Filtering of Colored Noise for Speech Enhancement and Coding

Jerry D. Gibson, *Senior Member, IEEE*, Boneung Koo, and Steven D. Gray

Abstract— **Scalar and vector Kalman filters are implemented for filtering speech contaminated by additive white noise or colored noise and an iterative signal and parameter estimator which can be used for both noise types is presented. Particular emphasis is placed on the removal of colored noise, such as helicopter noise, by using state-of-the-art colored noise assumption Kalman filters. The results indicate that the colored noise Kalman filters provide a significant gain in SNR, a visible improvement in the sound spectrogram, and an audible improvement in output speech quality, none of which are available with white noise assumption Kalman and Wiener filters. When the filter is employed as a prefilter for linear predictive coding, the coded output speech quality and intelligibility is greatly enhanced in comparison to direct coding of the noisy speech.**

I. Introduction

The effects of additive noise, particularly colored noise, on speech coders designed for noise-free speech can be substantial, even to the point of rendering the system unusable [1]–[3]. There have been numerous studies involving the enhancement of white noise contaminated speech, including the work of Lim and Oppenheim [4] and a recent investigation employing a white noise assumption Kalman filter [5]. As noted by several authors [6]–[10], minimum mean squared error (MMSE) encoding of a noisy source can be realized by the cascade of a MMSE filter followed by MMSE encoding of the filter output. Motivated by this result, Gibson *et al.* [11] utilized a white noise assumption Kalman filter cascaded with a vector quantizer for vector quantization of speech plus white noise. Ephraim and Gray [12] take an alternative approach of modifying the distortion measure to account for a signal plus noise input, and then use the Lloyd algorithm to design a vector quantizer for speech in additive white Gaussian noise.

Some unifying results for the general signal reconstruction/parameter identification problem given noisy measurements are provided by Musicus [13], [35], [36] by starting with the concept of cross entropy. There is a vast literature on signal reconstruction for general signals, with a recent contribution being that of Dembo [14].

Manuscript received February 17, 1989; revised June 21, 1990.

J. D. Gibson is with the Department of Electrical Engineering, Texas A&M University, College Station TX 77843.

B. Koo is with the Department of Electronic Engineering, Kyonggi University, Suwon, South Korea, 440-760.

S. D. Gray is with MITRE Corp., Bedford, MA.

IEEE Log Number 9100788.

The general method employed in this paper for coding noisy speech utilizes the cascade estimator/encoder structure and falls within the alphabet constrained approach to data compression developed by Gibson and Fischer [8], since overlapping, sliding block algorithms are used (unlike [4], [6], for example). The estimators examined include both white and colored noise assumption, scalar and vector output Kalman filters. The determination of the several Kalman filter parameters and the actual filtering of the speech are performed iteratively, analogous to [4]. We do not employ constraints on estimated spectra, as in [16], although such constraints could be incorporated in future work.

After developing the scalar and vector output filtering algorithms and considering the parameter identification iterations, we study the performance of these filters for the enhancement of speech contaminated by additive white or colored noise. These filters are then employed in the cascade estimator/encoder structure for linear predictive coding of speech additively distorted by helicopter noise. Performance comparisons are accomplished in terms of signal-to-noise ratio, sound spectrograms, and subjective listening tests.

II. Problem Statement

Let the noise-free source sequence $\{x(n)\}$ be generated according to the pth order autoregressive (AR) model

$$x(i) = \sum_{j=1}^{p} a_j x(i-j) + w(i) \tag{1}$$

where $w(i)$ is a zero mean, white Gaussian process with variance σ_w^2. The source sequence is then contaminated by zero mean additive Gaussian noise, which is either white or colored but independent of $x(i)$, so that the noisy source samples presented to an encoder are

$$s(i) = x(i) + v(i). \tag{2}$$

The structure we study here is shown in Fig. 1 and consists of an estimator cascaded with the encoder designed for noise-free speech. For the squared error distortion criterion, a similar structure that consists of an estimator cascaded with an encoder matched to the estimator output can be shown to be optimal (see [6], [8], [10]). The Gaussian assumptions are useful for developing algorithms, but neither the speech nor the noise may be Gauss-

1053-587X/91/0800-1732$01.00

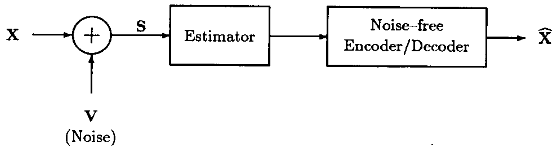


Fig. 1. Cascade estimator/coder configuration.

ian. The viability of the approach is demonstrated by experiment.

In the sequel, we present algorithms for estimating the source sequence $\{x(i)\}$, its AR parameters $\{a_1, \cdots, a_p\}$, and σ_w^2 given the noisy sequence $\{s(i)\}$ of length L. We assume that an AR model of the noise is known or that the autocorrelation properties of the noise can be measured, perhaps by a separate microphone. Details of the noise model are elaborated in subsequent sections.

The boldface symbols in Fig. 1 allow the possibility that these quantities are vectors. Any vector dimension appropriate for the application of interest can be used. Here we report results for algorithms that produce a single (scalar) filtered value at a time, as well as algorithms that generate vector-valued estimates.

III. Kalman Filter Algorithms

Assume that $N \geq 1$ consecutive samples are blocked together to form vectors $\boldsymbol{x}(n)$, $\boldsymbol{s}(n)$, and $\boldsymbol{v}(n)$ at time instant n. Then, the MMSE estimate is

$$\hat{\boldsymbol{x}}(n) = E[\boldsymbol{x}(n) \mid \boldsymbol{s}(n), \boldsymbol{s}(n-1), \cdots] \tag{3}$$

which can be computed by using a Kalman filtering algorithm. Since an estimate of the vector $\boldsymbol{x}(n)$ is produced at each time instant, we call a direct implementation of (3) the vector Kalman filter. Expanding (3), we get

$$E\{\boldsymbol{x}(n) \mid \boldsymbol{s}(n), \boldsymbol{s}(n-1), \cdots\} = \begin{bmatrix} E\{x(nN) \mid s(nN), s(nN-1), \cdots\} \\ E\{x(nN-1) \mid s(nN), s(nN-1), \cdots\} \\ \vdots \\ E\{x(nN-N+1) \mid s(nN), s(nN-1), \cdots\} \end{bmatrix} \tag{4}$$

which can be seen to be a smoother on a component-by-component basis, because all the vector elements except the first one use future observations to estimate the current sample. Note that the indexing on nN causes shifts of vector dimension or blocklength N. Thus, n is a vector time index.

We can also approximate (4) as

$$E\{\boldsymbol{x}(n) \mid \boldsymbol{s}(n), \boldsymbol{s}(n-1), \cdots\} \cong \begin{bmatrix} E\{x(nN) \mid s(nN), s(nN-1), \cdots\} \\ E\{x(nN-1) \mid s(nN-1), s(nN-2), \cdots\} \\ \vdots \\ E\{x(nN-N+1) \mid s(nN-N+1), s(nN-N), \cdots\} \end{bmatrix} \tag{5}$$

which on a component-by-component basis consists of N true filters since only the current and past observations are used for estimating the current element. We call the componentwise filtering implementation in (5) a scalar Kalman filter in contrast to the exact structure of (4). Comparing each element of (5) with that of (4), we see that while the first elements are the same, the given conditions on other elements in (5) are subsets of those in (4). Hence, the estimator in (5) can perform no better than that in (4), although the estimator based on (5) is simpler to implement.

