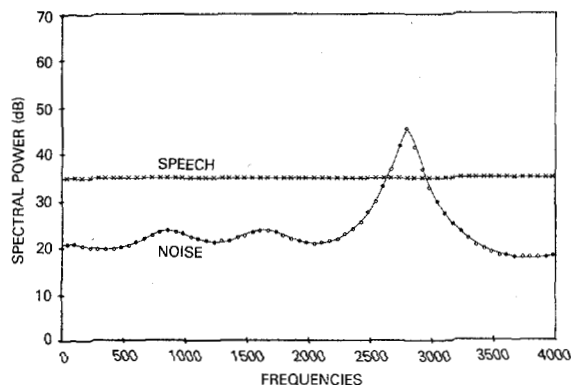


Fig. 10. Prior estimates of speech and noise spectra (third example).

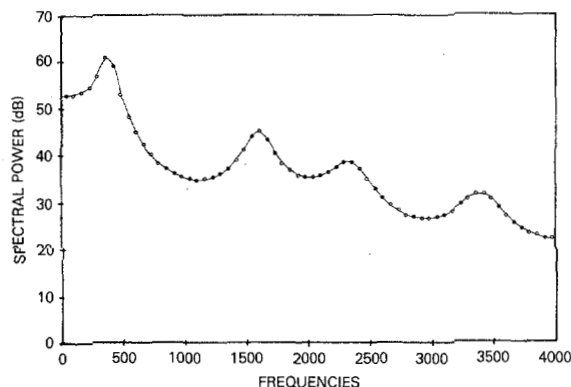


Fig. 11. MESA estimate of speech spectrum from noise-free data (third example).

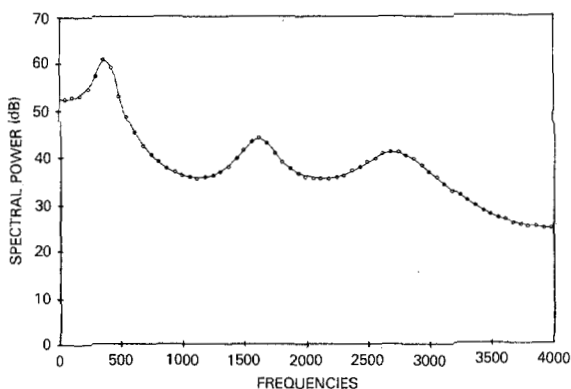


Fig. 12. MESA estimate of total spectrum (third example).

from the sums of speech and noise samples.

The MESA estimate for the noisy speech is shown in Fig. 12. This spectrum agrees rather well with the noise-free estimate in the band from 0 up to about 2000 Hz, which includes the first two formants. Above 2000 Hz, however, there is only a single maximum; the third and fourth formants have merged with the peak in the noise spectrum to form a single peak at about 2690 Hz.

We subtracted the noise prior (Fig. 10) from this result (Fig. 12). The difference, shown in Fig. 13, represents an attempt to estimate the speech spectrum by a MESA analysis and spectral subtraction. The subtracted MESA spectrum is fairly close to the unsubtracted MESA spectrum, except in the neighbor-

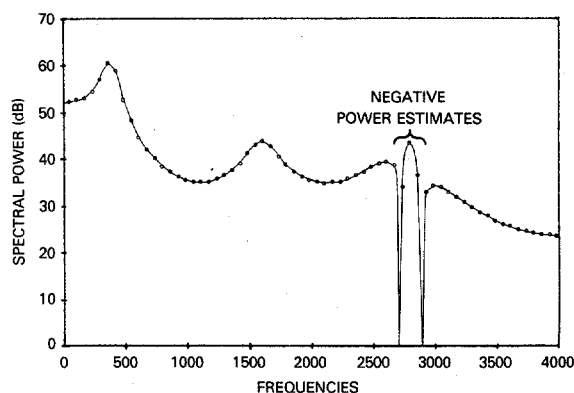


Fig. 13. Result of subtracting noise prior (see Fig. 10) from spectrum in Fig. 12 (third example).

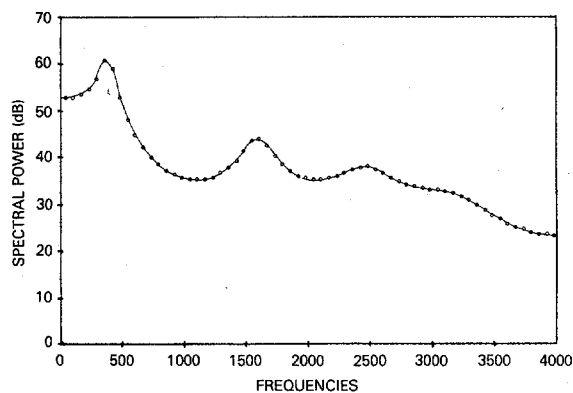


Fig. 14. Multisignal MCESA posterior estimate of speech spectrum (third example).

hood of the noise peak at 2780 Hz. Near that frequency, the subtraction so far overcompensates that the difference actually assumes rather large negative values. (Absolute values are plotted in the figure.)

