

Iterative and Sequential Algorithms for Multisensor Signal Enhancement

Ehud Weinstein, *Fellow, IEEE*, Alan V. Oppenheim, *Fellow, IEEE*, Meir Feder, *Senior Member, IEEE*, and John R. Buck, *Student Member, IEEE*

Abstract—In problems of enhancing a desired signal in the presence of noise, multiple sensor measurements will typically have components from both the signal and the noise sources. When the systems that couple the signal and the noise to the sensors are unknown, the problem becomes one of joint signal estimation and system identification. In this paper, we specifically consider the two-sensor signal enhancement problem in which the desired signal is modeled as a Gaussian autoregressive (AR) process, the noise is modeled as a white Gaussian process, and the coupling systems are modeled as linear time-invariant finite impulse response (FIR) filters. Our primary approach consists of modeling the observed signals as outputs of a stochastic dynamic linear system, and we apply the Estimate-Maximize (EM) algorithm for jointly estimating the desired signal, the coupling systems, and the unknown signal and noise spectral parameters. The resulting algorithm can be viewed as the time-domain version of our previously suggested frequency-domain approach [4], where instead of the noncausal frequency-domain Wiener filter, we use the Kalman smoother. This time-domain approach leads naturally to a sequential/adaptive algorithm by replacing the Kalman smoother with the Kalman filter, and in place of successive iterations on each data block, the algorithm proceeds sequentially through the data with exponential weighting applied to allow adaptation to nonstationary changes in the structure of the data. A computationally efficient implementation of the algorithm is developed by exploiting the structure of the Kalman filtering equations. An expression for the log-likelihood gradient based on the Kalman smoother/filter output is also developed and used to incorporate efficient gradient-based algorithms in the estimation process.

I. INTRODUCTION

IN problems of enhancing a desired signal in the presence of noise, multiple sensor measurements will typically have components from both the signal and the noise sources. Since the systems that couple the signal and the noise to the sensors

are unknown, one must deal with the more difficult problem of joint signal estimation and system identification.

An approach to the two-sensor signal enhancement problem is presented in [4]. In that approach, the desired (speech) signal is modeled as an autoregressive (AR) Gaussian process, the noise is modeled as a white Gaussian process, and the coupling systems are modeled as linear time-invariant finite impulse response (FIR) filters. The problem is formulated as a maximum likelihood (ML) estimation problem, and the iterative estimate-maximize (EM) algorithm is applied for its solution. The resulting algorithm consists of iteratively applying the Wiener filter to the two-sensor data to estimate the signal and the noise and using these estimates to identify the coupling systems and the unknown spectral parameters of the signal and the noise. When applied to speech signals in room acoustic environment, this algorithm has shown encouraging results that improve upon Widrow's LMS noise cancellation method [21].

In order to deal with the nonstationarity of the signal, the noise, and the coupling systems, it is suggested in [4] that the algorithm be applied on consecutive time frames using a sliding window. This approach involves two contradicting requirements: The window should be short enough so that the algorithm will respond to nonstationary changes in the signal and noise statistics. However, the window should be long in order to improve the statistical stability of the resulting signal and parameter estimates and in order to obtain a computationally tractable algorithm in which noncausal frequency-domain Wiener filtering can be applied.

In this paper, we present a time-domain approach to the two-sensor signal enhancement problem based on the development in [20]. In this approach, we model the observed signals as outputs of an unknown stochastic dynamic linear system and apply the EM algorithm for jointly estimating the signal, the noise, the coupling systems, and the unknown signal and noise spectral parameters. The proposed algorithm is similar in structure to that in [4]. However, in place of the noncausal Wiener filter, we employ a Kalman smoother. In this way, many of the computational and conceptual difficulties associated with the prior frequency-domain approach in [4] are avoided. Furthermore, the time-domain formulation leads naturally to a sequential/adaptive algorithm by replacing the Kalman smoother with the Kalman filter, and in place of successive iterations on each data block, the algorithm proceeds sequentially through the data with exponential weighting applied to allow adaptation to nonstationary changes in the

Manuscript received September 23, 1992; revised March 15, 1993. The associate editor coordinating the review of this paper and approving it for publication was Dr. James Zeidler. This paper was supported by the Wolfson Research Awards administered by the Israel Academy of Sciences and Humanities, by the U.S. Air Force-Office of Scientific Research under Grant number AFOSR-91-0034, and by the Office of Naval Research under Grant numbers N00014-91-J-1628 and N00014-93-1-0686.

E. Weinstein is with the Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv, Israel; he is also with the Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

A. V. Oppenheim and J. R. Buck are with the Department of Electrical Engineering and Computer Science and with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

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

IEEE Log Number 9214649.

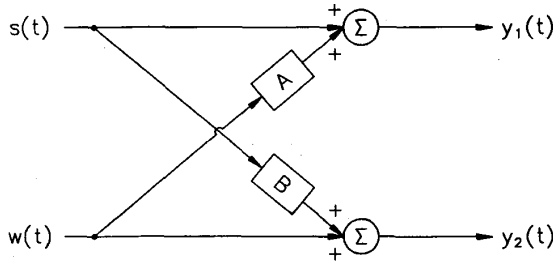


Fig. 1. Signal enhancement problem.

structure of the data. A computationally efficient implementation of the algorithm is developed by exploiting the structure of the Kalman filtering equations. An expression for the log-likelihood gradient based on the Kalman smoother/filter output is also developed and used to efficiently incorporate gradient-based algorithms in the estimation process.

We note that the approach in this paper is not restricted to the two-sensor signal enhancement problem. As shown in [20] and [13], respectively, this approach can also be applied to single-sensor signal-enhancement and noise-cancellation problems. More generally, it can be applied to multisensor signal enhancement problems in which the number of sensors is not necessarily equal to the number of signal sources.

II. PROBLEM FORMULATION

The basic problem of interest is illustrated in Fig. 1. We assume that a desired signal source and a noise source exist in some environment. We want to install two sensors in such a way that one of them (the primary sensor) measures the desired signal while the other one (the reference sensor) measures the noise. However, the signal and the noise are both coupled into each sensor through the environment. To simplify the exposition, we shall assume that the coupling systems A and B are causal linear time-invariant FIR filters, in which case the signals $z_1(t)$ and $z_2(t)$ measured at the two sensor outputs are modeled as

$$z_1(t) = s(t) + \sum_{k=0}^q a_k w(t-k) + e_1(t) \quad (1)$$

$$z_2(t) = w(t) + \sum_{k=0}^r b_k s(t-k) + e_2(t) \quad (2)$$

where $s(t)$ is the desired signal, $w(t)$ is the noise, $\{a_k\}_{k=0}^q$ and $\{b_k\}_{k=0}^r$ are the unit sample response coefficients of A and B , respectively, and the additional noise sources $e_1(t)$ and $e_2(t)$ are included to represent modeling errors, sensor noise, and measurement noise. The independent variable t represents normalized sampling time.

We shall assume that $e_1(t)$ and $e_2(t)$ are statistically independent zero-mean white Gaussian processes with variances g_1 and g_2 , respectively. The desired signal $s(t)$ is modeled as an AR process of order p , satisfying the difference equation

$$s(t) = -\sum_{k=1}^p \alpha_k s(t-k) + \sqrt{g_s} u_s(t) \quad (3)$$

and the noise $w(t)$ is modeled by

$$w(t) = \sqrt{g_w} u_w(t) \quad (4)$$

where $u_s(t)$ and $u_w(t)$ are normalized (zero-mean, unit variance) white Gaussian processes. We assume that $u_s(t)$, $u_w(t)$, $e_1(t)$, and $e_2(t)$ are mutually independent.

We shall find it convenient to define:

$$\mathbf{s}_r(t) = [s(t-r) \ s(t-r+1) \ \cdots \ s(t)]^T, \quad (5)$$

$$\mathbf{w}_q(t) = [w(t-q) \ w(t-q+1) \ \cdots \ w(t)]^T, \quad (6)$$

$$\boldsymbol{\alpha} = [\alpha_p, \ \alpha_{p-1}, \ \cdots, \ \alpha_1]^T, \quad (7)$$

$$\mathbf{a} = [a_q, a_{q-1}, \ \cdots, a_0]^T \quad (8)$$

and

$$\mathbf{b} = [b_r, b_{r-1}, \ \cdots, b_0]^T. \quad (9)$$

Then, (1) and (2) can be written in the form

$$z_1(t) = s(t) + \mathbf{a}^T \mathbf{w}_q(t) + e_1(t) \quad (10)$$

$$z_2(t) = w(t) + \mathbf{b}^T \mathbf{s}_r(t) + e_2(t) \quad (11)$$

and (3) can be written in the form

$$s(t) = -\boldsymbol{\alpha}^T \mathbf{s}_{p-1}(t-1) + \sqrt{g_s} u_s(t). \quad (12)$$

Denote by $\boldsymbol{\theta}$ the vector of unknown parameters:

$$\boldsymbol{\theta} = [\boldsymbol{\alpha}^T \ \mathbf{a}^T \ \mathbf{b}^T \ g_s g_w g_1 g_2]^T. \quad (13)$$

We note that some of the components of $\boldsymbol{\theta}$, e.g., some of the signal variances, may be known *a priori*, depending on the application.

Given the observed data

$$\mathbf{z} = \{z_1(t), z_2(t) : t = 1, 2, \dots, N\} \quad (14)$$

we want to find the best possible estimate of the desired signal $s(t)$. If one interprets "best" in the usual sense of minimizing the mean square error (mse), the optimal signal estimate is obtained by performing the conditional expectation of $s(t)$ given the observed data \mathbf{z} . However, this conditional expectation requires prior knowledge of $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is unknown, we must deal with the more complicated problem of joint signal estimation and parameter identification.

One approach would be to compute the ML estimate of $\boldsymbol{\theta}$ and use it to generate the signal estimate. The ML estimate $\boldsymbol{\theta}_{ML}$ of $\boldsymbol{\theta}$ is obtained by solving

$$\boldsymbol{\theta}_{ML} = \arg \max_{\boldsymbol{\theta}} \log f_Z(\mathbf{z}; \boldsymbol{\theta}) \quad (15)$$

where $\log f_Z(\mathbf{z}; \boldsymbol{\theta})$ is the log-likelihood, that is, the logarithm of the probability density of the observed data \mathbf{z} . Unfortunately, the maximization in (15) is a complicated multidimensional optimization that is very difficult to solve.

In the next section, we develop a computationally efficient iterative method based on the EM algorithm for solving the joint signal and parameter estimation problem indicated above.

III. SIGNAL ENHANCEMENT BASED ON THE EM ALGORITHM

The EM algorithm [3] is basically an iterative method for finding ML parameter estimates. It works with the notion of "complete" data and iterates between estimating the log likelihood of the complete data using the observed (incomplete) data and the current parameter estimate (E-step) and maximizing the estimated log-likelihood function to obtain the new parameter estimate (M-step).

More specifically, let \mathbf{y} denote the "complete" data related to the observed data \mathbf{z} by some noninvertible (many-to-one) transformation. Let $\theta^{(l)}$ denote the current estimate of θ after l iterations of the algorithm. Then, the next iteration cycle is specified in two steps as follows:

E-step: Compute

$$Q(\theta, \theta^{(l)}) = E_{\theta^{(l)}} \{\log f_{\mathbf{Y}}(\mathbf{y}; \theta) | \mathbf{z}\}. \quad (16)$$

M-step:

$$\max_{\theta} Q(\theta, \theta^{(l)}) \rightarrow \theta^{(l+1)} \quad (17)$$

where $\log f_{\mathbf{Y}}(\mathbf{y}; \theta)$ is the log-likelihood of \mathbf{y} , and $E_{\theta^{(l)}} \{\cdot | \mathbf{z}\}$ denotes the conditional expectation given $\mathbf{Z} = \mathbf{z}$ computed with respect to the current parameter estimate $\theta^{(l)}$.

