

# Subband Kalman Filtering for Speech Enhancement

Wen-Rong Wu, *Member, IEEE*, and Po-Cheng Chen

**Abstract**—Kalman filtering is an effective speech-enhancement technique, in which speech signals are usually modeled as autoregressive (AR) processes and represented in the state-space domain. Since AR coefficients identification and Kalman filtering require extensive computations, real-time implementation of this approach is difficult. This paper proposes a simple and practical scheme that overcomes these obstacles. Speech signals are first decomposed into subbands. Subband speech signals are then modeled as low-order AR processes, such that low-order Kalman filters can be applied. Enhanced fullband speech signals are finally obtained by combining the enhanced subband speech signals. To identify AR coefficients, prediction-error filters adapted by the LMS algorithm are applied. Due to noisy inputs, the LMS algorithm converges to biased solutions. The performance of the Kalman filter with biased parameters is analyzed. It is shown that accurate estimates of AR coefficients are not required when the driving-noise variance is properly estimated. New methods for making such estimates are proposed. Thus, we can tolerate biased AR coefficients and take advantage of the LMS algorithm's simple structure. Simulation results show that speech enhancement in the subband domain not only greatly reduces the computational complexity, but also achieves better performance compared to that in the fullband domain.

**Index Terms**—AR modeling, Kalman filtering, LMS algorithm, speech enhancement, subband filtering.

## I. INTRODUCTION

SPEECH enhancement has many applications in voice communication, speech recognition, and hearing aids. Speech enhancement often aims to reduce noise levels, increase intelligibility, or reduce auditory fatigue. Many studies have been done using techniques such as short-time spectral amplitude estimation [1]–[4], iterative Wiener filtering [5]–[7], audio-based filtering [8], [9], signal-subspace processing [10], [11], and hidden Markov modeling (HMM) [12], [13]. Although significant results have been achieved, most of them are not suitable for real-time implementation because their computational complexities are generally too high.

The Kalman filter is well known in signal processing for its efficient structure. In [14], Paliwal and Basu used a Kalman filter to enhance speech corrupted by white noise. On a short-time base, speech signals were modeled as stationary AR processes and AR parameters were assumed to be known. Gibson, Koo, and Gray considered speech enhancement with colored noise in [15]. They modeled both speech and colored noise as AR processes and developed scalar and vector

Kalman filtering algorithms. To estimate the AR coefficients, an EM-based algorithm was employed. In [16], Lees and Ann proposed a non-Gaussian AR model for speech signals. They modeled the distribution of the driving-noise as a Gaussian mixture and applied a decision-directed nonlinear Kalman filter. Again, an EM-based algorithm was used to identify unknown parameters. Niedźwiecki and Cisowki [17] assumed that speech signals are nonstationary AR processes and used a random-walk model for the AR coefficients. An extended Kalman filter was then used to simultaneously estimate speech and AR coefficients. Note that the stability of the extended Kalman filter is not guaranteed and dimensions of the Kalman filter are greatly increased.

The aforementioned Kalman filtering algorithms still require extensive computations for two reasons: first, to identify AR coefficients using EM algorithms, and second, to carry out filtering using Kalman filters. To overcome these drawbacks, we suggest modeling and filtering speech signals in the subband domain. Since the power spectral densities (PSD's) of subband speech signals are simpler and flatter than their fullband signals, low-order AR models are sufficient, and only lower-order Kalman filters will be required. Specifically, we focus on first- and zeroth-order modeling. In this case, the Kalman filter involves only scalar operations, thus saving a considerable amount of computation. For identification of AR coefficients, we use a prediction-error filter adapted by the LMS algorithm. The LMS algorithm is well known for its simplicity and robustness, however, its slow convergence precludes its use in many practical applications. Since the PSD's of subband speech signals are relatively flat and there is at most one AR coefficient, the LMS algorithm will thus converge more rapidly.

The other parameter required by the Kalman filter is AR model's driving-noise variance. Through extensive simulations, we found that this variance plays a crucial role in the filtering process and propose effective methods for identifying it. Note that the input to the prediction-error filter is noisy, which means the LMS algorithm will converge to biased solutions. We analyzed the performance of the Kalman filter with biased AR coefficients and were able to show that using our variance identification methods, accurate estimation of AR coefficients is not required. Our driving-noise variance estimates not only have the advantage of simplicity, but can also compensate for performance degradation due to biased AR coefficients. Thus, we can take advantage of the LMS algorithm's simple structure, which facilitates real-time implementation.