The multisignal MCESA posteriors are shown in Figs. 14 and 15; Fig. 14 is the speech, and Fig. 15 is the noise. Fig. 15 shows a maximum near 2440 Hz, about 130 Hz higher than the third formant, and a suggestion of the fourth formant is discernible. Except for frequencies near the noise peak, the multisignal speech spectrum (Fig. 14) and the subtracted MESA result (Fig. 13) are quite close, the multisignal result usually being the closer of the two to the estimate based on noise-free data (Fig. 11). Near 2780 Hz, the multisignal result is substantially closer, and where the subtracted MESA becomes negative, the multisignal estimate only takes physically meaningful positive values. Both methods underestimate the total power near 2780 Hz (cf. Fig. 15); however, the multisignal method apportions the total between speech and noise in a somewhat reasonable way, whereas the other does not.

V. ALGORITHMS

The computations for the examples in Section IV were done with a program that uses the Newton-Raphson method to find values for the β_r in (14) such that (15) is satisfied. The program is fully general in that neither the frequencies f_k nor the lags t_r need be equispaced. It is also, consequently, much too slow to be practical, except for small numbers of examples. FFT methods are precluded, and the program, in fact, runs through several iterations, each time passing between the time and frequency domains by what amounts to a slow Fourier transformation.

We have not yet tried to find the most efficient possible multisignal MCESA algorithm, but we have recently written an algorithm that is considerably faster than the one just mentioned for input data that satisfy two assumptions: the autocorrelation lags are equispaced, and the prior spectra are all of the MESA form [(1) without the term $1/P(f)$ in the denominator]. The assumptions imply that the posterior spectra are also of the MESA form. They also permit us to avoid the frequency domain altogether: all the spectra can be represented by various sets of standard LPC parameters such as autoregressive coefficients or reflection coefficients. Speech output

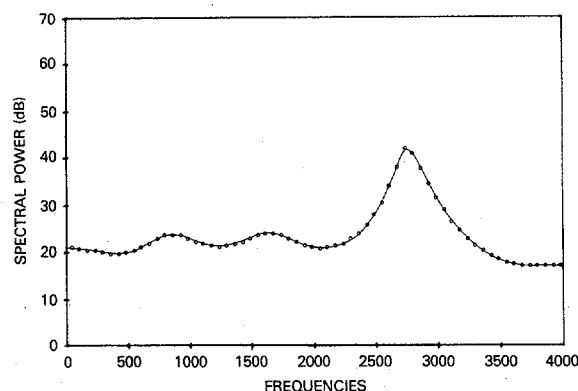


Fig. 15. Multisignal MCESA posterior estimate of noise spectrum (third example).

can be obtained in a form suitable for driving a conventional LPC speech synthesizer. A 12th-order, two-signal computation takes about 7 s/frame on a PDP 11/45™ equipped with floating-point hardware.

We have recently been able to demonstrate effective noise suppression in noisy speech by processing entire sentences and listening to the synthesized output. The results will be reported in a future paper.

VI. DISCUSSION AND CONCLUSIONS

Multisignal MCESA is a new spectrum-estimation method based on a provably optimal information-theoretic inductive-inference procedure. When separate prior estimates are available for the power spectra of two or more processes, and new information is obtained in the form of values of the autocorrelation function of their sum, the method yields separate posterior estimates. One application is separating the spectrum of a signal from that of additive noise. Preliminary experiments with speech synthesis are encouraging. By incorporating prior estimates for both signal and noise spectra, the multisignal method offers considerable scope and flexibility for tailoring an estimator to the characteristics of a signal or noise.

In the third example in Section IV, we contrasted this method with a more ad hoc method for taking a prior noise

™Trademark of Digital Equipment Corporation.

estimate into account—estimate the sum of signal and noise spectra from autocorrelations and then subtract the prior noise estimate. The latter method seems to imply an unwarranted absolute commitment to the noise-spectrum estimate: adjustments to the signal-spectrum estimate are made solely responsible for fitting the autocorrelation of the sum to measured values. The multisignal method, by contrast, adjusts both noise and signal estimates in fitting the autocorrelation of the sum. We saw that the multisignal method could thereby avoid nonphysical (negative) estimates that can result from spectral subtraction.