In the following subsections, we develop the scalar and vector Kalman filter algorithms for both the white and colored measurement noise cases. We begin with the white noise algorithms.

A. *Scalar Kalman Filter—White Measurement Noise*

The measurement or observation noise sequence $\{v(i)\}$ is assumed in this section to be white and Gaussian with zero mean and variance σ_v^2. To develop white noise assumption scalar Kalman filtering algorithms, we begin with the assumed AR source model in (1) and the observation model in (2) and write canonical state space models of the form

$$\boldsymbol{x}(n) = \boldsymbol{F}\boldsymbol{x}(n-1) + \boldsymbol{g}w(n) \tag{6}$$

$$s(n) = \boldsymbol{h}^T\boldsymbol{x}(n) + v(n) \tag{7}$$

with $\boldsymbol{x}(n) \triangleq [x(n-p+1)x(n-p+2) \cdots x(n)]$

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_2 & a_1 \end{bmatrix}$$

$$\boldsymbol{g} = \boldsymbol{h} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

where the driving term $\{w(n)\}$ is a white Gaussian sequence with zero mean and variance σ_w^2, and the measurement noise $\{v(n)\}$ is a white Gaussian noise sequence with zero mean and variance σ_v^2. Note that we let $N = p$ so that the filter order is the same as the order of the AR source model.

From standard Kalman filtering theory for white message and measurement noises [17], the state vector estimate is

$$\hat{\boldsymbol{x}}(n) = \boldsymbol{F}\hat{\boldsymbol{x}}(n-1) + \boldsymbol{k}(n)[s(n) - \boldsymbol{h}^T\boldsymbol{F}\hat{\boldsymbol{x}}(n-1)] \quad (8)$$

with the initial condition $\hat{\boldsymbol{x}}^T(0) = \boldsymbol{0}$. The gain and error covariance equations are

$$\boldsymbol{k}(n) = \boldsymbol{P}(n|n-1)\boldsymbol{h}[\boldsymbol{R} + \boldsymbol{h}^T\boldsymbol{P}(n|n-1)\boldsymbol{h}]^{-1} \quad (9)$$

$$\boldsymbol{P}(n|n-1) = \boldsymbol{F}\boldsymbol{P}(n-1)\boldsymbol{F}^T + \boldsymbol{g}\boldsymbol{Q}\boldsymbol{g}^T \quad (10)$$

$$\boldsymbol{P}(n) = [\boldsymbol{I} - \boldsymbol{k}(n)\boldsymbol{h}^T]\boldsymbol{P}(n|n-1) \quad (11)$$

where $\boldsymbol{k}(n)$ is a Kalman gain vector, $\boldsymbol{P}(n|n-1)$ is an *a priori* error covariance matrix, $\boldsymbol{P}(n)$ is an error covariance matrix, $\boldsymbol{R} = \sigma_v^2$ is the variance of the noise sequence $\{v(n)\}$, and $\boldsymbol{Q} = \sigma_w^2$ is the variance of the driving term $\{w(n)\}$. With the initial condition, $\boldsymbol{P}(0) = [0]_{N \times N}$, (10) is processed first, followed by (9), and then (8) and (11). The speech sample estimate at time instant n is finally obtained by

$$\hat{x}(n) = \boldsymbol{h}^T\hat{\boldsymbol{x}}(n).$$

The parameters $\boldsymbol{a} \triangleq [a_1 \cdots a_p]$ and σ_w^2 are to be estimated from the noisy observations when the true values are not known. This step is treated in Section IV.

B. Vector Kalman Filter—White Measurement Noise

To implement the white noise assumption vector Kalman filter, we rewrite the AR(p) source model in (1) into an appropriate vector form. For any vector dimension $N = pm$ with $m = 1, 2, \cdots$, the pth order Gauss–Markov source model can be written as

$$\boldsymbol{x}(n) = \Phi\boldsymbol{x}(n-1) + \Gamma\boldsymbol{w}(n) \quad (12)$$

$$s(n) = \boldsymbol{x}(n) + \boldsymbol{v}(n) \quad (13)$$

where $\boldsymbol{x}(n) \triangleq [x(nN)x(nN-1) \cdots x(nN-N+1)]^T$, $\boldsymbol{w}(n) \triangleq [w(nN)w(nN-1) \cdots w(nN-N+1)]^T$

$$\Phi = \begin{bmatrix} A^m & \boldsymbol{0} \\ A^{m-1} & \boldsymbol{0} \\ \vdots & \vdots \\ A^1 & \boldsymbol{0} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} B & AB & \cdots & A^{m-1}B \\ \boldsymbol{0} & B & \cdots & A^{m-2}B \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & B \end{bmatrix} \quad (14)$$

the N-vector $\boldsymbol{w}(n)$ is a white Gaussian process with zero mean and covariance matrix $\boldsymbol{Q} = \sigma_w^2\boldsymbol{I}$, and the N-vector $\boldsymbol{v}(n)$ is a white Gaussian process with zero mean and covariance matrix $\boldsymbol{R} = \sigma_v^2\boldsymbol{I}$. The matrices A and B can be obtained recursively as

$$A = \begin{bmatrix} A_p^T \\ A_{p-1}^T \\ \vdots \\ A_1^T \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 0 & \beta_1 & \cdots & \beta_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_1 \end{bmatrix} \quad (15)$$

where $A_1^T = [a_1 \ a_2 \ \cdots \ a_p]$, $\beta_1 = 1$, and for $k = 2, \cdots, p$,

$$A_k^T = \sum_{i=1}^{k-1} a_i A_{k-i}^T + \Big[a_k \ a_{k+1} \ \cdots \ a_p \ \underbrace{0 \cdots 0}_{(k-1)}\Big]$$

$\beta_k = \sum_{i=1}^{k-1} a_i\beta_{k-i}$. The state estimates are computed by

$$\hat{\boldsymbol{x}}(n) = \hat{\boldsymbol{x}}(n|n-1) + \boldsymbol{K}(n)[\boldsymbol{s}(n) - \hat{\boldsymbol{x}}(n|n-1)] \quad (16)$$

where $\hat{\boldsymbol{x}}(n|n-1) = \Phi\hat{\boldsymbol{x}}(n-1)$ and $\hat{\boldsymbol{x}}^T(0) = \boldsymbol{0}$. The gain and error covariance equations are

$$\boldsymbol{K}(n) = \boldsymbol{P}(n|n-1)[\boldsymbol{R} + \boldsymbol{P}(n|n-1)]^{-1} \quad (17)$$

$$\boldsymbol{P}(n|n-1) = \Phi\boldsymbol{P}(n-1)\Phi^T + \Gamma\boldsymbol{Q}\Gamma^T \quad (18)$$

$$\boldsymbol{P}(n) = [\boldsymbol{I} - \boldsymbol{K}(n)]\boldsymbol{P}(n|n-1) \quad (19)$$

where $\boldsymbol{K}(n)$ is a Kalman gain matrix, $\boldsymbol{P}(n|n-1)$ is an *a priori* error covariance matrix, and $\boldsymbol{P}(n)$ is an error covariance matrix. In this work, $m = 1$ or $N = p$ is used so that $\Phi = A$ and $\Gamma = B$.