This paper is organized as follows. Section II states the formulation of speech enhancement using the Kalman filter. In

Manuscript received December 9, 1996; revised November 4, 1997. This work was supported by the National Science Council, Taiwan, R.O.C., under Contract NSC 87-2213-E-009-120.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsin-Chu, 30050 Taiwan, R.O.C. (e-mail: wrwu@cc.nctu.edu.tw).

Publisher Item Identifier S 1057-7130(98)04674-6.

Section III, we describe how the Kalman filter can be applied in the subband domain. Section IV gives a performance analysis of the Kalman filter with biased parameters. Experimental results are reported in Section V and conclusions are drawn in Section VI.

## II. CONVENTIONAL KALMAN FILTERING

### A. White Noise Filtering

On a short-time basis, a speech sequence  $\{x(n)\}$  can be represented as an AR process, which is essentially the output of an all-pole linear system driven by a white noise sequence:

$$x(n) = \sum_{i=1}^p a_i x(n-i) + w(n) \quad (1)$$

where  $w(n)$  is a zero-mean white Gaussian process with variance  $\sigma_w^2$ . The observed speech signal  $s(n)$  is assumed to be contaminated by a zero-mean additive Gaussian noise  $v(n)$ , i.e.,  $s(n) = x(n) + v(n)$ . Let  $v(n)$  be white and  $\mathbf{x}(n) \triangleq [x(n) \ x(n-1) \ \dots \ x(n-p+1)]^T$ . Equation (1) and the corrupted speech  $s(n)$  can be reformulated in the state-space domain as

$$\mathbf{x}(n) = \mathbf{F}\mathbf{x}(n-1) + \mathbf{g}w(n) \quad (2)$$

$$s(n) = \mathbf{h}^T \mathbf{x}(n) + v(n) \quad (3)$$

$$\mathbf{F} = \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}_{p \times p}$$

$$\mathbf{g} = \mathbf{h} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}_{p \times 1}.$$

If we assume that parameters  $\mathbf{a} \triangleq [a_1 \ a_2 \ \dots \ a_p]^T$ ,  $\sigma_w^2$ , and  $\sigma_v^2$  are precisely known, the optimal estimate of  $\mathbf{x}(n)$  can be obtained from the Kalman filter. The Kalman equations for (2) and (3) are given as follows [15]:

$$\hat{\mathbf{x}}(n) = \mathbf{F}\hat{\mathbf{x}}(n-1) + \mathbf{k}(n)[s(n) - \mathbf{h}^T \mathbf{F}\hat{\mathbf{x}}(n-1)] \quad (5)$$

$$\mathbf{k}(n) = \mathbf{P}(n|n-1)\mathbf{h}[\mathbf{R} + \mathbf{h}^T \mathbf{P}(n|n-1)\mathbf{h}]^{-1} \quad (6)$$

$$\mathbf{P}(n|n-1) = \mathbf{F}\mathbf{P}(n-1)\mathbf{F}^T + \mathbf{g}\mathbf{Q}\mathbf{g}^T \quad (7)$$

$$\mathbf{P}(n) = [\mathbf{I} - \mathbf{k}(n)\mathbf{h}^T]\mathbf{P}(n|n-1) \quad (8)$$

where  $\hat{\mathbf{x}}(n)$  is the estimate of  $\mathbf{x}(n)$ ,  $\mathbf{k}(n)$  is the Kalman gain,  $\mathbf{P}(n|n-1) = E\{[\mathbf{x}(n) - \mathbf{F}\hat{\mathbf{x}}(n-1)][\mathbf{x}(n) - \mathbf{F}\hat{\mathbf{x}}(n-1)]^T\}$  is the prediction-error covariance matrix,  $\mathbf{P}(n) = E\{[\mathbf{x}(n) - \hat{\mathbf{x}}(n)][\mathbf{x}(n) - \hat{\mathbf{x}}(n)]^T\}$  is the filtering-error covariance matrix,  $\mathbf{R} = \sigma_v^2$  is the measurement-noise variance, and  $\mathbf{Q} = \sigma_w^2$  is the driving-noise variance. A speech sample estimate at time instant  $n$  can then be obtained using  $\hat{x}(n) = \mathbf{h}^T \hat{\mathbf{x}}(n)$ .