In the same example, a prominent noise peak was present in the sum spectrum. Most of the power in it was properly attributed to the noise spectrum in the posteriors, but substantial leakage a few decibels lower into the signal (speech) spectrum occurred. The relative apportionment of the power in that peak between the signal and noise posteriors would be substantially altered by a change in the level of the uniform spectrum that was used as the speech prior. This is in contrast to single-signal MCESA where all uniform priors give equivalent results (as long as one of the constrained autocorrelations is the total power). How best to choose the level of this uniform prior relative to the noise prior is a question not yet answered. Indeed, since the signal is known to be speech, it would undoubtedly be beneficial to replace the uniform signal prior with one tailored to the characteristics of speech. How best to do this tailoring is another unanswered question. In short, there is much to be learned about how to choose the prior estimates to reflect our prior knowledge of signals and noise in practical situations.

ACKNOWLEDGMENT

We thank G. Kang for supplying the speech and noise data on which the third numerical example was based, and for a number of helpful discussions. We are indebted to J. Buck for writing the programs that implement the algorithm mentioned in Section V, and for much other useful software.

REFERENCES

- [1] J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 230-237, Apr. 1981.
- [2] J. P. Burg, "Maximum entropy spectral analysis," presented at the 37th Annu. Meet. Soc. of Exploration Geophys., Oklahoma City, OK, 1967.
- [3] —, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975, University Microfilms 75-25,499.
- [4] W. M. Elsasser, "On quantum measurements and the role of the uncertainty relations in statistical mechanics," *Phys. Rev.*, vol. 52, pp. 987-999, Nov. 1937.
- [5] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, pp. 227-241, 1968.
- [6] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959, and Dover, 1969.
- [7] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26-37, Jan. 1980.

- [8] I. J. Good, "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables," *Ann. Math. Statist.*, vol. 34, pp. 911-934, 1963.
- [9] A. Hobson and B. Cheng, "A comparison of the Shannon and Kullback information measures," *J. Statist. Phys.*, vol. 7, pp. 301-310, 1973.
- [10] I. Csiszar, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.*, vol. 3, pp. 146-158, 1975.
- [11] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964.
- [12] R. Johnson and J. Shore, "Minimum-cross-entropy spectral analysis of multiple signals," Naval Res. Lab., Washington, DC, NRL Memo. Rep. 4492, Apr., 1981.
- [13] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Rep. 6th Int. Congr. Acoust.*, Y. Yonsai, Ed., Tokyo, Japan, 1968.
- [14] J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 472-482, July 1981.
- [15] Y. Yaglom, *An Introduction to the Theory of Stationary Random Functions*. Englewood Cliffs, NJ: Prentice-Hall, 1962.
- [16] J. E. Shore, "Minimum cross-entropy spectral analysis," Naval Res. Lab., Washington, DC, NRL Memo. Rep. 3921, Jan., 1979.
- [17] R. M. Gray, A. Buzo, A. H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 367-376, Aug. 1980.



Rodney W. Johnson was born in Scottsbluff, NE, on March 8, 1940. He received the B.S. degree in physics from Stanford University, Stanford, CA, in 1962, and the Ph.D. degree in mathematics from Stanford University in 1967.

He held Research Associateships in Physics at Princeton University, Princeton, NJ, from 1966 to 1968 and at Syracuse University, Syracuse, NY, from 1968 to 1970. From 1970 to 1973 he was an Assistant Professor in the Department of Mathematics, Fordham University, New York, NY. Following a year at the Center for Naval Analyses, he joined the Naval Research Laboratory, Washington, DC, in 1974, where he is a member of the Computer Science Section in the Computer Science and Systems Branch of the Information Technology Division. His current and former interests include information theory, signal processing, speech synthesis, and mathematical physics.



John E. Shore (M'72-SM'82) was born in England on September 2, 1946. He received the B.S. degree in physics from Yale University, New Haven, CT, in 1968, and the Ph.D. degree in theoretical physics (statistical mechanics) from the University of Maryland, College Park, in 1974.

In 1968 he joined the Naval Research Laboratory, Washington, DC, where he is currently Head of the Computer Science Section in the Computer Science and Systems Branch of the Information Technology Division. His previous research interests include computer architecture, dynamic memory allocation, programming language design, software engineering, and text-to-speech translation. His current interests include information theory, queueing theory, system modeling, pattern recognition, spectral analysis, and speech processing.