The heuristic idea is that we want to choose θ to maximize the log likelihood of the complete data \mathbf{y} . However, since $\log f_{\mathbf{Y}}(\mathbf{y}; \theta)$ is not available to us (because the complete data is unavailable), we maximize instead its expectation given the observed data \mathbf{z} and the current parameter estimate $\theta^{(l)}$. Since we have used $\theta^{(l)}$ rather than the actual (true) value of θ , the conditional expectation is not exact. Thus, the algorithm iterates, using the current parameter estimate to improve the conditional expectation on the next iteration cycle (E-step) and, thus, to improve the next parameter estimate (M-step). If $Q(\theta, \theta')$ is continuous in both θ and θ' , the algorithm converges monotonically to a stationary point of $\log f_{\mathbf{Z}}(\mathbf{z}; \theta)$, that is, the observed log-likelihood function (see [22]), where each iteration increases the likelihood value. Of course, as in all "hill climbing" algorithms, the stationary point may not be the global maximum, and thus, several starting points or an initial grid search may be needed.

In order to apply the EM algorithm, we must specify the "complete" data \mathbf{y} . Following the considerations in [4], let \mathbf{s} denote the $(N + r)$ -dimensional vector of signal samples

$$\mathbf{s} = \{s(t) : -r + 1 \leq t \leq N\} \quad (18)$$

and let \mathbf{w} denote the $(N + q)$ -dimensional vector of noise samples

$$\mathbf{w} = \{w(t) : -q + 1 \leq t \leq N\}. \quad (19)$$

We assume that the orders r and q of the transfer functions are greater than the order p of the desired signal. This would typically be the case if, for example, the signal is speech and the transfer functions represent room acoustics. Under this assumption, the vectors \mathbf{s} and \mathbf{w} contain all the signal and the noise samples that affect the observed data \mathbf{z} .

Let the "complete" data \mathbf{y} be specified by

$$\mathbf{y} = \begin{bmatrix} \mathbf{z} \\ \mathbf{s} \\ \mathbf{w} \end{bmatrix}. \quad (20)$$

Invoking Bayes' rule

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}; \theta) &= f_{\mathbf{S}}(\mathbf{s}; \theta) \cdot f_{\mathbf{Z}|\mathbf{S}, \mathbf{W}}(\mathbf{z} | \mathbf{s}, \mathbf{w}; \theta) \\ &= f_{\mathbf{S}}(\mathbf{s}; \theta) \cdot f_{\mathbf{W}}(\mathbf{w}; \theta) \cdot f_{\mathbf{Z}|\mathbf{S}, \mathbf{W}}(\mathbf{z} | \mathbf{s}, \mathbf{w}; \theta) \end{aligned} \quad (21)$$

where $f_{\mathbf{S}}(\mathbf{s}; \theta)$ is the p.d.f. of \mathbf{s} , $f_{\mathbf{W}}(\mathbf{w}; \theta)$ is the p.d.f. of \mathbf{w} , and $f_{\mathbf{Z}|\mathbf{S}, \mathbf{W}}(\mathbf{z} | \mathbf{s}, \mathbf{w}; \theta)$ is the conditional p.d.f. of \mathbf{z} given \mathbf{s} and \mathbf{w} . In the transition from the first line of (21) to its second line, we invoked the statistical independence between \mathbf{s} and \mathbf{w} .

Taking the logarithm on both sides of (21)

$$\begin{aligned} \log f_{\mathbf{Y}}(\mathbf{y}; \theta) &= \log f_{\mathbf{S}}(\mathbf{s}; \theta) \\ &+ \log f_{\mathbf{W}}(\mathbf{w}; \theta) + \log f_{\mathbf{Z}|\mathbf{S}, \mathbf{W}}(\mathbf{z} | \mathbf{s}, \mathbf{w}; \theta). \end{aligned} \quad (22)$$

By (12)

$$\begin{aligned} \log f_{\mathbf{S}}(\mathbf{s}; \theta) &= \log f(s_{p-1}(0)) \\ &- \frac{N}{2} \log 2\pi g_s \\ &- \frac{1}{2g_s} \sum_{t=1}^N [s(t) + \alpha^T s_{p-1}(t-1)]^2 \end{aligned} \quad (23)$$

by (4)

$$\begin{aligned} \log f_{\mathbf{W}}(\mathbf{w}; \theta) &= \log f(w_q(0)) \\ &- \frac{N}{2} \log 2\pi g_w - \frac{1}{2g_w} \sum_{t=1}^N w^2(t) \end{aligned} \quad (24)$$

and by (10) and (11)

$$\begin{aligned} \log f_{\mathbf{Z}|\mathbf{S}, \mathbf{W}}(\mathbf{z} | \mathbf{s}, \mathbf{w}; \theta) &= -\frac{N}{2} \log 2\pi g_1 - \frac{1}{2g_1} \sum_{t=1}^N [z_1(t) - s(t) - \mathbf{a}^T \mathbf{w}_q(t)]^2 \\ &- \frac{N}{2} \log 2\pi g_2 - \frac{1}{2g_2} \sum_{t=1}^N [z_2(t) - w(t) - \mathbf{b}^T \mathbf{s}_r(t)]^2. \end{aligned} \quad (25)$$

Substituting (23)–(25) into (22) and assuming that $N \gg p, q$ so that the contributions of $\log f(s_{p-1}(0))$ and $\log f(w_q(0))$ are negligible

$$\begin{aligned} \log f_{\mathbf{Y}}(\mathbf{y}; \theta) &= C - \frac{N}{2} \log g_s - \frac{1}{2g_s} \sum_{t=1}^N [s(t) + \alpha^T s_{p-1}(t-1)]^2 \\ &- \frac{N}{2} \log g_w - \frac{1}{2g_w} \sum_{t=1}^N w^2(t) \\ &- \frac{N}{2} \log g_1 - \frac{1}{2g_1} \sum_{t=1}^N [z_1(t) - s(t) - \mathbf{a}^T \mathbf{w}_q(t)]^2 \\ &- \frac{N}{2} \log g_2 - \frac{1}{2g_2} \sum_{t=1}^N [z_2(t) - w(t) - \mathbf{b}^T \mathbf{s}_r(t)]^2 \end{aligned} \quad (26)$$

where C is a constant independent of θ . Taking the conditional expectation given z at a parameter value $\theta^{(l)}$, we obtain

$$\begin{aligned} Q(\theta, \theta^{(l)}) &= E_{\theta^{(l)}} \{ \log f_Y(y; \theta) | z \} \\ &= C + \sum_{i=1}^4 Q_i(\theta, \theta^{(l)}) \end{aligned} \quad (27)$$

where

$$\begin{aligned} Q_1(\theta, \theta^{(l)}) &= -\frac{N}{2} \log g_s - \frac{1}{2g_s} \sum_{t=1}^N \\ &\quad \cdot [\widehat{s^2(t)}^{(l)} + 2\alpha^\top \widehat{s_{p-1}(t-1)s(t)}^{(l)} \\ &\quad + \alpha^\top \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(l)} \alpha] \end{aligned} \quad (28)$$

$$\begin{aligned} Q_2(\theta, \theta^{(l)}) &= -\frac{N}{2} \log g_w \\ &\quad - \frac{1}{2g_w} \sum_{t=1}^N \widehat{w^2(t)}^{(l)} \end{aligned} \quad (29)$$

$$\begin{aligned} Q_3(\theta, \theta^{(l)}) &= -\frac{N}{2} \log g_1 \\ &\quad - \frac{1}{2g_1} \sum_{t=1}^N [z_1^2(t) - 2\widehat{s^{(l)}(t)z_1(t)} + \widehat{s^2(t)}^{(l)} \\ &\quad - 2\alpha^\top \widehat{w_q^{(l)}(t)z_1(t)} + 2\alpha^\top \widehat{w_q(t)s(t)}^{(l)} \\ &\quad + \alpha^\top \widehat{w_q(t)w_q^\top(t)}^{(l)} \alpha] \end{aligned} \quad (30)$$

$$\begin{aligned} Q_4(\theta, \theta^{(l)}) &= -\frac{N}{2} \log g_2 - \frac{1}{2g_2} \sum_{t=1}^N \\ &\quad \cdot [z_2^2(t) - 2\widehat{w(t)z_2(t)} + \widehat{w^2(t)}^{(l)} \\ &\quad - 2\mathbf{b}^\top \widehat{s_r^{(l)}(t)z_2(t)} + 2\mathbf{b}^\top \widehat{s_r(t)w(t)}^{(l)} \\ &\quad + \mathbf{b}^\top \widehat{s_r(t)s_r^\top(t)}^{(l)} \mathbf{b}] \end{aligned} \quad (31)$$

where we define $\widehat{(\cdot)}^{(l)} \triangleq E_{\theta^{(l)}} \{ \cdot | z \}$.

Thus, the computation of $Q(\theta, \theta^{(l)})$ (E-step) only requires the computation of $\widehat{x^{(l)}(t)}$ and $\widehat{x(t)x^\top(t)}^{(l)}$, where $x(t)$ is defined by

$$\begin{aligned} x(t) &= [s_r^\top(t); w_q^\top(t)]^\top \\ &= [s(t-r)s(t-r+1) \cdots s(t); \\ &\quad w(t-q)w(t-q+1) \cdots w(t)]^\top. \end{aligned} \quad (32)$$

Since $Q_1(\theta, \theta^{(l)})$ depends only on α and g_s , $Q_2(\theta, \theta^{(l)})$ depends only on g_w , $Q_3(\theta, \theta^{(l)})$ depends only on \mathbf{a} and g_1 , and $Q_4(\theta, \theta^{(l)})$ depends only on \mathbf{b} and g_2 , the maximization of $Q(\theta, \theta^{(l)})$ (M-step) decouples into the maximizations of each one of the terms in (27) with respect to the corresponding parameters. Furthermore, these maximizations can be solved analytically. The resulting algorithm is:

E-step: For $t = 1, 2, \dots, N$ compute

$$\widehat{x^{(l)}(t)} = E_{\theta^{(l)}} \{ x(t) | z \} \quad (33)$$

$$\widehat{x(t)x^\top(t)}^{(l)} = E_{\theta^{(l)}} \{ x(t)x^\top(t) | z \}. \quad (34)$$

M-step: Compute

$$\begin{aligned} \hat{\alpha}^{(l+1)} &= - \left[\sum_{t=1}^N \widehat{s_{p-1}(t-1)s_{p-1}^\top(t-1)}^{(l)} \right]^{-1} \\ &\quad \cdot \sum_{t=1}^N \widehat{s_{p-1}(t-1)s(t)}^{(l)} \end{aligned} \quad (35)$$

$$\begin{aligned} \hat{g}_s^{(l+1)} &= \frac{1}{N} \left\{ \sum_{t=1}^N \widehat{s^2(t)}^{(l)} + \hat{\alpha}^{(l+1)\top} \right. \\ &\quad \cdot \sum_{t=1}^N \widehat{s_{p-1}(t-1)s(t)}^{(l)} \left. \right\} \end{aligned} \quad (36)$$

$$\hat{g}_w^{(l+1)} = \frac{1}{N} \sum_{t=1}^N \widehat{w^2(t)}^{(l)} \quad (37)$$

$$\begin{aligned} \hat{\mathbf{a}}^{(l+1)} &= \left[\sum_{t=1}^N \widehat{w_q(t)w_q^\top(t)}^{(l)} \right]^{-1} \\ &\quad \cdot \sum_{t=1}^N [\widehat{w_q^{(l)}(t)z_1(t)} - \widehat{w_q(t)s(t)}^{(l)}] \end{aligned} \quad (38)$$

$$\begin{aligned} \hat{g}_1^{(l+1)} &= \frac{1}{N} \sum_{t=1}^N [z_1^2(t) - 2\widehat{s^{(l)}(t)z_1(t)} + \widehat{s^2(t)}^{(l)}] \\ &\quad - \hat{\mathbf{a}}^{(l+1)\top} \cdot \frac{1}{N} \sum_{t=1}^N [\widehat{w_q^{(l)}(t)z_1(t)} - \widehat{w_q(t)s(t)}^{(l)}] \end{aligned} \quad (39)$$

$$\begin{aligned} \hat{\mathbf{b}}^{(l+1)} &= \left[\sum_{t=1}^N \widehat{s_r(t)s_r^\top(t)}^{(l)} \right]^{-1} \\ &\quad \cdot \sum_{t=1}^N [\widehat{s_r^{(l)}(t)z_2(t)} - \widehat{s_r(t)w(t)}^{(l)}] \end{aligned} \quad (40)$$

$$\begin{aligned} \hat{g}_2^{(l+1)} &= \frac{1}{N} \left\{ \sum_{t=1}^N [z_2^2(t) - 2\widehat{w(t)z_2(t)} + \widehat{w^2(t)}^{(l)}] \right. \\ &\quad \left. - \hat{\mathbf{b}}^{(l+1)\top} \sum_{t=1}^N [\widehat{s_r^{(l)}(t)z_2(t)} - \widehat{s_r(t)w(t)}^{(l)}] \right\}. \end{aligned} \quad (41)$$