C. Scalar Kalman Filter—Colored Noise

To model the colored noise in this section and Section III-D, we assume that the noise is wide sense stationary and is adequately described by the AR(q) model

$$v(n) = \sum_{i=1}^{q} b_i v(n-i) + \eta(n) \quad (20)$$

where $\{\eta(n)\}$ is a white Gaussian sequence with zero mean and variance σ_η^2. Rewriting (20) into canonical state space form, we get

$$\boldsymbol{v}(n) = \boldsymbol{F}_v\boldsymbol{v}(n-1) + \boldsymbol{g}_v\eta(n) \quad (21)$$

$$v(n) = \boldsymbol{h}_v^T\boldsymbol{v}(n) \quad (22)$$

where $\boldsymbol{v}(n) = [v(n-q+1)\,v(n-q+2) \cdots v(n)]^T$

$$\boldsymbol{F}_v = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ b_q & b_{q-1} & b_{q-2} & \cdots & b_2 & b_1 \end{bmatrix}$$

$$\boldsymbol{g}_v = \boldsymbol{h}_v = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (23)$$

and $\{\eta(n)\}$ is a white Gaussian driving term with zero mean and variance σ_η^2.

We now augment (6) and (7) with (21) and (22) to obtain

$$\overline{\boldsymbol{x}}(n) = \overline{\boldsymbol{F}}\overline{\boldsymbol{x}}(n-1) + \overline{\boldsymbol{G}}\overline{\boldsymbol{w}}(n) \quad (24)$$

$$s(n) = \bar{h}^T \bar{x}(n) \tag{25}$$

where

$$\bar{x}(n) = \begin{bmatrix} x(n) \\ v(n) \end{bmatrix}, \quad \bar{w}(n) = \begin{bmatrix} w(n) \\ \eta(n) \end{bmatrix} \tag{26}$$

and

$$\bar{F} = \begin{bmatrix} F & 0 \\ 0 & F_v \end{bmatrix}, \quad \bar{G} = \begin{bmatrix} g & 0 \\ 0 & g_v \end{bmatrix}, \quad \bar{h}^T = [h^T \quad h_v^T]. \tag{27}$$

This is in the form of a linear system driven by a white Gaussian vector $\bar{w}(n)$, but from which are available perfect measurements of certain linear combinations of states. This is the so-called, noise-free [17] or perfect measurements [18] problem in the estimation literature. We assume that $w(n)$ and $\eta(n)$ are uncorrelated such that

$$Q^\star \triangleq E[\bar{w}(n)\bar{w}^T(n)] = \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & \sigma_\eta^2 \end{bmatrix}. \tag{28}$$

The optimal algorithm for noise-free estimation has the same form as that for the white noise case given in (8)–(11) except that we have $R = 0$ and different vectors and matrices here. Before using that algorithm, we have to introduce a coordinate transformation in order to reduce the dimension of the optimal filter [19]. Also, by reducing the order, the invertibility of $h^T P(n|n-1) h$ in (9) is guaranteed [18].

To derive the reduced order optimal filter, one has to define a transformation appropriately so that the perfect measurements are in fact the direct measurements of individual state variables. Following [18], we define a $(p+q) \times (p+q)$ transformation matrix T as

$$T = \begin{bmatrix} \bar{h}^T \\ T_2 \end{bmatrix}$$

where T_2 can be any $(p+q-m) \times (p+q)$ matrix that yields nonsingular T. The dimension of $s(n)$ is denoted by m, which is 1 in (25). Here, we choose $T_2 = [I_{p+q-1} \; 0_1]$ so that the new state vector becomes identical with part of the original state vector of interest. (0_1 denotes a zero vector of $p+q-1$ elements.)

By using this transformation, we have

$$\tilde{x}(n) = T\bar{x}(n), \quad \bar{x}(n) = T^{-1}\tilde{x}(n). \tag{29}$$

Premultiplying (24) with T, we obtain

$$\tilde{x}(n) = \tilde{F}\tilde{x}(n-1) + \tilde{G}\bar{w}(n) \tag{30}$$

$$s(n) = \tilde{h}^T \tilde{x}(n) \tag{31}$$

with $\tilde{F} = T\bar{F}T^{-1}$, $\tilde{G} = T\bar{G}$, and $\tilde{h}^T = \bar{h}^T T^{-1} = [1 \; 0 \cdots 0]$. The Kalman filter estimate becomes

$$\hat{\tilde{x}}(n) = \hat{\tilde{x}}(n|n-1) + k(n)\{s(n) - \tilde{h}^T \hat{\tilde{x}}(n|n-1)\} \tag{32}$$

with $\hat{\tilde{x}}(n|n-1) = \tilde{F}\hat{\tilde{x}}(n-1)$, and the gain and error covariance equations evolve according to

$$k(n) = P(n|n-1)\tilde{h}[\tilde{h}^T P(n|n-1)\tilde{h}]^{-1} \tag{33}$$

$$P(n|n-1) = \tilde{F}P(n-1)\tilde{F}^T + \tilde{G}Q^\star\tilde{G}^T \tag{34}$$

$$P(n) = [I - k(n)\tilde{h}^T]P(n|n-1) \tag{35}$$

where $Q^\star$ is given by (28). With initial conditions $\hat{\tilde{x}}(0) = 0$ and $P(0) = 0$, (34) is processed first, followed by (33) and then (32) and (35). The state estimate (32) has to be transformed back to the original form by (29), so

$$\hat{\bar{x}}(n) = T^{-1}\hat{\tilde{x}}(n) \tag{36}$$

then our final estimate becomes by (26)

$$\hat{x}(n) = [h^T \quad 0]\hat{\bar{x}}(n). \tag{37}$$

The minimal order of the optimal Kalman filter for the noise-free problem equals $p + q - m$, which is, in this case, $p + q - 1$. This can be verified in (29) by noting that the first element of the vector $\tilde{x}(n)$ is exactly $s(n)$ and hence, does not have to be estimated in (32). However, because (33)–(35) still have dimension $p + q$ and the reduction of dimension in (32) is only $m = 1$, the computational burden is only slightly reduced.

D. Vector Kalman Filter—Colored Measurement Noise

Similar to the white noise case in (12) and (13), one can express the vector models for the source and measurement as

$$x(n) = A_x x(n-1) + B_x w(n) \tag{38}$$

$$s(n) = C_x x(n) + v(n) \tag{39}$$

where

$$x(n) \triangleq [x(nN)x(nN-1) \cdots x(nN-N+1)]^T$$

$$w(n) \triangleq [w(nN)w(nN-1) \cdots w(nN-N+1)]^T$$

the N-vector $w(n)$ is a white Gaussian process with zero mean and covariance matrix $Q_W = \sigma_w^2 I$, and with $N = p$ in (14) for simplicity, $N \times N$ matrices A_x and B_x are given by (15) and $C_x = I_N$.