### B. Colored Noise Filtering

We assume that colored noise is stationary and can also be described by a  $q$ -order AR model

$$v(n) = \sum_{i=1}^q b_i v(n-i) + \eta(n) \quad (9)$$

where  $\{\eta(n)\}$  is a zero-mean white Gaussian process with variance  $\sigma_\eta^2$ . The AR parameters  $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_q]^T$  and  $\sigma_\eta^2$  can be estimated during nonspeech intervals and are assumed to be known. There are two types of formulation. One is called state augmentation, and the other is measurement difference. State augmentation expresses (9) as a state-space representation and incorporates that into the state equations (2) and (3). The state-space representation of  $v(n)$  is similar to that in (1). Let  $\mathbf{v}(n) = [v(n) \ v(n-1) \ \dots \ v(n-q+1)]^T$ . Then,

$$\mathbf{v}(n) = \mathbf{F}_v \mathbf{v}(n-1) + \mathbf{g}_v \eta(n) \quad (10)$$

$$v(n) = \mathbf{h}_v^T \mathbf{v}(n) \quad (11)$$

where  $\mathbf{F}_v$ ,  $\mathbf{g}_v$ , and  $\mathbf{h}_v$  are identical to those in (4), except that  $a$ ,  $p$  are replaced by  $b$ ,  $q$ . Combining (10), (11), (2), and (3), we then have

$$\bar{\mathbf{x}}(n) = \bar{\mathbf{F}}\bar{\mathbf{x}}(n-1) + \bar{\mathbf{G}}\bar{\mathbf{w}}(n) \quad (12)$$

$$s(n) = \bar{\mathbf{h}}^T \bar{\mathbf{x}}(n) \quad (13)$$

where

$$\bar{\mathbf{x}}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{v}(n) \end{bmatrix}, \quad \bar{\mathbf{w}}(n) = \begin{bmatrix} w(n) \\ \eta(n) \end{bmatrix}, \quad \bar{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_v \end{bmatrix},$$

$$\bar{\mathbf{G}} = \begin{bmatrix} \mathbf{g} & \mathbf{0} \\ \mathbf{0} & \mathbf{g}_v \end{bmatrix}, \quad \bar{\mathbf{h}}^T = [\mathbf{h}^T \ \mathbf{h}_v^T]. \quad (14)$$

The covariance matrix of  $\bar{\mathbf{w}}(n)$  is defined as  $\mathbf{Q} \triangleq E[\bar{\mathbf{w}}(n)\bar{\mathbf{w}}^T(n)] = \text{diag}(\sigma_w^2, \sigma_\eta^2)$ . The Kalman equations for (12) and (13) are then obtained by setting  $\mathbf{R} = 0$  and replacing  $\hat{\mathbf{x}}(n)$ ,  $\mathbf{F}$ ,  $\mathbf{h}$ , and  $\mathbf{g}$  with  $\hat{\bar{\mathbf{x}}}(n)$ ,  $\bar{\mathbf{F}}$ ,  $\bar{\mathbf{h}}$ , and  $\bar{\mathbf{G}}$  in (5)–(8). The speech estimate is then  $\hat{x}(n) = [\mathbf{h}^T \ \mathbf{0}] \hat{\bar{\mathbf{x}}}(n)$ . Note that  $\hat{\bar{\mathbf{x}}}(n)$  is of dimension  $(p+q) \times 1$ , the computational complexity of the Kalman filter is increased when this approach is used.

The idea behind measurement difference is to perform some measurement transformations such that measurement noise becomes white. Assume  $p \geq q$ , let  $\mathbf{s}(n) = [s(n) \ s(n-1) \ \dots \ s(n-q+1)]^T$ , and  $\mathbf{F}_s = [I_{q \times q} \ \mathbf{0}_{q \times (p-q)}]^T$ .

The transformed measurement is defined as

$$\bar{s}(n) = s(n+1) - \mathbf{g}_v^T \mathbf{F}_v \mathbf{s}(n)$$

$$= \mathbf{g}_v^T (\mathbf{F}_s^T \mathbf{F} - \mathbf{F}_v \mathbf{F}_s^T) \mathbf{x}(n) + w(n+1) + \eta(n+1)$$

$$= \mathbf{H}^T \mathbf{x}(n) + e(n) \quad (15)$$

where  $\mathbf{H}^T = \mathbf{g}_v^T (\mathbf{F}_s^T \mathbf{F} - \mathbf{F}_v \mathbf{F}_s^T)$  and  $e(n) = w(n+1) + \eta(n+1)$  is white Gaussian noise with a variance of  $\sigma_e^2 = \sigma_w^2 + \sigma_\eta^2$ . Now, (2) and (15) become the new state equations. Although measurement noise has been made white, it is correlated with driving noise. Fortunately, an optimal filter is available for

such a situation. The Kalman equations for the measurement-difference sequence  $\{\bar{s}(n)\}$  are given by [26]

$$\tilde{\mathbf{x}}(n|n-1) = (\mathbf{F} - \mathbf{g}\sigma_w^2\sigma_e^{-2}\mathbf{H}^T)\tilde{\mathbf{x}}(n-1) + \mathbf{g}\sigma_w^2\sigma_e^{-2}\bar{s}(n-1) \quad (16)$$