This algorithm has a nice intuitive form. In the E-step, we use the current parameter estimate $\theta^{(l)}$ to estimate the sufficient statistics of the desired signal and the noise. The M-step of the algorithm decouples as follows: Equation (35) is the Yule-Walker solution for the AR parameters where the sufficient statistics of the signal are replaced by their current estimates. Equations (38) and (40) are, respectively, the least squares solutions for \mathbf{a} and \mathbf{b} based on the estimated sufficient statistics. The gain parameters in (36), (37), (39), and (41) are the sample averages of the corresponding power levels. We note that if some of the gain parameters are known *a priori* and need not be estimated, we simply eliminate the corresponding equations.

Since the algorithm is based on the EM method, it converges monotonically to the ML estimate of θ or, at least, to a stationary point of the log-likelihood function. As a by product, it also provides the desired signal estimate $\widehat{s^{(l)}(t)}$, which is

the $(r+1)$ st component of $\hat{\mathbf{x}}^{(l)}(t)$. For the purpose of signal enhancement, it is the signal estimate that we are primarily interested in.

The computation of the conditional expectations in (33) and (34) can be carried out using the Kalman smoothing equations. To do that, let us represent (10)–(12) in a state-space form:

$$\mathbf{x}(t) = \Phi \mathbf{x}(t-1) + G \mathbf{u}(t) \quad (42)$$

$$\mathbf{z}(t) = H \mathbf{x}(t) + \mathbf{e}(t) \quad (43)$$

where $\mathbf{x}(t)$, which is the state vector, is defined in (32), and the other terms in (42) and (43) are defined by

$$\mathbf{z}(t) = [z_1(t) z_2(t)]^T, \quad (44)$$

$$\mathbf{u}(t) = [u_s(t) u_w(t)]^T, \quad (45)$$

$$\mathbf{e}(t) = [e_1(t) e_2(t)]^T, \quad (46)$$

$$G = \begin{bmatrix} 0 & \cdots & 0 & \sqrt{g_s} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \sqrt{g_w} \end{bmatrix}^T, \quad (47)$$

$$H = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 & a_q & a_{q-1} & \cdots & a_0 \\ b_r & b_{r-1} & \cdots & \cdots & b_0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \quad (48)$$

$$\Phi = \begin{bmatrix} \Phi_s & 0 \\ 0 & \Phi_w \end{bmatrix} \begin{matrix} \uparrow r+1 \\ \uparrow q+1 \end{matrix}$$

$\begin{matrix} \leftarrow & \leftarrow \end{matrix}$
 $\begin{matrix} r+1 & q+1 \end{matrix}$

where

$$\Phi_s = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & \vdots \\ \vdots & & & & & \ddots & \vdots \\ \vdots & & & & & & 0 \\ \vdots & & & & & & 0 & 1 \end{bmatrix} \quad (50)$$

and

$$\Phi_w = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & \ddots & \vdots \\ \vdots & & & & & \ddots & \vdots \\ \vdots & & & & & & 0 \\ \vdots & & & & & & 0 & 1 \\ \vdots & & & & & & & 0 \end{bmatrix} \quad (51)$$

We also define by R the 2×2 covariance matrix of $\mathbf{e}(t)$:

$$R = \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix}. \quad (52)$$

Now, let the conditional expectations required in (33) and (34) (E-step) be represented as

$$\hat{\mathbf{x}}^{(l)}(t) = \boldsymbol{\mu}_{t|N}^{(l)} \quad (53)$$

$$\mathbf{x}(t) \mathbf{x}^T(t) = \boldsymbol{\mu}_{t|N}^{(l)} \boldsymbol{\mu}_{t|N}^{(l)\top} + P_{t|N}^{(l)} \quad (54)$$

where we define

$$\boldsymbol{\mu}_{t|n}^{(l)} = E_{\boldsymbol{\theta}^{(l)}} \{ \mathbf{x}(t) | z(1), z(2), \dots, z(n) \} \quad (55)$$

$$P_{t|n}^{(l)} = E_{\boldsymbol{\theta}^{(l)}} \{ [\mathbf{x}(t) - \boldsymbol{\mu}_{t|n}^{(l)}][\mathbf{x}(t) - \boldsymbol{\mu}_{t|n}^{(l)}]^\top | z(1), z(2), \dots, z(n) \}. \quad (56)$$

Denote by $\Phi^{(l)}$, $G^{(l)}$, $H^{(l)}$, and $R^{(l)}$ the matrices Φ , G , H , and R computed at the current parameter estimate $\boldsymbol{\theta} = \boldsymbol{\theta}^{(l)}$. Then, using the Kalman smoothing equations, $\boldsymbol{\mu}_{t|n}^{(l)}$ and $P_{t|n}^{(l)}$ may be computed, in three stages, as follows.

Propagation Equations: For $t = 1, 2, \dots, N$ compute

$$\boldsymbol{\mu}_{t|t-1}^{(l)} = \Phi^{(l)} \boldsymbol{\mu}_{t-1|t-1}^{(l)} \quad (57)$$

$$P_{t|t-1}^{(l)} = \Phi^{(l)} P_{t-1|t-1}^{(l)} \Phi^{(l)\top} + G^{(l)} G^{(l)\top} \quad (58)$$

with initial conditions $\boldsymbol{\mu}_{0|0}^{(l)}$ and $P_{0|0}^{(l)}$.

Updating Equations: For $t = 1, 2, \dots, N$ compute

$$\boldsymbol{\mu}_{t|t}^{(l)} = \boldsymbol{\mu}_{t|t-1}^{(l)} + K_t^{(l)} [z(t) - H^{(l)} \boldsymbol{\mu}_{t|t-1}^{(l)}] \quad (59)$$

$$P_{t|t}^{(l)} = [I - K_t^{(l)} H^{(l)}] P_{t|t-1}^{(l)} \quad (60)$$

where I is the identity matrix, and $K_t^{(l)}$ is the Kalman gain:

$$K_t^{(l)} = P_{t|t-1}^{(l)} H^{(l)\top} [H^{(l)} P_{t|t-1}^{(l)} H^{(l)\top} + R^{(l)}]^{-1}. \quad (61)$$

Smoothing Equations: For $t = N, N-1, \dots, 1$ compute

$$\boldsymbol{\mu}_{t-1|N}^{(l)} = \boldsymbol{\mu}_{t-1|t-1}^{(l)} + S_{t-1}^{(l)} [\boldsymbol{\mu}_{t|N}^{(l)} - \Phi^{(l)} \boldsymbol{\mu}_{t-1|t-1}^{(l)}] \quad (62)$$

$$P_{t-1|N}^{(l)} = P_{t-1|t-1}^{(l)} + S_{t-1}^{(l)} [P_{t|N}^{(l)} - P_{t|t-1}^{(l)}] S_{t-1}^{(l)\top} \quad (63)$$

where

$$S_{t-1}^{(l)} \triangleq P_{t-1|t-1}^{(l)} \Phi^{(l)\top} P_{t|t-1}^{(l)-1}. \quad (64)$$

To initialize the Kalman smoothing equations, we must specify $\boldsymbol{\mu}_{0|0}^{(l)}$ and $P_{0|0}^{(l)}$. In the case of weak coupling we may use the actual sensor data, i.e., the first $(r+1)$ samples of $z_1(t)$ and the first $(q+1)$ samples of $z_2(t)$ to specify $\boldsymbol{\mu}_{0|0}^{(0)}$, and the initial covariance $P_{0|0}^{(0)}$ can be assessed by computing sample covariances. These initial estimates can then be iteratively improved by using the final estimates from the previous iteration cycle, i.e., $\boldsymbol{\mu}_{0|0}^{(l)} = \boldsymbol{\mu}_{0|N}^{(l)}$ and $P_{0|0}^{(l+1)} = P_{0|N}^{(l)}$.

Comments: The EM algorithm for joint state and parameter estimation in linear dynamic state-space models has previously been developed in [12] and [17]. The proposed algorithm can therefore be viewed as an extension of these methods to the two-channel signal enhancement problem.

The EM algorithm is only guaranteed to converge to a local maximum of the likelihood function. Therefore, in order to ensure convergence to the global maximum, a good initialization procedure may be required. Under the assumption of weak coupling of the desired signal $s(t)$ to the second

(reference) sensor, we may initialize \mathbf{b} to be a zero vector and apply the LMS method [21] to obtain an initial estimate of \mathbf{a} . In case of strong coupling of the desired signal to the reference sensor, use of the method in [19] is suggested to extract initial estimates for both \mathbf{a} and \mathbf{b} . These are the most crucial parameters for the purpose of estimating, or enhancing, the desired signal.

The rate of convergence of the EM algorithm is determined by the portion of the covariance of the complete data that can be predicted using the incomplete (observed) data (see [3] and [11]). In our problem it is determined by the power levels g_1 and g_2 of the additive noises. Therefore, in order to obtain fast convergence, overestimating g_1 and g_2 is recommended at the initial stage of the algorithm.

All other parameters α , g_s , and g_w may be initialized by a least-squares fit of the signal and the noise to the actual sensor data in case of weak coupling, or they may be arbitrarily initialized. Simulation results indicate that the initialization of the signal and noise spectral parameters has little effect on the convergence behaviour of the algorithm.

The computational complexity of the proposed EM algorithm is determined by the orders $(q+1)$ and $(r+1)$ of the coupling systems that are the dimensions of the vectors \mathbf{a} and \mathbf{b} , respectively. For example, in the problem of enhancing speech, the coupling systems that represent the acoustic transfer functions in a room environment are typically modeled as FIR filters of at least several hundred coefficients. Thus, in the M-step of the algorithm, (38) and (40) are computationally the most demanding since they require the inversion of $(q+1) \times (q+1)$ and $(r+1) \times (r+1)$ matrices, respectively. In the E-step of the algorithm, we need to implement the Kalman smoothing equations for estimating a $(q+r+2)$ -dimensional state vector. The smoothing equations (62) and (63) are computationally the most expensive since they require the inversion of matrices of dimension $(q+r+2) \times (q+r+2)$. We may exploit the structure of the matrices Φ , G , H , and R in order to simplify the computations involved (see considerations in the sequel).

We also note that there are a variety of methods, e.g., the square-root algorithm, that can be used to implement the Kalman equations more efficiently (e.g., see [1]).