One can also express the colored noise model in (20) with $q = p$ in vector form similar to (38) as

$$\xi(n) = A_v \xi(n-1) + B_v \nu(n) \tag{40}$$

$$v(n) = C_v \xi(n) \tag{41}$$

where

$$\xi(n) \triangleq [v(nN)v(nN-1) \cdots v(nN-N+1)]^T,$$

$$\nu(n) \triangleq [\eta(nN)\eta(nN-1) \cdots \eta(nN-N+1)]^T$$

the N-vector $\nu(n)$ is a white Gaussian process with zero mean and covariance matrix $Q_\eta = \sigma_\eta^2 I$, and $N \times N$ matrices A_v and B_v are given by (15) with $\{a_i\}$ replaced by $\{b_i\}$, and $C_v = I$.

Augmenting the state and measurement vectors given in (38)–(41), we obtain

$$\underbrace{\begin{bmatrix} x(n) \\ \xi(n) \end{bmatrix}}_{\bar{x}(n)} = \underbrace{\begin{bmatrix} A_x & 0 \\ 0 & A_v \end{bmatrix}}_{\bar{A}} \underbrace{\begin{bmatrix} x(n-1) \\ \xi(n-1) \end{bmatrix}}_{\bar{x}(n-1)} + \underbrace{\begin{bmatrix} B_x & 0 \\ 0 & B_v \end{bmatrix}}_{\bar{B}} \underbrace{\begin{bmatrix} w(n) \\ \nu(n) \end{bmatrix}}_{\bar{w}(n)} \tag{42}$$

$$s(n) = \underbrace{[C_x \quad C_v]}_{\bar{C}} \underbrace{\begin{bmatrix} x(n) \\ \xi(n) \end{bmatrix}}_{\bar{x}(n)} \tag{43}$$

which is a linear system driven by a white Gaussian vector $\bar{w}(n)$ from which is available perfect measurements. Since $w(n)$ and $\nu(n)$ are uncorrelated by assumption, the $2N \times 2N$ covariance matrix is

$$Q \triangleq E[\bar{w}(n)\bar{w}^T(n)] = \begin{bmatrix} Q_W & 0 \\ 0 & Q_\eta \end{bmatrix} = \begin{bmatrix} \sigma_w^2 I & 0 \\ 0 & \sigma_\eta^2 I \end{bmatrix}. \tag{44}$$

As in the scalar case, we introduce a coordinate transformation in order to remove the singularity in the error covariance recursion and to reduce the dimension of the optimal filter. Let us define a $2N \times 2N$ transformation matrix T as

$$T = \begin{bmatrix} \bar{C} \\ T_2 \end{bmatrix}$$

and choose $T_2 = [I_N \quad 0_N]$ to yield

$$T = \begin{bmatrix} I_N & I_N \\ I_N & 0_N \end{bmatrix}, \quad \text{and } T^{-1} = \begin{bmatrix} 0_N & I_N \\ I_N & -I_N \end{bmatrix}.$$

By using this transformation, we have

$$\tilde{x}(n) = T\bar{x}(n) = \begin{bmatrix} s(n) \\ x(n) \end{bmatrix}, \quad \text{and } \bar{x}(n) = T^{-1}\tilde{x}(n). \tag{45}$$

Premultiplying (42) with T, we obtain

$$\tilde{x}(n) = \underbrace{T\bar{A}T^{-1}}_{\tilde{A}}\tilde{x}(n-1) + \underbrace{T\bar{B}}_{\tilde{B}}\bar{w}(n) \tag{46}$$

$$s(n) = \underbrace{\bar{C}T^{-1}}_{\tilde{C}}\tilde{x}(n) \tag{47}$$

where

$$\tilde{C} = [I_N \quad 0_N]. \tag{48}$$

The optimal Kalman filter estimate becomes

$$\hat{\tilde{x}}(n) = \hat{\tilde{x}}(n|n-1) + K(n)\{s(n) - \tilde{C}\hat{\tilde{x}}(n|n-1)\} \tag{49}$$

with $\hat{\tilde{x}}(n|n-1) = \tilde{A}\hat{\tilde{x}}(n-1)$, and the gain and error covariance equations evolve according to

$$K(n) = P(n|n-1)\tilde{C}^T[\tilde{C}P(n|n-1)\tilde{C}^T]^{-1} \tag{50}$$

$$P(n|n-1) = \tilde{A}P(n-1)\tilde{A}^T + \tilde{B}Q\tilde{B}^T \tag{51}$$

$$P(n) = [I - K(n)\tilde{C}]P(n|n-1) \tag{52}$$

where Q is given by (44).

As noted in the discussion of the scalar Kalman filter, the optimal filter order for the colored noise is $p + q -$ dim $s(n)$. In the scalar case, the computational saving attainable by filter order reduction is trivial because dim $s(n) = 1$, which is much smaller than $p + q$. However, it is significant in this case because we have $N = p = q$ and dim $s(n) = N$, yielding the filter order N.

We proceed to obtain the reduced order filter. If we denote the $2N \times 2N$ matrix $P(n|n-1)$ as four $N \times N$ submatrices

$$P(n|n-1) \triangleq \begin{bmatrix} P_{11}(n|n-1) & P_{12}(n|n-1) \\ P_{21}(n|n-1) & P_{22}(n|n-1) \end{bmatrix}$$

then with (48), the $2N \times N$ matrix $K(n)$ becomes

$$K(n) = \begin{bmatrix} P_{11}(n|n-1) \\ P_{21}(n|n-1) \end{bmatrix} P_{11}^{-1}(n|n-1) = \begin{bmatrix} I_N \\ P_{21}(n|n-1)P_{11}^{-1}(n|n-1) \end{bmatrix} \triangleq \begin{bmatrix} I_N \\ K_2(n) \end{bmatrix}. \tag{53}$$

By denoting the augmented state vector estimate as

$$\hat{\tilde{x}}(n) \triangleq \begin{bmatrix} \hat{s}(n) \\ \hat{x}(n) \end{bmatrix}$$

the augmented prediction vector can be written as

$$\begin{aligned} \hat{\tilde{x}}(n|n-1) &\triangleq \tilde{A}\hat{\tilde{x}}(n-1) \\ &= \begin{bmatrix} A_x\hat{x}(n-1) + A_v\{\hat{s}(n-1) - \hat{x}(n-1)\} \\ A_x\hat{x}(n-1) \end{bmatrix} \\ &\triangleq \begin{bmatrix} \hat{s}(n|n-1) \\ \hat{x}(n|n-1) \end{bmatrix}. \end{aligned} \tag{54}$$

Then, $\hat{\tilde{x}}(n)$ in (49) becomes, with $K(n)$ in (53)

$$\hat{\tilde{x}}(n) \triangleq \begin{bmatrix} \hat{s}(n) \\ \hat{x}(n) \end{bmatrix} = \begin{bmatrix} s(n) \\ \hat{x}(n|n-1) + K_2(n)[s(n) - \hat{s}(n|n-1)] \end{bmatrix}$$

from which we have

$$\hat{s}(n) = s(n) \tag{55}$$

and the reduced order vector estimate

$$\hat{x}(n) = \hat{x}(n|n-1) + K_2(n)[s(n) - \hat{s}(n|n-1)] \tag{56}$$

where $\hat{x}(n|n-1) = A_x\hat{x}(n-1)$, directly from (54), and $\hat{s}(n|n-1) = A_x\hat{x}(n-1) + A_v[s(n-1) - \hat{x}(n-1)]$ by (54) and (55).