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{x}}(n|n-1) + \mathbf{k}(n)[\bar{s}(n) - \mathbf{H}^T\tilde{\mathbf{x}}(n|n-1)] \quad (17)$$

$$\mathbf{k}(n) = \mathbf{P}(n|n-1)\mathbf{H}[\mathbf{H}^T\mathbf{P}(n|n-1)\mathbf{H} + \sigma_w^2\sigma_e^{-2}]^{-1} \quad (18)$$

$$\mathbf{P}(n|n-1) = (\mathbf{F} - \mathbf{g}\sigma_w^2\sigma_e^{-2}\mathbf{H}^T)\mathbf{P}(n-1) \cdot (\mathbf{F} - \mathbf{g}\sigma_w^2\sigma_e^{-2}\mathbf{H}^T)^T + \mathbf{g}\sigma_w^2(1 - \sigma_w^2\sigma_e^{-2})\mathbf{g}^T \quad (19)$$

$$\mathbf{P}(n) = [\mathbf{I} - \mathbf{k}(n)\mathbf{H}^T]\mathbf{P}(n|n-1) \quad (20)$$

where  $\tilde{\mathbf{x}}(n|n-1)$  is the optimal prediction  $\mathbf{x}(n)$  based on  $\{\bar{s}(1), \bar{s}(2), \dots, \bar{s}(n-1)\}$ , and  $\tilde{\mathbf{x}}(n)$  is the optimal estimate based on  $\{\bar{s}(1), \bar{s}(2), \dots, \bar{s}(n)\}$ . Note that we use  $\tilde{\mathbf{x}}(n)$  instead of  $\hat{\mathbf{x}}(n)$  to denote the state estimate in the measurement-difference method. As we can see, the measurement-difference-type Kalman filter is somewhat more complex than the state-augmentation type. However, the main advantage of the measurement-difference approach is that state dimensions do not increase. If the AR order of the colored noise is high, this method can save considerable computation. The speech estimate is the same as that for white noise, i.e.,  $\tilde{\mathbf{x}}(n) = \mathbf{h}^T\tilde{\mathbf{x}}(n)$ . Note that

$$\begin{aligned} \tilde{x}(n) &= E\{x(n)|\bar{s}(1), \bar{s}(2), \dots, \bar{s}(n)\} \\ &= E\{x(n)|s(1), s(2), \dots, s(n+1)\} \end{aligned} \quad (21)$$

where  $E\{\cdot\}$  denotes the expectation operation. Thus,  $\tilde{x}(n)$  corresponds to a smoothing rather than a filtering result.

The AR coefficients and driving-noise variance must be estimated to apply the Kalman filter to speech enhancement. Many algorithms for performing this task have been presented, however, most of them require extensive computations and are not suitable for real-time implementation. In this paper, we focus on an adaptive prediction-error filter using an LMS-type algorithm. It is known that the convergence rate of the LMS algorithm is slow when the input correlation matrix has a large eigen-spread [25]. In addition to the convergence problem, direct application of the algorithm to noisy signals gives biased AR parameters. These may explain why the LMS algorithm is rarely used in speech enhancement. In what follows, we present new approaches that overcome these drawbacks.

### III. SUBBAND KALMAN FILTERING

#### A. Formulation

Our approach is motivated by the idea of curve-fitting using spline functions, which allows arbitrary curves to be fitted using polynomial functions. For better results, we usually need high-order polynomials. However, it is known that polynomials are inflexible; making them behave a certain way in one place may cause them to misbehave elsewhere. A more flexible approach is to use low-order piecewise polynomials. The curve is first divided into consecutive segments and each segment is fitted using a low-order polynomial. Constraints on segment

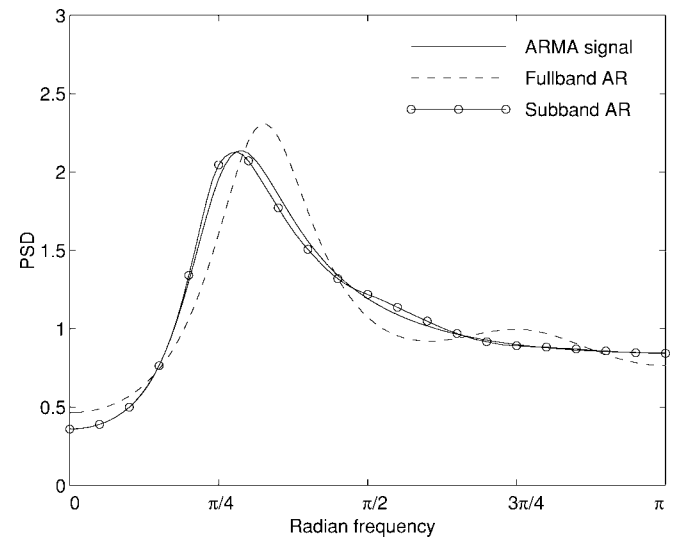


Fig. 1. PSD's of an ARMA and its AR-modeled signals; AR(4) in fullband and AR(1) in subband.