IV. SEQUENTIAL/ADAPTIVE ALGORITHMS

The algorithm in the previous section was developed under the assumption that the signal and the noise are stationary processes and that the coupling systems are time invariant. However, in many practical situations, the desired signal (e.g., speech) is more typically a nonstationary process with time-varying spectral parameters, and the parameters a_k and b_k characterizing the coupling filters may also be time varying, perhaps because of changes in the propagation conditions in the medium or because of changes in the source-receiver geometry during the observation interval. Basically, we have a situation in which the state vector $\mathbf{x}(t)$ that contains the signal and the noise samples depends on a time-varying parameter vector $\theta(t)$, and we want an adaptive algorithm that is capable of tracking the varying parameters.

One approach is to assume that the observed signals are stationary over a fixed time window, and apply the algorithm on consecutive data blocks. An alternative, and certainly more attractive, approach is suggested by the structure of the algorithm. As it stands, the algorithm iterates between state estimation using the Kalman smoother and parameter identification. To obtain a sequential/adaptive algorithm, we replace the Kalman smoother by the Kalman filter, requiring only the propagation equations followed by the updating equations. In that way, the state at a particular time instant t is estimated using only the past and current data samples, and there is no need for the smoothing equations that are computationally the most expensive. Then, we suggest incorporating exponential weighting into the parameter estimation update in order to reduce the effect of past data samples relative to new data. The exponential weighting is effectively equivalent to a sliding window operation. Finally, since we are interested in an on-line adaptive algorithm, we suggest restricting the number of iteration cycles. In particular, if we perform only one iteration per data sample (i.e., replacing the iteration index by the time index), then we obtain a fully sequential algorithm in which the state (signal) estimate is generated using a forward Kalman filter whose parameters are continuously updated.

Specifically, denote the estimates of $\mathbf{x}(t)$ and $\mathbf{x}(t)^T(t)$ based on the observed data to time t and the current parameter estimate $\hat{\theta}(t)$ by

$$\hat{\mathbf{x}}(t|t) = E_{\hat{\theta}(t)} \{ \mathbf{x}(t) | z(1), z(2), \dots, z(t) \} \triangleq \boldsymbol{\mu}_{t|t} \quad (65)$$

$$\mathbf{x}(t|t) \widehat{\mathbf{x}}^T(t|t) = E_{\hat{\theta}(t)} \{ \mathbf{x}(t) \mathbf{x}^T(t) | z(1), z(2), \dots, z(t) \} \triangleq \boldsymbol{\mu}_{t|t}^T + P_{t|t}. \quad (66)$$

In addition, denote by $\hat{\Phi}_t$, \hat{G}_t , \hat{H}_t , and \hat{R}_t the matrices Φ , G , H , and R computed at $\theta = \hat{\theta}(t)$. Then, $\boldsymbol{\mu}_{t|t}$ and $P_{t|t}$ are computed recursively in t as follows.

Propagation Equations:

$$\boldsymbol{\mu}_{t|t-1} = \hat{\Phi}_t \boldsymbol{\mu}_{t-1|t-1} \quad (67)$$

$$P_{t|t-1} = \hat{\Phi}_t P_{t-1|t-1} \hat{\Phi}_t^T + \hat{G}_t \hat{G}_t^T \quad (68)$$

with initial conditions $\boldsymbol{\mu}_{0|0}$ and $P_{0|0}$.

Updating Equations:

$$\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \hat{K}_t [z(t) - \hat{H}_t \boldsymbol{\mu}_{t|t-1}] \quad (69)$$

$$P_{t|t} = [I - \hat{K}_t \hat{H}_t] P_{t|t-1}, \quad (70)$$

where

$$\hat{K}_t = P_{t|t-1} \hat{H}_t^T [\hat{H}_t P_{t|t-1} \hat{H}_t^T + \hat{R}_t]^{-1}. \quad (71)$$

In Appendix A, we have exploited the structure of the matrices, Φ , G , H , and R in order to efficiently compute the Kalman filtering equations. Instead of having to multiply $(q+r+2) \times (q+r+2)$ matrices as suggested in (68), the computationally most expensive term in the efficient

implementation is a quadratic form $\mathbf{u}^T \mathbf{V} \mathbf{u}$, where \mathbf{u} is a vector of dimension q or r . For the specific application of speech enhancement in a noisy acoustic environment subject to reverberant or multipath effects, the dimensions r and q of the FIR filters characterizing the acoustic transfer functions tend to be very large (on the order of several hundred coefficients), in which case, the development in Appendix A results in a very significant savings in computations.

The parameter estimate $\hat{\theta}(t)$ is obtained from (35)–(41) by replacing the iteration index l by the time index t , using data only up to the current time t and incorporating exponential weighting. Thus

$$\hat{\alpha}(t+1) = - \left[\sum_{\tau=1}^t \gamma_s^{t-\tau} \overbrace{s_{p-1}(\tau-1|\tau) s_{p-1}^T(\tau-1|\tau)} \right]^{-1} \cdot \sum_{\tau=1}^t \gamma_s^{t-\tau} \overbrace{s_{p-1}(\tau-1|\tau) s(\tau|\tau)} \quad (72)$$

$$\hat{g}_s(t+1) = \frac{1}{\sum_{\tau=1}^t \gamma_s^{t-\tau}} \left[\sum_{\tau=1}^t \gamma_s^{t-\tau} \hat{s}^2(\tau|\tau) + \hat{\alpha}^T(t+1) \cdot \sum_{\tau=1}^t \gamma_s^{t-\tau} \overbrace{s_{p-1}(\tau-1|\tau) s(\tau|\tau)} \right] \quad (73)$$

$$\hat{g}_w(t+1) = \frac{1}{\sum_{\tau=1}^t \gamma_w^{t-\tau}} \cdot \sum_{\tau=1}^t \gamma_w^{t-\tau} \hat{w}^2(\tau|\tau) \quad (74)$$

$$\hat{\mathbf{a}}(t+1) = \left[\sum_{\tau=1}^t \gamma_a^{t-\tau} \overbrace{\mathbf{w}_q(\tau|\tau) \mathbf{w}_q^T(\tau|\tau)} \right]^{-1} \cdot \sum_{\tau=1}^t \gamma_a^{t-\tau} [\hat{\mathbf{w}}_q(\tau|\tau) z_1(\tau) - \overbrace{\mathbf{w}_q(\tau|\tau) s(\tau|\tau)}] \quad (75)$$

$$\hat{g}_1(t+1) = \frac{1}{\sum_{\tau=1}^t \gamma_a^{t-\tau}} \cdot \left\{ \sum_{\tau=1}^t \gamma_1^{t-\tau} [z_1^2(\tau) - 2\hat{s}(\tau|\tau) z_1(\tau) + \hat{s}^2(\tau|\tau)] - \hat{\mathbf{a}}^T(t+1) \sum_{\tau=1}^t \gamma_1^{t-\tau} [\hat{\mathbf{w}}_q(\tau|\tau) z_1(\tau) - \overbrace{\mathbf{w}_q(\tau|\tau) s(\tau|\tau)}] \right\} \quad (76)$$

$$\hat{\mathbf{b}}(t+1) = \left[\sum_{\tau=1}^t \gamma_b^{t-\tau} \overbrace{s_r(\tau|\tau) s_r^T(\tau|\tau)} \right]^{-1} \cdot \sum_{\tau=1}^t \gamma_b^{t-\tau} [\hat{s}_r(\tau|\tau) z_2(\tau) - \overbrace{s_r(\tau|\tau) w(\tau|\tau)}] \quad (77)$$

$$\hat{g}_2(t+1) = \frac{1}{\sum_{\tau=1}^t \gamma_2^{t-\tau}} \left\{ \sum_{\tau=1}^t \gamma_2^{t-\tau} [z_2^2(\tau) \right.$$

$$\left. - 2\hat{w}(\tau|\tau) z_2(\tau) + \hat{w}^2(\tau|\tau) \right\} - \hat{\mathbf{b}}^T(t+1) \sum_{\tau=1}^t \gamma_b^{t-\tau} [\hat{s}_r(\tau|\tau) z_2(\tau) - \overbrace{s_r(\tau|\tau) w(\tau|\tau)}] \quad (78)$$

where $\gamma_s, \gamma_w, \gamma_a, \gamma_b, \gamma_1$, and γ_2 are the exponential weights that are preselected real-valued numbers between 0 and 1, specifying the effective window in estimating the various parameters. Recursive formulae for computing these equations are developed in Appendix B. Altogether, we obtain a fully sequential algorithm that is capable of tracking nonstationary changes in the structure of the data.

Of course, this is no longer an EM algorithm, and there is no proof that it converges. However, since the algorithm essentially consists of a Kalman filter whose parameters are continuously updated, then under certain stationary/ergodic conditions it may be possible to claim convergence. This important issue must be explored in depth.

We note that this is not the only way to generate sequential/adaptive algorithms. One may use the state-space model in (42) and (43) as a starting point and apply the prediction error methods in [2] and [9] directly to obtain sequential and adaptive algorithms. It might also be interesting to apply these methods to the two-channel signal enhancement problem and to compare the resulting algorithms to the algorithm presented here.

V. GRADIENT-BASED ALGORITHMS

As an alternative to the EM algorithm, consider the class of gradient-based algorithms for finding the ML parameter estimate:

$$\theta^{(l+1)} = \theta^{(l)} + \Lambda(z; \theta^{(l)}) \frac{\partial \log f_Z(z; \theta)}{\partial \theta} \bigg|_{\theta=\theta^{(l)}} \quad (79)$$

where, as before, $\theta^{(l)}$ denotes the estimate of θ after l iteration cycles. If $\Lambda(z; \theta)$ is a constant diagonal matrix independent of z and θ , we obtain the conventional steepest-descent algorithm, where the diagonal elements of Λ are the step-sizes used for updating the estimates of the various parameters along the iterations. For sufficiently small step-sizes, the steepest-descent algorithm converges linearly to the desired ML solution or, at least, to a local maximum of the likelihood function. As shown in [18], the EM algorithm also possesses a linear convergence rate near the point of convergence. If $\Lambda(z; \theta) = -H^{-1}(z; \theta)$, where $H(z; \theta)$ is the Hessian matrix defined by

$$H(z; \theta) = \frac{\partial^2 \log f_Z(z; \theta)}{\partial \theta^2}$$

we obtain the well-known Newton-Raphson algorithm. If $\Lambda(z; \theta) = J^{-1}(\theta)$, where $J(\theta)$ is the Fisher information matrix (FIM) defined by

$$J(\theta) = E \left\{ - \frac{\partial^2 \log f_Z(z; \theta)}{\partial \theta^2} \right\}$$

$$= E \left\{ \frac{\partial \log f_Z(z; \theta)}{\partial \theta} \frac{\partial \log f_Z(z; \theta)}{\partial \theta} \right\} \quad (80)$$

we obtain the scoring algorithm (e.g., see [14]). The Newton-Raphson and the scoring algorithms possess a super-linear (quadratic) convergence rate; however, their convergence may depend more critically on initialization.

The computation of the log-likelihood gradient (score) $\partial \log f_Z(z; \theta) / \partial \theta$ by direct differentiation of the log-likelihood function $\log f_Z(z; \theta)$ is very complicated. It leads to the so-called sensitivity derivatives, which are effectively the derivatives of the Kalman filtering equations (i.e., the propagation and updating equations) with respect to the unknown parameters (e.g., see [2], [6]). The prediction-error methods in [2], [9] also require the computation of the sensitivity derivatives.