The reduced order gain equation is by (53),

$$K_2(n) = P_{21}(n|n-1)P_{11}^{-1}(n|n-1). \tag{57}$$

Dimensions of the error covariance matrices in (51) and (52) cannot be reduced because all four submatrices of $P(n|n-1)$ are involved in computing $P(n)$ and vice versa. The reduced order Kalman filter algorithm for colored noise iterates as follows: With initial conditions $\hat{x}(0) = 0$ and $P(0) = 0$, (51) is processed first, followed by (57) and then (56) and (52).

IV. Experimental Results

The filtering algorithms in Section III require the knowledge of $a = [a_1 \cdots a_p]^T$ and σ_w^2, as well as the correlation properties of the noise which we assume are known or can be calculated from some reference signal. Estimation of source model parameters from noisy speech is a long-standing research problem with most efforts being focused on the white noise case. Among the solutions for the white noise case, optimal methods as in [20], [21] require the solution of a nonlinear optimization problem, while suboptimal methods based on modified or extended Yule-Walker equations [22]–[24] are known to yield unsatisfactory performance for wide-band signals such as speech.

The estimation method presented here is a suboptimal solution that can be considered as a version of the EM algorithm based on a maximum likelihood argument [37]. At each iteration, we alternately estimate the parameters and filter the speech, which is the same idea used by Lim and Oppenheim [4]. However, the results in [4] are only for Wiener filtering of white noise.

We now evaluate the performance of the proposed Kalman filtering algorithms for speech enhancement alone and for coding of noisy speech when the additive noise is white or colored. The objective distortion measure used is the signal-to-noise ratio (SNR) defined by

$$\mathrm{SNR} \triangleq 10 \log_{10} \frac{\frac{1}{L}\sum_{n=1}^{L} x^2(n)}{\frac{1}{L}\sum_{n=1}^{L} [x(n) - \hat{x}(n)]^2} \text{ dB} \tag{58}$$

where $x(n)$ is the noise-free sequence and $\hat{x}(n)$ is the appropriate output sequence, both of length L, where L is the total number of samples in the utterance. For speech enhancement alone, $\hat{x}(n)$ is the filter output, while for the cascade estimator/coder configuration, we define an SNR_E when $\hat{x}(n)$ is the filter output and an SNR_0 when $\hat{x}(n)$ is the output of the coder as in Fig. 1. The limitations of the SNR in (58) for speech coding applications are well known, and hence, we augment the SNR performance indicator with sound spectrogram analyses and informal subjective listening tests. We did attempt to utilize the Itakura log likelihood ratio statistic [1], [29] for performance evaluation, but the results were inconsistent. We have not determined the limitations of this distance measure, and since this is a research topic in itself, we leave it to [30], [31] and future work.

The noise sequences used in the experiments have zero mean and are scaled to yield a 5-dB input SNR as defined by $1/L \sum_{n=1}^{L} x^2(n)/\sigma_v^2$. The white Gaussian noise sequence is generated by the polar method [32] and the primary colored noise considered is helicopter noise. The noise sequences are assumed wide sense stationary over the entire utterance and their autocorrelations are computed based upon the noise sequences alone. This presumes the availability of a reference signal at some other location or in some time interval when speech is not present. The other parameters $\{a_i, i = 1, 2, \cdots, p\}$ and σ_w^2 needed in the Kalman filters are computed over block lengths of 120 samples using the autocorrelation method [28]. The speech data for this work is described in the Appendix.

Note that for all of the simulations reported here, L = length of the entire sentence being processed, $N = p$ = order of the AR model for the speech, and the frame length used for calculating the parameters $\{a_i, i = 1, 2, \cdots, p\}$ and σ_w^2 is 120 samples. The frame length of 120 samples corresponds to 15 ms and was selected arbitrarily from within the common parameter update range of 10 to 30 ms often used in linear predictive coding systems. The frame length is not critical to the filter performance within these bounds.

A. Speech Enhancement

To develop bounds on the performance of the Kalman filters for speech enhancement and to determine suitable model orders p and q for the speech and noise models in (1) and (20), we consider two extreme cases. One case is where we compute the Kalman filter parameters $\{a_i, \sigma_w^2\}$ directly from the noisy speech without any iteration, which we designate as the "noisy (0)." Thus, we filter the speech within each frame and then proceed to the next frame. Neither the signal estimates from the previous frame nor the covariance matrix at the end of a frame is used for initial conditions in the succeeding frame. The other extreme is where we calculate the Kalman filter parameters from the noise-free speech, and then process the noisy data through the Kalman filter with the ideal parameters, which we call the "ideal" case. The SNR values for white and colored noise distortions are presented in Tables I and II, respectively. Note that we employ the appropriate Kalman filter in each situation; that is, we use white noise assumption algorithms for white noise and colored noise assumption algorithms for colored noise. Tables I and II also allow the comparison of the scalar and vector algorithms described in Section III. For compactness in the following discussion, we denote "Kalman filter" by KF. From the tables it is evident that the vector KF produces a consistently larger SNR than its scalar counterpart at any order. Further, the SNR's of both the scalar and vector KF algorithms increase with increasing order, although the improvements for the noisy (0) cases are 1 dB or less. However, for the ideal case, increasing

TABLE I
KALMAN FILTER SNR FOR THE NOISY AND IDEAL CASES (WHITE NOISE)

Kalman Filter Type	p	Sentence 1 Noisy (0)	Sentence 1 Ideal	Sentence 4 Noisy (0)	Sentence 4 Ideal
Scalar	4	8.073	9.925	7.798	9.171
	8	8.188	10.039	7.910	9.408
	10	8.188	10.094	7.928	9.509
	20	8.228	10.561	7.899	9.667
Vector	4	8.576	11.039	8.187	9.956
	8	8.751	11.580	8.379	10.539
	10	8.770	11.750	8.397	10.725
	20	8.736	12.253	8.368	11.072

TABLE II
KALMAN FILTER SNR FOR THE NOISY AND IDEAL CASES (COLORED NOISE)

Kalman Filter Type	(p, q)	Sentence 1 Noisy (0)	Sentence 1 Ideal	Sentence 4 Noisy (0)	Sentence 4 Ideal
Scalar	(4, 4)	8.029	9.417	7.657	8.822
	(4, 10)	8.461	9.983	8.032	9.227
	(10, 10)	8.240	10.153	8.079	9.562
	(14, 14)	8.480	10.490	8.187	9.662
	(20, 20)	8.605	10.976	8.253	9.875
Vector	(4, 4)	8.078	9.535	7.698	8.923
	(8, 8)	8.367	10.014	8.208	9.415
	(10, 10)	8.490	10.805	8.272	9.974
	(15, 15)	8.825	11.646	8.476	10.439
	(20, 20)	9.061	12.237	8.571	10.622

the filter orders improves the SNR by as much as 2.7 dB. We note here that sentence 1 was chosen because its correlation and spectral properties are somewhat representative of a majority of speech utterances, while sentence 4 was chosen because it has different correlation properties that vary rapidly during the utterance. Subjective listening tests generally track the SNR values, but differences of less than 0.5 dB are not audible.