end points may be introduced to control the smoothness of the fitting function. The idea of the piecewise fitting can be extended to our application. We can consider signal modeling as analogous to (but not identical) a PSD fitting problem. Using a high-order model will cause problems similar to those encountered in curve-fitting, so we divide the PSD into consecutive segments and use a low-order model for each segment. Due to its simplicity, we use the AR model although its PSD is not a polynomial function. Subband decomposition is a perfect way to realize the piecewise modeling scheme.

Here, we present an illustrative example. Suppose that the input signal  $x(n)$  is an ARMA signal obtained by passing white noise through a system with the following transfer function:

$$H(z) = \frac{1 - z^{-1} + 0.3z^{-2}}{1 - z^{-1} + 0.5z^{-2}}. \quad (22)$$

Let AR( $i$ ) denote the  $i$ th order AR model. We first use an AR(4) process to model this signal. The optimal coefficients are found by solving the Yule-Walker equations. The tree-structure QMF bank with a prototype filter of length 32 is applied to obtain a four-band decomposition [24]. Each subband signal is then modeled as an AR(1) process. Fig. 1 shows the PSD's of the fullband AR(4) model and the subband AR(1) model. Note that the overall subband AR orders are the same as that of the fullband. The averaged square error (ASE) between the PSD's of the AR models and the original signal are also calculated. The ASE is 0.0225 for the fullband AR(4) model, while the ASE is 0.00132 for the subband AR(1) model. It is clear that subband modeling is preferable to fullband modeling. It is worth noting that there may be sharp changes in the subband modeling spectrum. For instance, each subband signal can be modeled as AR(0). In this case, the optimal filter may also have sharp changes in its spectrum. Constraints can be imposed to improve this situation, however, this will complicate the whole scheme. A simpler alternative is to add a postfilter to smooth the optimal filter spectrum. Details of this approach will be presented in a subsequent work.

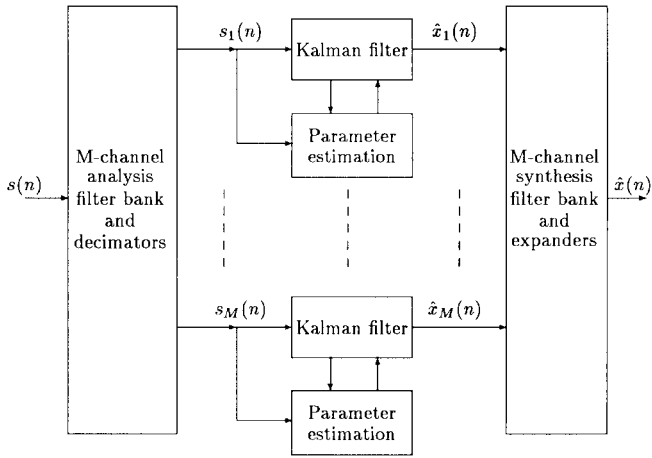


Fig. 2. Subband speech enhancement system.

The block diagram of the proposed speech enhancement system is shown in Fig. 2. Noisy speech  $s(n)$  is first split into a set of subband signals,  $s_i(n)$ , where  $i = 1, \dots, M$ , by an  $M$ -channel analysis filter bank and  $M$ -fold decimators. The subband signal  $s_i(n)$  can be expressed as follows:

$$s_i(n) = x_i(n) + v_i(n), \quad i = 1, \dots, M \quad (23)$$

where  $x_i(n)$  and  $v_i(n)$  are subband signals of  $x(n)$  and  $v(n)$ , respectively. If  $v(n)$  is white, we can approximate  $v_i(n)$  as white (if the subband filters are ideal,  $v_i(n)$  is exactly white). If  $v(n)$  is colored,  $v_i(n)$  is also colored. Thus, we model  $v_i(n)$  as an AR process. Since subband speech signals have simpler spectra than their fullband counterparts, they can be modeled as lower-order AR signals. We focus on the modeling of AR(1) and AR(0). The Kalman filtering operations will be greatly simplified in these cases. For AR(1),  $x_i(n)$  is expressed as follows.

$$x_i(n) = c_i x_i(n-1) + w_i(n) \quad (24)$$

where  $w_i(n)$  is a zero-mean white Gaussian process with a variance of  $\sigma_{w_i}^2$ . Equation (24) is the state equation for subband signals. Combining it with the measurement equation in (23), we can apply a bank of Kalman filters to subband speech signals. For AR(0), we can just set  $c_i = 0$  in (24).