An alternative approach for computing the score in linear dynamic state models is proposed in [15]. It is based on Fisher's identity [5]

$$\frac{\partial}{\partial \theta} \log f_Z(z; \theta) = E_{\theta} \left\{ \frac{\partial \log f_Y(y; \theta)}{\partial \theta} \middle| z \right\} \quad (81)$$

where, as before, z is the observed (incomplete) data and y is the complete data. Fisher's identity asserts that the observed (incomplete) data score is equal to the conditional expectation of the complete data

Substituting (26) into (81) and carrying out the indicated differentiation and expectation operations, the components of the score vector are given by

$$\frac{\partial}{\partial \alpha} \log f_Z(z; \theta) = -\frac{1}{g_s} \sum_{t=1}^N \left[\widehat{s(t) s_{p-1}^T(t-1)} + \alpha^T s_{p-1}(t-1) \widehat{s_{p-1}^T(t-1)} \right] \quad (82)$$

$$\frac{\partial}{\partial g_s} \log f_Z(z; \theta) = -\frac{N}{2g_s} + \frac{1}{2g_s^2} \sum_{t=1}^N \left[\widehat{s^2(t)} + 2\alpha^T s_{p-1}(t-1) \widehat{s(t)} + \alpha^T s_{p-1}(t-1) \widehat{s_{p-1}^T(t-1)} \alpha \right] \quad (83)$$

$$\frac{\partial}{\partial g_w} \log f_Z(z; \theta) = -\frac{N}{2g_w} + \frac{1}{2g_w^2} \sum_{t=1}^N \widehat{w^2(t)} \quad (84)$$

$$\frac{\partial}{\partial \mathbf{a}} \log f_Z(z; \theta) = \frac{1}{g_1} \sum_{t=1}^N \left[\widehat{z_1(t) \mathbf{w}_q^T(t)} - \widehat{s(t) \mathbf{w}_q^T(t)} - \alpha^T \widehat{\mathbf{w}_q(t) \mathbf{w}_q^T(t)} \right] \quad (85)$$

$$\frac{\partial}{\partial g_1} \log f_Z(z; \theta) = -\frac{N}{2g_1} + \frac{1}{2g_1^2} \sum_{t=1}^N \left[\widehat{z_1^2(t)} - 2z_1(t) \widehat{s(t)} + \widehat{s^2(t)} - 2\alpha^T \widehat{\mathbf{w}_q(t) z_1(t)} + 2\alpha^T \widehat{\mathbf{w}_q(t) s(t)} + \alpha^T \widehat{\mathbf{w}_q(t) \mathbf{w}_q^T(t)} \alpha \right] \quad (86)$$

$$\frac{\partial}{\partial \mathbf{b}} \log f_Z(z; \theta) = \frac{1}{g_2} \sum_{t=1}^N \left[\widehat{z_2(t) \mathbf{s}_r^T(t)} - \widehat{w(t) \mathbf{s}_r^T(t)} - \mathbf{b}^T \widehat{\mathbf{s}_r(t) \mathbf{s}_r^T(t)} \right] \quad (87)$$

$$\begin{aligned} \frac{\partial}{\partial g_2} \log f_Z(z; \theta) = & -\frac{N}{2g_2} + \frac{1}{2g_2^2} \sum_{t=1}^N \left[\widehat{z_2^2(t)} \right. \\ & - 2z_2(t) \widehat{w(t)} + \widehat{w^2(t)} - 2\mathbf{b}^T \widehat{\mathbf{s}_r(t) z_2(t)} \\ & \left. + 2\mathbf{b}^T \widehat{\mathbf{s}_r(t) w(t)} + \mathbf{b}^T \widehat{\mathbf{s}_r(t) \mathbf{s}_r^T(t)} \mathbf{b} \right] \end{aligned} \quad (88)$$

where $\widehat{(\cdot)} \triangleq E_{\theta} \{ \cdot | z \}$.

We observe that (82)–(88) depend on the data z only through the conditional expectations $\widehat{\mathbf{x}(t)} = E_{\theta} \{ \mathbf{x}(t) | z \}$ and $\widehat{\mathbf{x}(t) \mathbf{x}^T(t)} = E_{\theta} \{ \mathbf{x}(t) \mathbf{x}^T(t) | z \}$.

Thus, if we choose to apply the steepest-descent algorithm, then substituting (82)–(88) into (79), we obtain the following algorithm:

Signal Estimation: For $t = 1, 2, \dots, N$ compute

$$\widehat{\mathbf{x}}^{(l)}(t) = E_{\theta^{(l)}} \{ \mathbf{x}(t) | z \} \quad (89)$$

$$\widehat{\mathbf{x}(t) \mathbf{x}^T(t)}^{(l)} = E_{\theta^{(l)}} \{ \mathbf{x}(t) \mathbf{x}^T(t) | z \}. \quad (90)$$

Parameter Estimation:

$$\begin{aligned} \widehat{\alpha}^{(l+1)} = & \widehat{\alpha}^{(l)} - \frac{\delta_{\alpha}}{\widehat{g}_s^{(l)}} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{s_{p-1}(t-1) s(t)} \right. \\ & \left. + \widehat{s_{p-1}(t-1) s_{p-1}^T(t-1)} \widehat{\alpha}^{(l)} \right] \end{aligned} \quad (91)$$

$$\begin{aligned} \widehat{g}_s^{(l+1)} = & \left(1 - \frac{\delta_s}{2} \right) \widehat{g}_s^{(l)} + \frac{\delta_s}{2} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{s^2(t)} \right. \\ & \left. + 2\widehat{\alpha}^{(l)T} \widehat{s_{p-1}(t-1) s(t)} \right. \\ & \left. + \widehat{\alpha}^{(l)T} \widehat{s_{p-1}(t-1) s_{p-1}^T(t-1)} \widehat{\alpha}^{(l)} \right] \end{aligned} \quad (92)$$

$$\widehat{g}_w^{(l+1)} = \left(1 - \frac{\delta_w}{2} \right) \widehat{g}_w^{(l)} + \frac{\delta_w}{2} \cdot \frac{1}{N} \sum_{t=1}^N \widehat{w^2(t)} \quad (93)$$

$$\begin{aligned} \widehat{\mathbf{a}}^{(l+1)} = & \widehat{\mathbf{a}}^{(l)} + \frac{\delta_a}{\widehat{g}_1^{(l)}} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{w_q^{(l)}(t) z_1(t)} \right. \\ & \left. - \widehat{w_q(t) s(t)}^{(l)} - \widehat{w_q(t) \mathbf{w}_q^T(t)} \widehat{\mathbf{a}}^{(l)} \right] \end{aligned} \quad (94)$$

$$\begin{aligned} \widehat{g}_1^{(l+1)} = & \left(1 - \frac{\delta_1}{2} \right) \widehat{g}_1^{(l)} + \frac{\delta_1}{2} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{z_1^2(t)} \right. \\ & - 2z_1(t) \widehat{s(t)} + \widehat{s^2(t)} \\ & - 2\widehat{\mathbf{a}}^{(l)T} \widehat{w_q^{(l)}(t) z_1(t)} + 2\widehat{\mathbf{a}}^{(l)T} \widehat{w_q(t) s(t)}^{(l)} \\ & \left. + \widehat{\mathbf{a}}^{(l)T} \widehat{w_q(t) \mathbf{w}_q^T(t)} \widehat{\mathbf{a}}^{(l)} \right] \end{aligned} \quad (95)$$

$$\begin{aligned} \widehat{\mathbf{b}}^{(l+1)} = & \widehat{\mathbf{b}}^{(l)} + \frac{\delta_b}{\widehat{g}_2^{(l)}} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{s_r^{(l)}(t) z_2(t)} \right. \\ & \left. - \widehat{s_r(t) w(t)}^{(l)} - \widehat{s_r(t) \mathbf{s}_r^T(t)} \widehat{\mathbf{b}}^{(l)} \right] \end{aligned} \quad (96)$$

$$\begin{aligned} \widehat{g}_2^{(l+1)} = & \left(1 - \frac{\delta_2}{2} \right) \widehat{g}_2^{(l)} + \frac{\delta_2}{2} \cdot \frac{1}{N} \sum_{t=1}^N \left[\widehat{z_2^2(t)} - 2z_2(t) \widehat{w(t)} + \widehat{w^2(t)} \right. \\ & - 2\widehat{\mathbf{b}}^{(l)T} \widehat{s_r^{(l)}(t) z_2(t)} + 2\widehat{\mathbf{b}}^{(l)T} \\ & \left. \cdot \widehat{s_r(t) w(t)}^{(l)} + \widehat{\mathbf{b}}^{(l)T} \widehat{s_r(t) \mathbf{s}_r^T(t)} \widehat{\mathbf{b}}^{(l)} \right]. \end{aligned} \quad (97)$$

where $\delta_\alpha, \delta_s, \delta_w, \delta_a, \delta_b, \delta_1$, and δ_2 are the step-sizes used in the algorithm.

The most striking feature of this algorithm is that it has the same form as the EM algorithm specified by (33)–(41). It consists of a signal estimation step followed by a parameter estimation step. Furthermore, the signal estimation step is identical in both algorithms, and can be implemented using the Kalman smoothing equations. The only difference is in the parameter estimation update. Since (91)–(97) do not require any matrix inversions, they may be simpler to compute, particularly for the problem of speech enhancement in noisy acoustic environments, when the dimensions of the coupling filters representing the acoustic transfer functions tend to be very large, and the matrix inversions required in (38) and (40) of the M-step of the EM algorithm are computationally very expensive.

It is important to note that unlike the EM algorithm whose convergence rate depends on the complete data specification, the steepest descent is a gradient-based algorithm whose convergence path depends solely on the step-sizes being used. Here, we have used the notion of complete data only as a mechanism to compute the log-likelihood gradient.

We may choose to first apply the EM algorithm in order to guarantee monotonic convergence and then switch to the gradient algorithm for parameter updating, with the benefit of reduced computational requirement. We may also incorporate the Newton–Raphson or the scoring algorithm in order to accelerate the algorithm near the point of convergence. For that purpose, we need to compute the Hessian or the FIM (see considerations in [16]).

As with the EM algorithm, to convert the gradient algorithm into a sequential/adaptive algorithm, we suggest replacing the Kalman smoother used for state (signal) estimation by the Kalman filtering equations (67)–(71). Then, to obtain the parameter estimate $\hat{\theta}(t)$, we suggest replacing the cumulative averages $1/N \Sigma(\cdot)$ appearing in (91)–(97) by the most current term in the sum, and replacing the iteration index l by the time index t . With these modifications, we obtain the following recursive formulae for updating the components of $\hat{\theta}(t)$:

$$\hat{\alpha}(t+1) = \hat{\alpha}(t) - \frac{\delta_\alpha}{\hat{g}_s(t)} \cdot \left[\hat{s}_{p-1}(t-1|t)s(t|t) + \hat{s}_{p-1}(t-1|t)\hat{s}_{p-1}^\top(t-1|t)\hat{\alpha}(t) \right] \quad (98)$$

$$\hat{g}_s(t+1) = \left(1 - \frac{\delta_s}{2}\right)\hat{g}_s(t) + \frac{\delta_s}{2} \cdot \left[\hat{s}^2(t|t) + 2\hat{\alpha}^\top(t)\hat{s}_{p-1}(t-1|t)s(t|t) + \hat{\alpha}^\top(t)\hat{s}_{p-1}(t-1|t)\hat{s}_{p-1}^\top(t-1|t)\hat{\alpha}(t) \right] \quad (99)$$

$$\hat{g}_w(t+1) = \left(1 - \frac{\delta_w}{2}\right)\hat{g}_w(t) + \frac{\delta_w}{2}\hat{w}^2(t|t) \quad (100)$$

$$\hat{\mathbf{a}}(t+1) = \hat{\mathbf{a}}(t) + \frac{\delta_a}{\hat{g}_1(t)} \cdot \left[\hat{\mathbf{w}}_q(t|t)z_1(t) - \hat{\mathbf{w}}_q(t|t)s(t|t) - \hat{\mathbf{w}}_q(t|t)\hat{\mathbf{w}}_q^\top(t|t)\hat{\mathbf{a}}(t) \right] \quad (101)$$