To choose a model order, we note that for colored noise with $(p, q) = (4, 4)$, remnants of the colored noise spectrum for both the scalar and vector filters are audible, indicating the need to increase the order q of the AR noise model. The orders (4, 10) for the scalar KF and (10, 10) for the vector KF are the minimal values for effective filter performance.

Next we investigate the performance of the iterative parameter estimation method by alternately calculating the parameters $\{a_i, i = 1, \cdots, p, \sigma_w^2\}$ and filtering, and then iterating on this process. The noisy (0) case consists of calculating the parameters from the noisy speech and then using them in KF algorithms. Iteration 1 computes the parameters from the filtered output of iteration noisy (0) and substitutes these values into the KF algorithms for filtering. Subsequent iterations proceed similarly. The SNR values for white and colored noises at selected orders are presented in Tables III and IV, respectively. It is evident that iterating can improve the SNR, but that iterating beyond a point (iterations 2 or 3) can actually reduce the SNR. Informal subjective listening tests confirm the SNR values, with the speech quality at iteration 2 substantially improved over iterations 0 and 1, and in fact, relatively close to the ideal case. The increase in distortion with repeated iterations beyond a certain point also has been observed by Lim and Oppenheim [4].

TABLE III
KALMAN FILTER SNR FOR ITERATIVE PARAMETER ESTIMATION (WHITE NOISE)

Kalman Filter Type	Iteration	Sentence 1 p = 4	Sentence 1 p = 10	Sentence 4 p = 4	Sentence 4 p = 10
Scalar	Noisy (0)	8.073	8.188	7.798	7.928
	1	9.237	9.400	8.701	8.922
	2	9.531	9.735	8.867	9.126
	3	9.551	9.777	8.827	9.090
	4	9.532	9.767	8.772	9.026
	5	9.504	9.753	8.721	8.973
	Ideal	9.925	10.094	9.171	9.509
Vector	Noisy (0)	8.576	8.770	8.187	8.397
	1	10.252	10.655	9.423	9.849
	2	10.728	11.278	9.675	10.228
	3	10.771	11.383	9.628	10.223
	4	10.754	11.369	9.562	10.162
	5	10.737	11.342	9.509	10.111
	Ideal	11.039	11.750	9.956	10.725

TABLE IV
KALMAN FILTER SNR FOR ITERATIVE PARAMETER ESTIMATION (COLORED NOISE)

Kalman Filter Type	Iteration	Sentence 1 (p, q) = (4, 10)	Sentence 1 (p, q) = (10, 10)	Sentence 4 (p, q) = (4, 10)	Sentence 4 (p, q) = (10, 10)
Scalar	Noisy (0)	8.461	8.240	8.032	8.079
	1	9.541	9.409	8.833	9.008
	2	9.736	9.713	8.948	9.165
	3	9.728	9.753	8.817	9.115
	4	9.697	9.749	8.777	9.047
	5	9.671	9.738	8.737	8.978
	Ideal	9.983	10.153	9.227	9.562
Vector	Noisy (0)	8.078	8.490	7.698	8.272
	1	8.986	9.877	8.409	9.321
	2	9.092	10.249	8.399	9.492
	3	8.990	10.287	8.197	9.426
	4	8.888	10.270	8.057	9.336
	5	8.805	10.255	7.958	9.274
	Ideal	9.535	10.805	8.923	9.974

It is of interest to compare the computational requirements of the scalar KF algorithms to the vector KF algorithms. The general conclusion is that the scalar algorithms run much faster than their vector counterparts, with exactly how much faster depending upon the specific implementation. The reasons for this difference are evident from an examination of the algorithms in Section III. The relative simplicity of the scalar colored noise KF algorithms in Section III-C primarily accrues from the facts that $\tilde{\boldsymbol{h}}$ has only one nonzero component and $\bar{\boldsymbol{h}}$ has only two nonzero components. Since $\bar{\boldsymbol{h}}$ is part of the transformation $\boldsymbol{T}$ which is used to compute $\tilde{\boldsymbol{F}}$ and $\tilde{\boldsymbol{G}}$, these latter matrices are somewhat sparse. Perhaps more important is the effect of $\tilde{\boldsymbol{h}}$ on (32), (33), and (35)—especially (33). Note that because of the special form of $\tilde{\boldsymbol{h}}$, the inversion in (33) only involves a scalar. This can be contrasted to the vector KF algorithms that require the inversion of an $N \times N$ matrix in (57), which more than offsets any advantage in processing vectors of N samples at a time.

Table V establishes the difference in performance between using a white assumption KF algorithm and a colored noise assumption KF algorithm on three different colored noises. This comparison is important since the colored noise algorithms are not as straightforward nor as widely known as the white noise KF algorithms, and it is tempting to simply use the white noise algorithms. To minimize the number of variables in the comparisons, all results shown are for the ideal case, scalar filters, and $p = 10$ or $(p, q) = (4, 10)$. The AR(4) noise is taken from [33]. In all three instances, it is clear that the colored noise algorithms provide a significant gain in SNR over the white noise algorithms. Sound spectrogram analyses and listening tests again support the SNR results, although the audible difference between KF outputs A and B is not as dramatic for jeep noise as it is for the AR(4) noise and the helicopter noise. The value of the colored noise algorithms is thus clearly established.

When the noise model structure and parameters are known, the method should be equally effective on all noises. However, if the noise model structure is not accurate or the model parameters must be estimated, the performance may be degraded. For example, the AR(10) noise model is apparently not quite as accurate for the jeep noise as it is for the AR(4) and helicopter noises, and hence the filtering performance of the algorithms is not quite as good subjectively for the jeep noise as it is for the other two. Further, if the noise model parameters must be iteratively estimated, as in Tables III and IV, the method will be more effective against "nonspeech-like" noises such as narrow-band noise.

TABLE V
PERFORMANCE COMPARISON OF KALMAN FILTERS. FILTER TYPE A: STANDARD WHITE NOISE ASSUMPTION KALMAN FILTER. FILTER TYPE B: COLORED NOISE ASSUMPTION KALMAN FILTER

	Noise Type	SNR (dB)
KF Input Signal	AR(4) Noise	5.004
KF Output—A	AR(4) Noise	7.150
KF Output—B	AR(4) Noise	9.900
KF Input Signal	Jeep Noise	5.002
KF Output—A	Jeep Noise	8.342
KF Output—B	Jeep Noise	10.317
KF Input Signal	Helicopter Noise	5.004
KF Output—A	Helicopter Noise	8.343
KF Output—B	Helicopter Noise	9.983

TABLE VI
LPC INPUTS FOR THE TANDEM KF/LPC SYSTEM (COLORED NOISE)

LPC Input		SNR_E (dB)
Noise-free Sentence 6		∞
Sentence 6 plus Helicopter Noise		5.002
KF Output Scalar (4, 10)	Noisy (0)	8.401
	Iteration 2	9.834
	Ideal	10.180

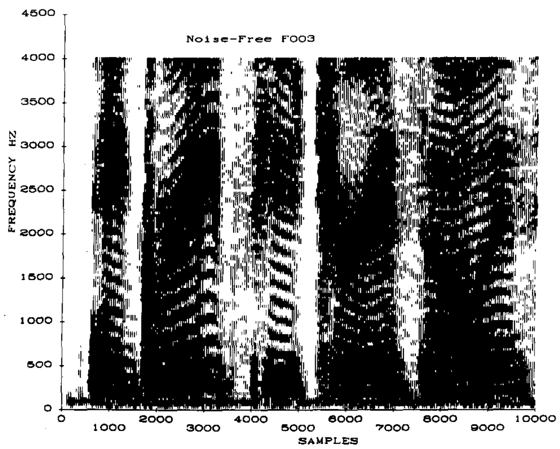


Fig. 2. Narrow-band spectrogram for original sentence 6.