To use the Kalman filter, the parameters  $c_i$  and  $\sigma_{w_i}^2$  must be estimated. The estimation method is described in the following subsection. The filtered subband signal, denoted as  $\hat{x}_i(n)$ , is up-sampled by expanders and then processed by an  $M$ -channel synthesis filter bank to reconstruct the filtered signal  $\hat{x}(n)$ .

### B. Aliasing Problem

Aliasing is an inherent problem that arises in subband processing systems. Consider the two-band case in Fig. 3. The output of the system is given by

$$\begin{aligned} \hat{x}(z) = & \frac{1}{2} [F_0(z)H_0(z^2)G_0(z) + F_1(z)H_1(z^2)G_1(z)]s(z) \\ & + \frac{1}{2} [F_0(-z)H_0(z^2)G_0(z) + F_1(-z)H_1(z^2)G_1(z)] \\ & \cdot s(-z) \end{aligned} \quad (25)$$

where  $H_0(z)$  and  $H_1(z)$  are transfer functions of subband processing filters. The second term on the right-hand side

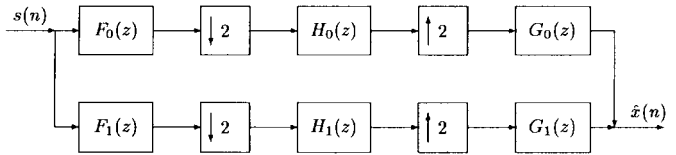


Fig. 3. A two-band processing system.

of (25) is the aliasing component. The perfect reconstruction conditions for a conventional QMF bank are known to be

$$F_0(z)G_0(z) + F_1(z)G_1(z) = Cz^{-L} \quad (26)$$

$$F_0(-z)G_0(z) + F_1(-z)G_1(z) = 0 \quad (27)$$

where  $L$  is an integer and  $C$  is a constant. It is apparent that even though (26) and (27) are satisfied, the aliasing component in (25) cannot be cancelled. To have a aliasing-free reconstruction for arbitrary  $H_0(z)$  and  $H_1(z)$ , we need a filter bank that satisfies the following conditions:

$$F_0(z)G_0(z) + F_1(z)G_1(z) = Cz^{-L} \quad (28)$$

$$F_0(-z)G_0(z) = 0 \quad \text{and} \quad F_1(-z)G_1(z) = 0. \quad (29)$$

Note that the conditions (28) and (29) are strict; it is difficult to design filters that satisfy these conditions. Many methods have been proposed to solve the problem [18]–[21], however, they are either computationally expensive or inappropriate for our use. Recently, an echo-cancellation algorithm, that combines an IIR filter bank and a notch filter, was proposed [22]. It was shown that the aliasing effect can be effectively reduced. Application of an IIR filter bank is beyond the scope of this paper, however, we found that in speech enhancement, if the input SNR is low, the filtering error due to aliasing tends to be masked by other effects. In Section V, we use simulation results to describe this.

### C. Parameter Estimation

In this subsection, we consider estimation of AR coefficients and driving-noise variances. As mentioned above, we use an adaptive prediction-error filter with the LMS algorithm for coefficient identification. For faster convergence, we adopt the normalized LMS algorithm. The update equation for the coefficient vector is given as follows:

$$\begin{aligned} \hat{\mathbf{a}}(n) = & \hat{\mathbf{a}}(n-1) + \frac{\mu}{\|\mathbf{x}(n-1)\|^2} \mathbf{x}(n-1) \\ & \cdot [\mathbf{x}(n) - \hat{\mathbf{a}}(n-1)^T \mathbf{x}(n-1)] \end{aligned} \quad (30)$$