$$\hat{g}_1(t+1) = \left(1 - \frac{\delta_1}{2}\right)\hat{g}_1(t) + \frac{\delta_1}{2}$$

$$\cdot \left[z_1^2(t) - 2z_1(t)\hat{s}(t) + \hat{s}^2(t) - 2\hat{\mathbf{a}}^\top(t)\hat{\mathbf{w}}_q(t|t)z_1(t) + 2\hat{\mathbf{a}}^\top(t)\hat{\mathbf{w}}_q(t|t)s(t|t) + \hat{\mathbf{a}}^\top(t)\hat{\mathbf{w}}_q(t|t)\hat{\mathbf{w}}_q^\top(t|t)\hat{\mathbf{a}}(t) \right] \quad (102)$$

$$\hat{\mathbf{b}}(t+1) = \hat{\mathbf{b}}(t) + \frac{\delta_b}{\hat{g}_2(t)} \left[\hat{s}_r(t|t)z_2(t) - \hat{s}_r(t|t)\hat{w}(t|t) - \hat{s}_r(t|t)\hat{s}_r^\top(t|t)\hat{\mathbf{b}}(t) \right] \quad (103)$$

$$\hat{g}_2(t+1) = \left(1 - \frac{\delta_2}{2}\right)\hat{g}_2(t) + \frac{\delta_2}{2} \left[z_2^2(t) - 2z_2(t)\hat{w}(t|t) + \hat{w}^2(t|t) - 2\hat{\mathbf{b}}^\top(t)\hat{s}_r(t|t)z_2(t) + 2\hat{\mathbf{b}}^\top(t)\hat{s}_r(t|t)\hat{w}(t|t) + \hat{\mathbf{b}}^\top(t)\hat{s}_r(t|t)\hat{s}_r^\top(t|t)\hat{\mathbf{b}}(t) \right] \quad (104)$$

Once again, the signal estimation is identical for both the sequential-EM and the sequential gradient algorithms. The difference is only in the parameter estimation update. If we are primarily interested in an on-line adaptive algorithm then from a computational viewpoint, the sequential-gradient algorithm may be preferable since it does not require any matrix inversions. We note that the efficient implementation of the Kalman filtering equations presented in Appendix A can be exploited to further simplify the form of (98)–(104).

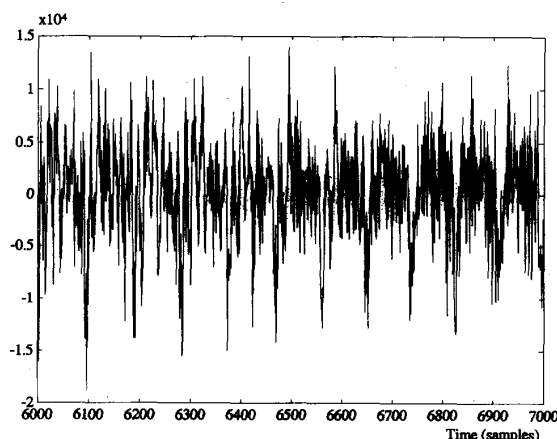
VI. EXPERIMENTAL RESULTS

To verify the proposed approach, we have implemented and tested the algorithms in the following context: The desired signal $s(t)$ is speech, corresponding to the sentence: “He has the bluest eyes.” The signals $w(t)$, $e_1(t)$, and $e_2(t)$ are computer generated white Gaussian noises. The coupling systems A and B are derived from the simulated room acoustics impulse responses used in [4]. For computational simplicity the impulse responses are assumed to be finite length with order 128, i.e., $r = q = 127$. The spectral level of $w(t)$ is set such that the SNR, that is, the average power of the desired signal divided by the average power of the residual signal, at the first (primary) sensor is 0 dB, and the SNR at the second (reference) sensor is -26 dB. The levels of the independent noise sources $e_1(t)$ and $e_2(t)$ are 34 dB below the level of $w(t)$. The measured signals $z_1(t)$ and $z_2(t)$, which are generated using (1) and (2), respectively, are shown in Fig. 2 (only 1000-point data segments are shown).

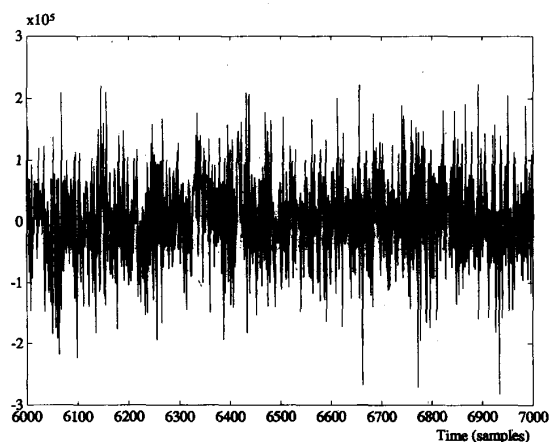
In general, speech is well-modeled by a tenth-order AR process. However, preliminary experimentation with various orders indicated that it is sufficient and sometimes preferable to use a lower order model. Specifically, in this experiment we used a second-order AR model for the speech signal.

In applying the algorithms, only the vector unit sample response coefficients \mathbf{a} of the coupling system A , and the desired signal parameters α and g_s are estimated. The other coupling system B and the spectral levels of $w(t)$, $e_1(t)$, and $e_2(t)$ are assumed to be known.

We first implemented the sequential-EM algorithm that consists of state estimation using the Kalman filtering equations (67)–(71) and parameter estimation using equations (72), (73), and (75). The inverse of the 128×128 matrix in (75) was



(a)



(b)

Fig. 2. (a) Signal measured at the first (primary) sensor; (b) signal measured at the second (reference) sensor.

updated every 100 data samples to save computations. The exponential weights γ_α , γ_s , and γ_a used were all equal to 0.995, corresponding to a sliding window of effective length 138 ($0.995^{138} = 0.5$). In this setting, we have an adaptive algorithm that is capable of tracking the varying characteristics of both the signal (since $\gamma_\alpha, \gamma_s < 1$) and the unknown coupling system (since $\gamma_a < 1$).

We applied the LMS method [18] (under the assumption that B is zero) using the first 300 data samples of the given signals $z_1(t)$ and $z_2(t)$ to obtain an initial estimate of \mathbf{a} and of the speech signal $s(t)$. An initial estimate of α is then obtained by a least squares fit of the initial signal estimate to a second-order AR model. The signal level was arbitrarily set to the initial value of $g_s = 100$. The state vector was initialized using the first $(r+1)$ samples of $z_1(t)$ and the first $(q+1)$ samples of $z_2(t)$, and the state covariance was initially set to be a diagonal matrix whose first $(r+1)$ diagonal elements are all equal to the initial value of g_s , and the other $(q+1)$ diagonal elements are all equal to the noise level g_w .

The results of this experiment are illustrated in Figs. 3–6. The actual and estimated speech signals are shown in Fig. 3.

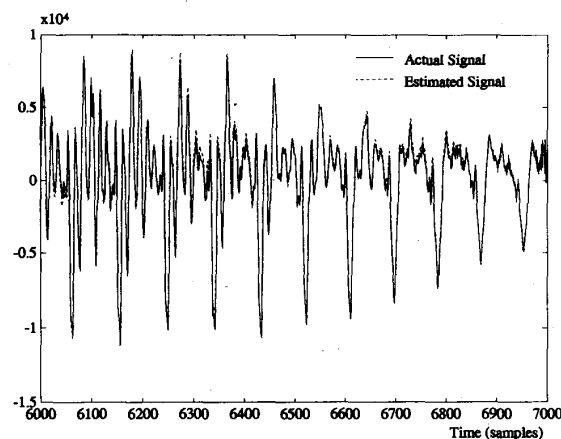


Fig. 3. Estimated signal using the sequential EM algorithm.

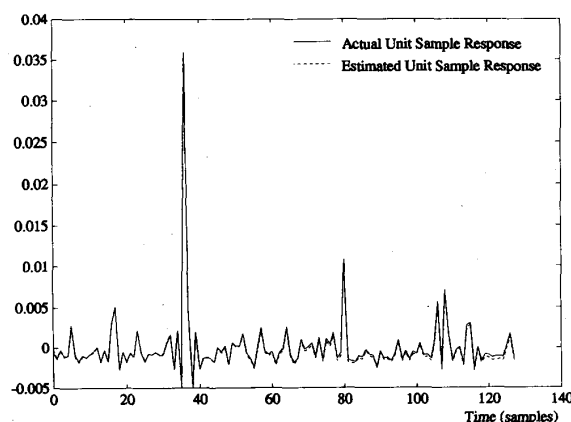


Fig. 4. Estimated unit sample response of the coupling system using the sequential EM algorithm.

The post-processing SNR, that is, the average power of the actual signal $s(t)$ divided by the average power of the error signal $[\hat{s}(t) - s(t)]$, was 23 dB, indicating a 23-dB enhancement relative to the original sensor data. In Fig. 4 we have shown the actual and estimated unit sample response coefficients of A , and in Fig. 5, we have shown the frequency response of the actual and estimated filter. The algorithm converged to an accurate estimate of the filter coefficients in roughly 1000 data samples, corresponding to 0.1 s for the 10-kHz sampling frequency used to generate the data. In Fig. 6(a), we have shown the estimate of the signal level g_s as a function of time. For reference, in Fig. 6(b), we have shown the original speech signal. We see that the estimate of the signal level tracks the signal envelope very closely.

Next, we implemented the sequential-gradient algorithm that consists of performing the parameter estimation update using (98), (99), and (101). The step sizes δ_α , δ_s , and δ_a used were all equal to 0.995. The state vector and its covariance were initialized as before. However, the coupling system A was initially set to zero (i.e., $\mathbf{a} = 0$), and the initial estimate of α was obtained by a least squares fit of a second-order

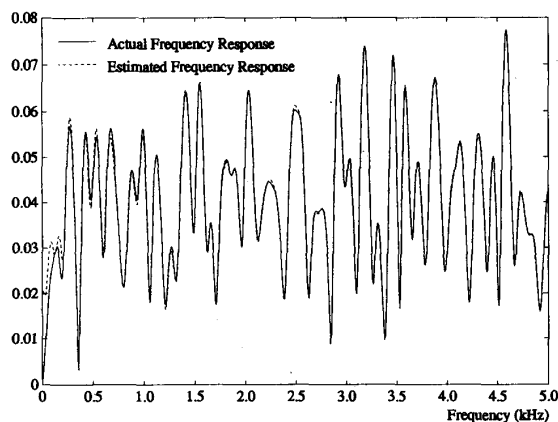


Fig. 5. Estimated frequency response of the coupling system using the sequential EM algorithm.