B. LPC of Noisy Speech

Additive noise presents a particularly difficult problem for linear predictive coding (LPC) systems operating at 4.8 kb/s and below. In this section, we study the performance of a cascade estimator/LPC system for coding noisy speech at 4.8 kb/s. The LPC system is 14th order, and the LPC synthesis is accomplished with accurate pitch and voicing calculated from noise-free speech, so that any distortion in the LPC output is due to errors in estimating the LPC coefficients. The autocorrelation method is used for coefficient calculation [28]. Sentence 6, taken from Texas Instrument's pitch data base and described in the Appendix, is the speech segment in this study.

The LPC inputs are listed in Table VI, along with the SNR_E at the KF output. For simplicity, a scalar, colored noise KF algorithm with orders (4, 10) is implemented here. Figs. 2 and 3 are narrow-band (45 Hz) spectrograms of the noise-free, unprocessed speech and the speech plus helicopter noise, respectively. Figs. 4–6 are narrow-band spectrograms of the LPC output for the unfiltered noisy speech, the noisy (0) case KF output, and the iteration 2

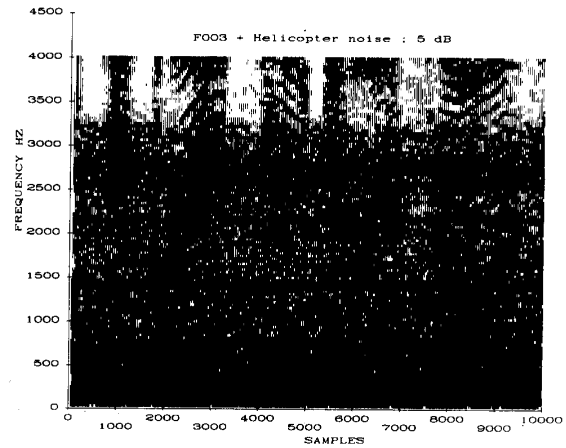


Fig. 3. Sentence 6 plus helicopter noise (SNR = 5.002 dB).

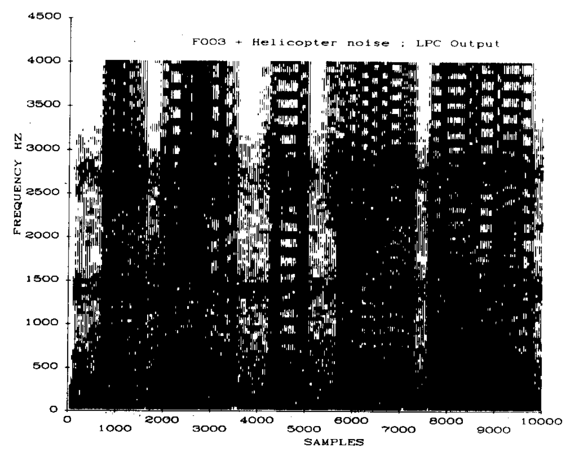


Fig. 4. LPC output for speech plus helicopter noise.

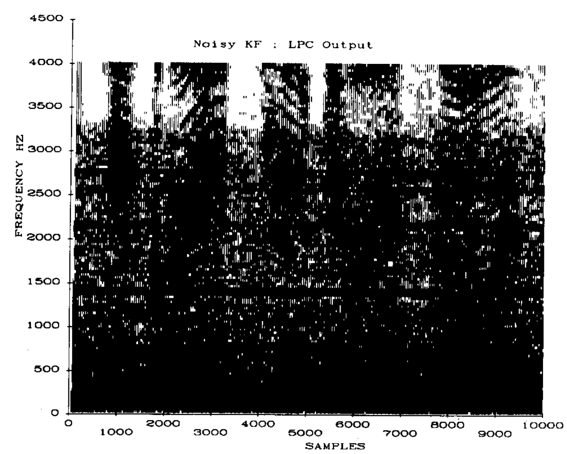


Fig. 5. LPC output for noisy (0) KF.

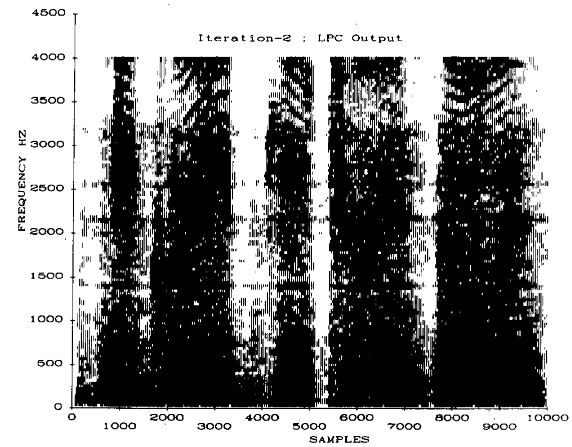


Fig. 6. LPC output for iteration 2 KF.

KF output, respectively. Narrow-band spectrograms are presented to emphasize the tones in the helicopter noise. A substantial improvement is evident in Fig. 6 compared to Figs. 4 and 5. Subjectively, the LPC output for the unfiltered noisy speech input is barely intelligible and very noisy. The intelligibility is improved for the noisy (0) filtering case, and the LPC output for the iteration 2 KF input has good intelligibility and much less noise, being only slightly noisier than the ideal case. The cascade estimator/LPC structure is clearly effective for improving LPC coefficient estimates.

V. Conclusions

Scalar and vector output, white and colored noise assumption Kalman filter algorithms have been developed for speech enhancement and for use in a cascade estimator/speech coder structure to code noisy speech. The Kalman algorithms provide increases in SNR and improved speech quality and intelligibility in all cases, with the vector Kalman algorithms yielding the greatest improvement. In the cascade estimator/coder configurations with LPC, the Kalman filters improve the output speech and change the LPC output from unusable to clearly intelligible. It is also demonstrated that white noise assumption Kalman algorithms are ineffective in colored noise environments, and that the lesser known colored noise Kalman algorithms are an effective tool for the enhancement and coding of noisy speech.

Appendix
Speech Data

The three sentences used in this work are as follows.

Sentence 1: "The pipe began to rust while new." (Female speaker.)

Sentence 4: "Thieves who rob friends deserve jail." (Male speaker.)

Sentence 6: "A great future is always provided the student of music." (Female speaker.)

All sentences are sampled 8000 times/s.

ACKNOWLEDGMENT

The authors appreciate the use of the Texas Instruments pitch data base provided by Dr. J. Picone, from which sentence 6 was taken.

REFERENCES

[1] M. R. Sambur and N. S. Jayant, "LPC analysis/synthesis from speech inputs containing quantization noise or additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 488-494, Dec. 1976.