where  $\hat{\mathbf{a}}(n) = [\hat{a}_1(n) \ \hat{a}_2(n) \ \dots \ \hat{a}_p(n)]^T$  is the estimate of AR coefficients in (1), and  $\|\mathbf{x}(n-1)\|$  is the Euclidean norm of  $\mathbf{x}(n-1)$ . The step size  $\mu$  in (30) determines the convergence rate and stability of  $\hat{\mathbf{a}}(n)$ . It has been shown [25] that when  $\mu$  is chosen properly, the expectation of  $\hat{\mathbf{a}}(n)$  will converge to  $\mathbf{a}$ . This type of identification is well known in AR modeling, however, clean speech is required in order to obtain an unbiased estimate of  $\mathbf{a}$ . In speech enhancement, only noisy speech  $s(n)$  is available for the normalized LMS filter. Thus,

the AR coefficient vector update equation becomes

$$\hat{\mathbf{a}}(n) = \hat{\mathbf{a}}(n-1) + \frac{\mu}{\|\mathbf{y}(n-1)\|^2} \mathbf{y}(n-1) \cdot [s(n) - \hat{\mathbf{a}}(n-1)^T \mathbf{y}(n-1)] \quad (31)$$

where  $\mathbf{y}(n) = [s(n) \ s(n-1) \ \dots \ s(n-p+1)]^T$ . Since  $s(n)$  is used instead of  $x(n)$ , the LMS algorithm will converge to a biased solution. Let  $R_x = E\{\mathbf{x}(n)\mathbf{x}(n)^T\}$ ,  $R_v = E\{\boldsymbol{\nu}(n)\boldsymbol{\nu}(n)^T\}$  where  $\boldsymbol{\nu}(n) = [v(n) \ v(n-1) \ \dots \ v(n-p+1)]^T$ , and  $P = E\{\mathbf{x}(n)x(n+1)\}$ . It can easily be shown that  $E\{\hat{\mathbf{a}}(n)\}$  will converge to  $R_x^{-1}P$  for (30) and to  $(R_x + R_v)^{-1}P$  for (31). Through extensive experimentation, we found that Kalman filtering performance degradation due to biased AR coefficients is small when driving-noise variance estimation is also biased in some way. In other words, some biased  $\sigma_w^2$  can compensate for the effect of biased AR coefficients. Thus, we will not pursue accurate estimation of  $\mathbf{a}$ , instead, we focus on estimating  $\sigma_w^2$ .

Here, we propose a method for identifying  $\sigma_w^2$ . First, we consider the white noise case. Denote  $k_i(n)$  as the  $i$ th element of  $\mathbf{k}(n)$ ,  $P_{i,j}(n)$  the  $ij$ th element of  $\mathbf{P}(n)$ , and  $P_{i,j}(n|n-1)$  the  $ij$ th element of  $\mathbf{P}(n|n-1)$ . We also define the prediction-error for noisy input as

$$\begin{aligned} \xi(n) &= s(n) - \mathbf{h}^T \mathbf{F} \hat{\mathbf{x}}(n-1) \\ &= x(n) - \mathbf{a}^T \hat{\mathbf{x}}(n-1) + v(n). \end{aligned} \quad (32)$$

The variance of  $\xi(n)$  can be obtained as follows:

$$\begin{aligned} \sigma_\xi^2 &= E\{\xi(n)^2\} = E\{[x(n) - \mathbf{a}^T \hat{\mathbf{x}}(n-1)]^2\} + \sigma_v^2 \\ &= P_{1,1}(n|n-1) + \sigma_v^2 \end{aligned} \quad (33)$$

where  $P_{1,1}(n|n-1) = E\{[x(n) - \mathbf{a}^T \hat{\mathbf{x}}(n-1)]^2\}$ . From (6), (7), and (33), we have

$$\begin{aligned} k_1(n) &= P_{1,1}(n|n-1)[P_{1,1}(n|n-1) + \sigma_v^2]^{-1} \\ &= P_{1,1}(n|n-1)\sigma_\xi^{-2} \\ &= [\mathbf{a}^T \mathbf{P}(n-1)\mathbf{a} + \sigma_w^2]\sigma_\xi^{-2}. \end{aligned} \quad (34)$$

Thus,  $\sigma_w^2$  can be obtained by rewriting (34) as

$$\sigma_w^2 = k_1(n)\sigma_\xi^2 - \mathbf{a}^T \mathbf{P}(n-1)\mathbf{a}. \quad (35)$$

In practice, we cannot obtain  $\mathbf{a}$ ; we use  $\hat{\mathbf{a}}(n)$  to approximate it. A fading-memory average is then used to recursively estimate  $\sigma_w^2$ :

$$\begin{aligned} \hat{\sigma}_w^2(n+1) &= \beta \hat{\sigma}_w^2(n) + (1-\beta) \\ &\cdot [k_1(n)\xi(n)^2 - \hat{\mathbf{a}}^T(n)\mathbf{P}(n-1)\hat{\mathbf{a}}(n)] \end{aligned} \quad (36)$$

where  $\beta$  is a forgetting factor and is set close to 1. This forgetting factor controls the variance of  $\hat{\sigma}_w^2$ . If  $\beta$  is larger, the variance will be smaller. However, the signal tracking ability becomes poorer. In practice, we found that a value around 0.95 provides a good compromise. The compensation capacity of the estimate in (36) for biased  $E\{\hat{\mathbf{a}}(n)\}$  is discussed in Section IV.