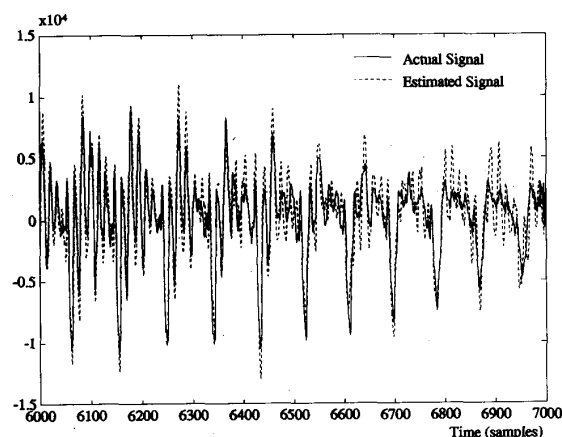
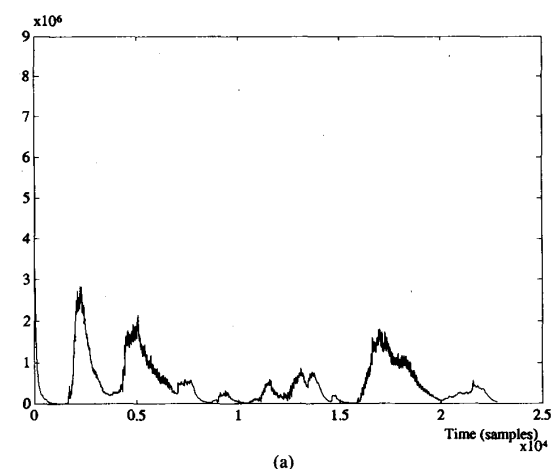
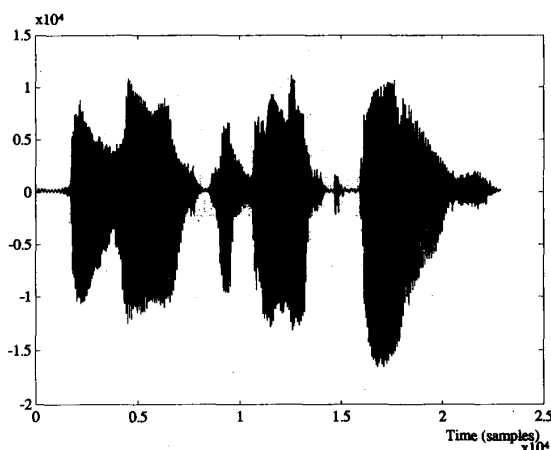


Fig. 7. Estimated speech signal using the sequential gradient algorithm.



(a)



(b)

Fig. 6. (a) Estimated signal level; (b) actual speech level.

AR model to $z_1(t)$ using the first 300 data samples. As before, the signal level was arbitrarily initialized to $g_s = 100$. The actual and estimated speech signals are shown in Fig.

7. The post-processing SNR in this case was 8 dB, which is considerably lower than the 23-dB enhancement achieved with the sequential-EM algorithm. However, a more careful calculation of the post-processing SNR over a sliding window of 100 time samples shows an enhancement of over 10 dB and sometimes 20 dB during the more stationary voiced segments of the speech. The algorithm did the worst during the silence periods and in the transitions into and out of the silence periods. In qualitative listening, the estimated speech exhibited noticeable improvement when compared with the actual sensor data. The background noise was more evident between words, but the actual words were clearly intelligible. However, the perceptual quality of the signal estimate is not as good as that obtained with the sequential-EM algorithm.

The main advantage of the sequential-gradient algorithm is in its computational simplicity. It does not involve any matrix inversion, and consequently, it requires only about 2% of the floating-point operations needed for the sequential-EM algorithm. It also seems to work better with arbitrary initialization. When the initial A was set to zero, the sequential-EM exhibited convergence problems, whereas the sequential-gradient algorithm converged very quickly within 2000 data samples.

Finally, we note that the iterative-batch EM algorithm is expected to perform similarly to the frequency-domain EM algorithm in [4], which resulted in an improvement of 28 dB for this data. Therefore, in this particular experiment, 5 dB of enhancement was sacrificed in making the algorithm sequential and adaptive.

APPENDIX A DEVELOPMENT OF THE EFFICIENT FORM OF THE KALMAN FILTERING EQUATIONS

Let $\mu_{t-1|t-1}$ be partitioned as follows:

$$\mu_{t-1|t-1} = \begin{bmatrix} \mu_s \\ \mu_w \end{bmatrix} \begin{matrix} \uparrow r+1 \\ \uparrow q+1 \end{matrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} \begin{matrix} \uparrow 1 \\ \uparrow r \\ \uparrow 1 \\ \uparrow q \end{matrix}. \quad (\text{A.1})$$

Let μ_p be the lower $p \times 1$ sub-vector of μ_2 (or the lower $p \times 1$ sub-vector of μ_s):

$$\mu_2 = \begin{bmatrix} - \\ \mu_p \end{bmatrix} \uparrow p. \quad (\text{A.2})$$

Let $P_{t-1|t-1}$ be partitioned as follows:

$$P_{t-1|t-1} = \begin{bmatrix} P_{ss} & P_{sw} \\ P_{sw}^\top & P_{ww} \end{bmatrix} \begin{matrix} \uparrow r+1 \\ \uparrow q+1 \end{matrix}$$

$$= \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{12}^\top & P_{22} & P_{23} & P_{24} \\ P_{13}^\top & P_{23}^\top & P_{33} & P_{34} \\ P_{14}^\top & P_{24}^\top & P_{34}^\top & P_{44} \end{bmatrix} \begin{matrix} \uparrow 1 \\ \uparrow r \\ \uparrow 1 \\ \uparrow q \end{matrix} \quad (\text{A.3})$$

$$\begin{matrix} \longleftrightarrow & \longleftrightarrow & \longleftrightarrow & \longleftrightarrow \\ 1 & r & 1 & q \end{matrix}$$

Let Γ_p be the following submatrix of P_{22} :

$$P_{22} = \begin{bmatrix} & \Gamma_p \end{bmatrix} \uparrow r \quad (\text{A.4})$$

$$\begin{matrix} \longleftrightarrow \\ p \end{matrix}$$

and let Γ_{pp} be the following submatrix of Γ_p :

$$\Gamma_p = \begin{bmatrix} - \\ \Gamma_{pp} \end{bmatrix} \uparrow p \quad (\text{A.5})$$

$$\begin{matrix} \longleftrightarrow \\ p \end{matrix}$$

(so that Γ_{pp} is the lower right $p \times p$ submatrix of P_{22}).

Let Λ_p be the following submatrix of P_{24} :

$$P_{24} = \begin{bmatrix} - \\ \Lambda_p \end{bmatrix} \uparrow p \quad (\text{A.6})$$

$$\begin{matrix} \longleftrightarrow \\ q \end{matrix}$$

Let \mathbf{a} and \mathbf{b} be partitioned as follows:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} \begin{matrix} \uparrow q \\ \uparrow 1 \end{matrix} \quad (\text{A.7})$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_0 \end{bmatrix} \begin{matrix} \uparrow r \\ \uparrow 1 \end{matrix} \quad (\text{A.8})$$

For convenience, we shall use θ instead of $\hat{\theta}(t)$, and Φ, G, H , and R instead of $\hat{\Phi}_t, \hat{G}_t, \hat{H}_t$, and \hat{R}_t , respectively.

Substituting (49) and (A.1) into (67)

$$\mu_{t|t-1} = \begin{bmatrix} \Phi_s & 0 \\ 0 & \Phi_w \end{bmatrix} \begin{bmatrix} \mu_s \\ \mu_w \end{bmatrix} = \begin{bmatrix} \Phi_s \mu_s \\ \Phi_w \mu_w \end{bmatrix} \quad (\text{A.9})$$

where

$$\Phi_s \mu_s = \begin{bmatrix} 0 & I \\ 0 & 0^\top & -\alpha^\top \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_2 \\ -\alpha^\top \mu_p \end{bmatrix} \quad (\text{A.10})$$

$$\Phi_w \mu_w = \begin{bmatrix} 0 & I \\ 0 & 0^\top \end{bmatrix} \begin{bmatrix} \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} \mu_4 \\ 0 \end{bmatrix}. \quad (\text{A.11})$$

Substituting (A.10) and (A.11) into (A.9), we obtain

$$\mu_{t|t-1} = \begin{bmatrix} \mu_2 \\ -\alpha^\top \mu_p \\ \mu_4 \\ 0 \end{bmatrix} \begin{matrix} \uparrow r \\ \uparrow 1 \\ \uparrow q \\ \uparrow 1 \end{matrix} \quad (\text{A.12})$$

Substituting (49), (A.3), and (47) into (68)

$$P_{t|t-1} = \begin{bmatrix} \Phi_s & 0 \\ 0 & \Phi_w \end{bmatrix} \begin{bmatrix} P_{ss} & P_{sw} \\ P_{sw}^\top & P_{ww} \end{bmatrix} \begin{bmatrix} \Phi_s^\top & 0 \\ 0 & \Phi_w^\top \end{bmatrix} + GG^\top$$

$$= \begin{bmatrix} \Phi_s P_{ss} \Phi_s^\top & \Phi_s P_{sw} \Phi_w^\top \\ \Phi_w P_{sw}^\top \Phi_s^\top & \Phi_w P_{ww} \Phi_w^\top \end{bmatrix} + GG^\top \quad (\text{A.13})$$

where

$$\Phi_s P_{ss} \Phi_s^\top = \begin{bmatrix} 0 & I \\ 0 & 0^\top & -\alpha^\top \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^\top & P_{22} \end{bmatrix}$$

$$= \begin{bmatrix} P_{22} & -\Gamma_p \alpha \\ -\alpha^\top \Gamma_p & \alpha^\top \Gamma_{pp} \alpha \end{bmatrix}, \quad (\text{A.14})$$

$$\Phi_s P_{sw} \Phi_w^\top = \begin{bmatrix} 0 & I \\ 0 & 0^\top & -\alpha^\top \end{bmatrix} \begin{bmatrix} P_{13} & P_{14} \\ P_{23} & P_{24} \end{bmatrix}$$

$$= \begin{bmatrix} P_{24} & 0 \\ -\alpha^\top \Lambda_p & 0 \end{bmatrix}, \quad (\text{A.15})$$

$$\Phi_w P_{ww} \Phi_w^\top = \begin{bmatrix} 0 & I \\ 0 & 0^\top \end{bmatrix} \begin{bmatrix} P_{33} & P_{34} \\ P_{34}^\top & P_{44} \end{bmatrix}$$

$$= \begin{bmatrix} P_{44} & 0 \\ 0^\top & 0 \end{bmatrix}, \quad (\text{A.16})$$

and GG^\top is the matrix having only two nonzero elements

$$GG^\top = \begin{bmatrix} & g_s & & \\ & & g_w & \\ & & & \\ & & & \end{bmatrix} \begin{matrix} \leftarrow r+1 \\ \leftarrow r+q+2 \\ \\ \end{matrix} \quad (\text{A.17})$$

$$\begin{matrix} \uparrow & \uparrow \\ r+1 & r+q+2 \end{matrix}$$

Substituting (A.14)–(A.17) into (A.13), we obtain

$$P_{t|t-1} = \begin{bmatrix} P_{22} & -\Gamma_p \alpha & P_{24} & 0 \\ \alpha^\top \Gamma_{pp} \alpha + g_s & -\alpha^\top \Lambda_p & 0 & 0 \\ & & P_{44} & 0 \\ & & & g_w \end{bmatrix} \begin{matrix} \uparrow r \\ \uparrow 1 \\ \uparrow q \\ \uparrow 1 \end{matrix} \quad (A.18)$$

$\begin{matrix} \longleftrightarrow & \longleftrightarrow & \longleftrightarrow & \longleftrightarrow \\ r & 1 & q & 1 \end{matrix}$

where we note that $P_{t|t-1}$ is a symmetric matrix; therefore, the lower blocks in (A.18) can be completed accordingly.

Substituting (A.1) and (A.18) into (69)–(71), and following straightforward matrix manipulations, we obtain

$$\mu_{t|t} = \mu_{t|t-1} + D_t F_t^{-1} \cdot \begin{bmatrix} z_1(t) + \alpha^\top \mu_p - a_1^\top \mu_4 \\ z_2(t) + b_0 \alpha^\top \mu_p - b_1^\top \mu_2 \end{bmatrix} \quad (A.19)$$

$$P_{t|t} = P_{t|t-1} - D_t F_t^{-1} D_t^\top \quad (A.20)$$

where we have defined D_t in (A.21), which appears at the bottom of this page, and F_t is the 2×2 symmetric matrix:

$$F_t = \begin{bmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{bmatrix} \quad (A.22)$$

where

$$f_{11} = a_1^\top P_{44} a_1 - 2a_1^\top \Lambda_p^\top \alpha + \alpha^\top \Gamma_{pp} \alpha + g_s + a_0^2 g_w + g_1 \quad (A.23)$$

$$f_{22} = b_1^\top P_{22} b_1 - 2b_0 \cdot b_1^\top \Gamma_p \alpha + b_0^2 (\alpha^\top \Gamma_{pp} \alpha + g_s) + g_w + g_2 \quad (A.24)$$

$$f_{12} = a_1^\top P_{24} b_1 - b_0 \cdot a_1^\top \Lambda_p^\top \alpha - b_1^\top \Gamma_p \alpha + b_0 (\alpha^\top \Gamma_{pp} \alpha + g_s) + a_0 g_w. \quad (A.25)$$

Equations (A.12) and (A.18) constitute the propagation equations, and (A.19) and (A.20) constitute the updating equations.