[2] C. F. Teacher and D. Coulter, "Performance of LPC vocoders in a noisy environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Washington, DC), Apr. 2-4, 1979, pp. 216-219.

[3] J. D. Gibson, "Theory of LPC analysis for distorted inputs," in *Proc. Eighteenth Annu. Allerton Conf. Commun., Contr., Computing*, (Monticello, IL), Oct. 8-10, 1980, p. 955.

[4] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 197-210, June 1978.

[5] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering, " in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Dallas, TX), Apr. 6-9, 1987, pp. 177-180.

[6] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 406-411, July 1970.

[7] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 518-521, Sept. 1980.

[8] J. D. Gibson and T. R. Fischer, "Alphabet-constrained data compression," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 443-457, May 1982.

[9] K. Kroschel, "A comparison of quantizers optimized for corrupted and uncorrupted input signals," *Signal Processing*, vol. 12, pp. 169-176, 1987.

[10] E. Ayanoglu, "Optimal quantization of noisy sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (New York, NY), Apr. 11-14, 1988, pp. 565-568.

[11] J. D. Gibson, T. R. Fischer, and B. Koo, "Estimation and vector quantization of noisy speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (New York, NY), Apr. 11-14, 1988, pp. 541-544.

[12] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory*, vol. 34, pp. 826-834, July 1988.

[13] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Boston, MA), Apr. 14-16, 1983, pp. 235-238.

[14] A. Dembo, "Signal reconstruction from noisy partial information of its transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 65-72, Jan. 1989.

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

[16] J. H. L. Hansen and M. A. Clements, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Trans. Signal Processing*, vol. 39, no. 4, pp. 795-805, Apr. 1991.

[17] A. P. Sage and J. L. Melsa, *Estimation Theory with Applications to Communications and Control*. New York: McGraw-Hill, 1971.

[18] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. 1. New York: Academic, 1979.

[19] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[20] M. Pagano, "Estimation of models of autoregressive signals plus white noise," *Ann. Stat.*, vol. 2, pp. 99-108, 1974.

[21] B. Friedlander and B. Porat, "A spectral matching technique for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 338-343, Apr. 1984.

[22] W. J. Done and C.K. Rushforth, "Estimating the parameters of a noisy all-pole process using pole-zero modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Washington, DC), Apr. 2-4, 1979, pp. 228-231.

[23] K. Bry and J. LeRoux, "Comparison of some algorithms for identifying autoregressive signals in the presence of noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Paris, France), May 3-5, 1982, pp. 224-227.

[24] K. K. Paliwal, "A noise-compensated long correlation matching method for AR spectral estimation of noisy signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (Tokyo, Japan), Apr. 8-11, 1986, pp. 1369-1372.

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

[26] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1965.

[27] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, pp. 95-103, 1983.

[28] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[29] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67-72, Feb. 1975.

[30] P. de Souza and P. J. Thomson, "LPC distance measures and statistical tests with particular reference to the likelihood ratio," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 304-315, Apr. 1982.

[31] B.-H. Juang, "On using the Itakura-Saito measures for speech coder performance evaluation," *AT&T Bell Lab. Tech. J.*, vol. 63, pp. 1477-1498, Oct. 1984.

[32] D. E. Knuth, *The Art of Computer Programming*, vol. 2, 2nd ed. Reading, MA: Addison-Wesley, 1981.

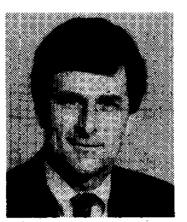
[33] B. Friedlander and B. Porat, "A spectral matching technique for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 338-343, Apr. 1984.

[34] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84-95, 1980.

[35] B. R. Musicus, "An iterative technique for maximum likelihood estimation with noisy data," M. S. thesis, Dep. Elec. Eng. Comput. Sci., M.I.T., Cambridge, MA, Feb. 20, 1979.

[36] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," Ph.D. dissertation, Dep. Elec. Eng., M.I.T., Cambridge, MA, Aug. 1982.

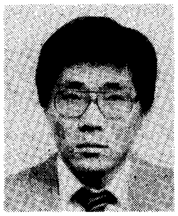
[37] B. Koo, "Structures for coding noisy speech at medium-to-low bit rates," Ph.D. dissertation, Dep. Elec. Eng., Texas A&M Univ., Dec. 1988.



Jerry D. Gibson (S'73-M'73-SM'83) was born in Fort Worth, TX, on May 12, 1946. He received the B.S. degree in electrical engineering from the University of Texas at Austin in 1969 and the M.S. and Ph.D. degrees from Southern Methodist University, Dallas, TX, in 1971 and 1973, respectively.

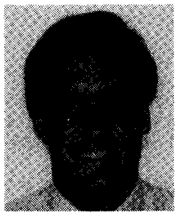
He has held positions at General Dynamics, Fort Worth, TX (1969 to 1972), the University of Notre Dame (1973 to 1974), the Defense Communications Agency (Summer 1974), and the University of Nebraska—Lincoln (1974 to 1976). In 1976 he joined Texas A&M University where he currently holds the J. W. Runyon, Jr., Professorship in the Department of Electrical Engineering. He is coauthor of the book *Introduction to Nonparametric Detection with Applications* (Academic Press, 1975) and was Associate Editor for Speech Processing for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1981 to 1985. He wrote the textbook, *Principles of Digital and Analog Communications* (Macmillan, 1989), and a chapter (with K. Sayood) entitled "Lattice Quantization" in *Advances in Electronics and Electron Physics* (Academic Press, 1988). He is currently Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY and is a member of the IEEE Information Theory Society Board of Governors (1990 to 1992). He is General Cochairman of the 1993 International Symposium on Information Theory to be held in San Antonio, TX. His research interests include speech processing, data compression, digital communications, and image processing.

Dr. Gibson received the 1990 Frederick Emmons Terman Award from the American Society for Engineering Education. He is a member of Eta Kappa Nu and Sigma Xi.



Boneung Koo was born in Seoul, Korea, in 1953. He received the B.S. degree in 1975 from the Seoul National University, Seoul, Korea, and the M.S. and Ph.D. degrees in 1984 and 1988, respectively, from Texas A&M University, College Station, TX, all in electrical engineering.

From April 1975 to May 1976, he served in the Korean Air Force, Seoul, Korea. From June 1976 to December 1976, he was with the Gold-Star Electric Co., Osan, Korea, where he worked on the development of communication equipments. From 1977 to 1982, he was with the Korea Advanced Energy Research Institute, Seoul, Korea. From 1983 to 1988, he was a Teaching Assistant or a Research Assistant in the Department of Electrical Engineering, Texas A&M University. He is currently an Assistant Professor in the Department of Electronic Engineering at Kyonggi University, Suwon, South Korea. His research interests include noise filtering, speech analysis/synthesis and coding, data compression, digital signal processing, and digital communication systems, in general.



Steven D. Gray was born in Kenosha, WI, on April 3, 1962. He received the B.S.E.E. and M.S.E.E. degrees from Texas A&M University, College Station, TX, in 1985 and 1986, respectively.

He started employment at Sandia National Laboratories in Albuquerque, NM, in 1986. He then joined E-Systems Inc., Garland, TX, and is presently with the MITRE Corporation, Bedford, MA. His main research interests are in estimation theory and adaptive filtering.