APPENDIX B

RECURSIVE IMPLEMENTATION OF (72)–(78)

Define

$$R_{11}(t) = \sum_{\tau=1}^t \gamma_s^{t-\tau} \widehat{s_{p-1}(\tau-1|\tau)} s_{p-1}^\top(\tau-1|\tau) = \gamma_s R_{11}(t-1) + \widehat{s_{p-1}(t-1|t)} s_{p-1}^\top(t-1|t) \quad (B.1)$$

$$R_{12}(t) = \sum_{\tau=1}^t \gamma_s^{t-\tau} \widehat{s_{p-1}(\tau-1|\tau)} s(\tau|\tau) = \gamma_s R_{12}(t-1) + \widehat{s_{p-1}(t-1|t)} s(t|t) \quad (B.2)$$

$$R_{22}(t) = \sum_{\tau=1}^t \gamma_s^{t-\tau} \widehat{s^2(\tau|\tau)} = \gamma_s R_{22}(t-1) + \widehat{s^2(t|t)} \quad (B.3)$$

$$Q_{11}(t) = \sum_{\tau=1}^t \gamma_w^{t-\tau} \widehat{w^2(\tau|\tau)} = \gamma_w Q_{11}(t-1) + \widehat{w^2(t|t)} \quad (B.4)$$

$$A_{11}(t) = \sum_{\tau=1}^t \gamma_a^{t-\tau} \widehat{w_q(\tau|\tau)} w_q^\top(\tau|\tau) = \gamma_a A_{11}(t-1) + \widehat{w_q(t|t)} w_q^\top(t|t) \quad (B.5)$$

$$A_{12}(t) = \sum_{\tau=1}^t \gamma_a^{t-\tau} [\widehat{w_q(\tau|\tau)} z_1(\tau) - \widehat{w_q(\tau|\tau)} s(\tau|\tau)] = \gamma_a A_{12}(t-1) + \widehat{w_q(t|t)} z_1(t) - \widehat{w_q(t|t)} s(t|t) \quad (B.6)$$

$$A_{22}(t) = \sum_{\tau=1}^t \gamma_a^{t-\tau} [z_1^2(\tau) - 2\hat{s}(\tau|\tau) z_1(\tau) + \widehat{s^2(\tau|\tau)}] = \gamma_a A_{22}(t-1) + \widehat{z_1^2(t|t)} - 2\hat{s}(t|t) z_1(t) + \widehat{s^2(t|t)} \quad (B.7)$$

$$B_{11}(t) = \sum_{\tau=1}^t \gamma_b^{t-\tau} \widehat{s_r(\tau|\tau)} s_r^\top(\tau|\tau) = \gamma_b B_{11}(t-1) + \widehat{s_r(t|t)} s_r^\top(t|t) \quad (B.8)$$

$$B_{12}(t) = \sum_{\tau=1}^t \gamma_b^{t-\tau} [\widehat{s_r(\tau|\tau)} z_2(\tau) - \widehat{s_r(\tau|\tau)} w(\tau|\tau)] = \gamma_b B_{12}(t-1) + \widehat{s_r(t|t)} z_2(t) - \widehat{s_r(t|t)} w(t|t) \quad (B.9)$$

$$B_{22}(t) = \sum_{\tau=1}^t \gamma_b^{t-\tau} [z_2^2(\tau) - 2\hat{w}(\tau|\tau) z_2(\tau) + \widehat{w^2(\tau|\tau)}] = \gamma_b B_{22}(t-1) + \widehat{z_2^2(t|t)} - 2\hat{w}(t|t) z_2(t) + \widehat{w^2(t|t)}. \quad (B.10)$$

$$D_t = \begin{bmatrix} -\Gamma_p \alpha + P_{24} a_1 & P_{22} b_1 - \Gamma_p \alpha \cdot b_0 \\ \alpha^\top \Gamma_{pp} \alpha + g_s - \alpha^\top \Gamma_p a_1 & -\alpha^\top \Gamma_p^\top b_1 + b_0 (\alpha^\top \Gamma_{pp} \alpha + g_s) \\ -\Lambda_p^\top \alpha + P_{44} a_1 & P_{24}^\top b_1 - \Lambda_p^\top \alpha \cdot b_0 \\ a_0 \cdot g_w & g_w \end{bmatrix} \begin{matrix} \uparrow r \\ \uparrow 1 \\ \uparrow q \\ \uparrow 1 \end{matrix} \quad (A.21)$$

$\begin{matrix} \longleftrightarrow & \longleftrightarrow \\ 1 & 1 \end{matrix}$

Then, using (B.1)–(B.10) in (72)–(78), we obtain the following recursive formulas for updating the parameter estimates:

$$\begin{aligned}\hat{\alpha}(t+1) &= -R_{11}^{-1}(t)R_{12}(t) \\ &= \hat{\alpha}(t) - R_{11}^{-1}(t)[\overbrace{s_{p-1}(t-1|t)s(t|t)}^{\text{cross-term}}] \\ &\quad + \overbrace{s_{p-1}(t-1|t)s_{p-1}^T(t-1|t)\hat{\alpha}(t)}^{\text{self-term}}\end{aligned}\quad (\text{B.11})$$

$$\hat{g}_s(t+1) = \frac{1-\gamma_s}{1-\gamma_s^t}[R_{22}(t) + \hat{\alpha}^T(t+1)R_{12}(t)] \quad (\text{B.12})$$

$$\hat{g}_w(t+1) = \frac{1-\gamma_w}{1-\gamma_w^t}Q_{11}(t) \quad (\text{B.13})$$

$$\begin{aligned}\hat{\mathbf{a}}(t+1) &= A_{11}^{-1}(t)A_{12}(t) \\ &= \hat{\mathbf{a}}(t) + A_{11}^{-1}(t)[\overbrace{\hat{\mathbf{w}}_q(t|t)z_1(t)}^{\text{cross-term}} \\ &\quad - \overbrace{\mathbf{w}_q(t|t)s(t|t) - \mathbf{w}_q(t|t)\mathbf{w}_q^T(t|t)\hat{\mathbf{a}}(t)}^{\text{self-term}}]\end{aligned}\quad (\text{B.14})$$

$$\hat{g}_1(t+1) = \frac{1-\gamma_a}{1-\gamma_a^t}[A_{22}(t) - \hat{\mathbf{a}}^T(t+1)A_{12}(t)] \quad (\text{B.15})$$

$$\begin{aligned}\hat{\mathbf{b}}(t+1) &= B_{11}^{-1}(t)B_{12}(t) \\ &= \hat{\mathbf{b}}(t) + B_{11}^{-1}(t)[\overbrace{\hat{\mathbf{s}}_r(t|t)z_2(t)}^{\text{cross-term}} \\ &\quad - \overbrace{\mathbf{s}_r(t|t)\mathbf{s}_r^T(t|t)\hat{\mathbf{b}}(t)}^{\text{self-term}}]\end{aligned}\quad (\text{B.16})$$

$$\hat{g}_2(t) = \frac{1-\gamma_b}{1-\gamma_b^t}[B_{22}(t) - \hat{\mathbf{b}}^T(t+1)B_{12}(t)]. \quad (\text{B.17})$$

REFERENCES

- [1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [2] K. J. Åström, "Maximum likelihood and prediction error methods," in *Trends and Progress in System Identification*, (P. Eykhoff, Ed.) Oxford: Pergamon, 1981, pp. 145–168, ch. 5.
- [3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc.*, pp. 1–38, 1977, ser. 3g.
- [4] M. Feder, A. V. Oppenheim, and E. Weinstein, "Maximum likelihood noise cancellation using the EM algorithm," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 37, no. 2, pp. 204–216, Feb. 1989.
- [5] R. A. Fisher, "Theory of statistical estimation," *Proc. Cambridge Phil. Soc.*, vol. 22, pp. 700–725, 1925.
- [6] N. K. Gupta and R. K. Mehra, "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 774–783, Dec. 1974.
- [7] J. S. Lim and A. V. Oppenheim, "All pole modelling of degraded speech," *IEEE Trans. Acoustics Speech Signal Processing*, vol. ASSP-26, pp. 197–210, 1978.
- [8] J. S. Lim (Ed.), *Speech Enhancement*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [9] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
- [10] T. A. Louis, "Finding the observed information matrix when using the EM algorithm," *J. Royal Stat. Soc.*, vol. 44, no. 2, pp. 226–233, 1982, ser. B.
- [11] J. Meilijson, "A fast improvement to the EM algorithm on its own terms," *J. Royal Stat. Soc.*, vol. 51, no. 1, pp. 127–138, 1989, ser. B.
- [12] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," Ph.D. dissertation, Mass. Inst. Technol., Sept. 1982.
- [13] A. V. Oppenheim, E. Weinstein, K. Zangi, M. Feder, and D. Gauger, "Single sensor active noise cancellation based on the EM algorithm," *IEEE Trans. Signal Processing*, Apr. 1994.
- [14] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973, 2nd ed.
- [15] M. Segal and E. Weinstein, "A new method for evaluating the log-likelihood gradient (score) of linear dynamic systems," *IEEE Trans. Automat. Contr.*, vol. 33, no. 8, pp. 763–766, Aug. 1988.
- [16] M. Segal and E. Weinstein, "A new method for evaluating the score, Hessian, and the Fisher's information matrix of discrete-time dynamic systems," *IEEE Trans. Inform. Theory*, vol. 35, no. 3, pp. 682–687, May 1989.
- [17] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.
- [18] D. M. Titterton, "Recursive parameter estimation using incomplete data," *J. Royal Stat. Soc.*, vol. 46, no. 2, pp. 257–267, 1984.
- [19] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation based on decorrelation," *IEEE Trans. Audio Speech*, Oct. 1993, pp. 504–519 (also RLE Techn. Rep. No. 573, M.I.T., Cambridge, MA., Mar. 1992).
- [20] E. Weinstein, A. V. Oppenheim, and M. Feder, "Signal enhancement using single and multi-sensor measurements," RLE Techn. Rep. No. 560, M.I.T., Cambridge, MA, Nov. 1990.
- [21] B. Widrow *et al.*, "Adaptive noise canceling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, 1975.
- [22] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Annals Stat.*, vol. 11, pp. 95–103, 1983.

Ehud Weinstein (F'94), for photograph and biography, please see p. 413 of the October 1993 issue of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING.

Alan V. Oppenheim (F'77), for photograph and biography, please see p. 413 of the October 1993 issue of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING.

Meir Feder (SM'94), for photograph and biography, please see p. 413 of the October 1993 issue of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING.

John R. Buck (S'91) received S.B. degrees in electrical engineering and english literature from the Massachusetts Institute of Technology (MIT) in 1989. He received the S.M. and E.E. degrees jointly from M.I.T. and the Woods Hole Oceanographic Institution (WHOI) in 1991 and 1992, respectively. Since 1990 he has been enrolled in the MIT-WHOI Joint Program for Applied Ocean Science and Engineering. Mr. Buck is a member of Sigma Xi and Phi Beta Kappa.