The above method can be extended to estimate  $\sigma_w^2$  in colored noise. Consider the state augmentation formulation. We first redefine  $\xi(n)$  in (32) as

$$\xi(n) = s(n) - \bar{\mathbf{h}}^T \bar{\mathbf{F}} \bar{\hat{\mathbf{x}}}(n-1) = \bar{\mathbf{h}}^T [\bar{\mathbf{x}}(n) - \bar{\mathbf{F}} \bar{\hat{\mathbf{x}}}(n-1)]. \quad (37)$$

The variance of  $\xi(n)$  is then

$$\begin{aligned} \sigma_\xi^2(n) &= \bar{\mathbf{h}}^T E\{[\bar{\mathbf{x}}(n) - \bar{\mathbf{F}} \bar{\hat{\mathbf{x}}}(n-1)][\bar{\mathbf{x}}(n) - \bar{\mathbf{F}} \bar{\hat{\mathbf{x}}}(n-1)]^T\} \bar{\mathbf{h}} \\ &= \bar{\mathbf{h}}^T \mathbf{P}(n|n-1) \bar{\mathbf{h}}. \end{aligned} \quad (38)$$

From (38), the gain vector  $\mathbf{k}(n)$  is written as follows:

$$\begin{aligned} \mathbf{k}(n) &= \mathbf{P}(n|n-1) \bar{\mathbf{h}} [\bar{\mathbf{h}}^T \mathbf{P}(n|n-1) \bar{\mathbf{h}}]^{-1} \\ &= [\bar{\mathbf{F}} \mathbf{P}(n-1) \bar{\mathbf{F}} + \bar{\mathbf{G}} \bar{\mathbf{Q}} \bar{\mathbf{G}}] \bar{\mathbf{h}} \sigma_\xi^{-2}. \end{aligned} \quad (39)$$

The first element of  $\mathbf{k}(n)$  is

$$k_1(n) = \{[\mathbf{a}^T \ \mathbf{b}^T] \mathbf{P}(n-1) [\mathbf{a}^T \ \mathbf{0}]^T + \sigma_w^2\} \sigma_\xi^{-2}. \quad (40)$$

Thus,  $\sigma_w^2$  can be obtained by rewriting (40) as

$$\sigma_w^2 = k_1(n)\sigma_\xi^2 - [\mathbf{a}^T \ \mathbf{b}^T] \mathbf{P}(n-1) [\mathbf{a}^T(n) \ \mathbf{0}]^T. \quad (41)$$

Again, the fading-memory technique is used to recursively estimate  $\sigma_w^2$ :

$$\begin{aligned} \hat{\sigma}_w^2(n+1) &= \beta \hat{\sigma}_w^2(n) + (1-\beta) \\ &\cdot \{k_1(n)\xi^2(n) - [\hat{\mathbf{a}}(n)^T \ \mathbf{b}^T] \mathbf{P}(n-1) [\hat{\mathbf{a}}^T(n) \ \mathbf{0}]^T\}. \end{aligned} \quad (42)$$

Unfortunately, the method for estimating  $\sigma_w^2$  described above cannot be used in the measurement-difference formulation. So, we developed another method. Using (21), we can express (32) as

$$\xi(n) = x(n) - \mathbf{a}^T \tilde{\mathbf{x}}(n-1|n-2) + v(n). \quad (43)$$

Using (2), we can describe the driving noise as

$$w(n) = x(n) - \mathbf{a}^T(n) \mathbf{x}(n-1). \quad (44)$$

Thus, from (43) and (44), the variance of  $\xi(n)$  can be written as

$$\begin{aligned} \sigma_\xi^2 &= E\{\xi(n)^2\} \\ &= E\{[x(n) - \mathbf{a}^T \tilde{\mathbf{x}}(n-1|n-2) + v(n)]^2\} \\ &= \sigma_w^2 + \sigma_v^2 + E\{[\mathbf{a}^T \mathbf{x}(n-1) - \mathbf{a}^T \tilde{\mathbf{x}}(n-1|n-2)]^2\}. \end{aligned} \quad (45)$$

This shows that  $\sigma_\xi^2 - \sigma_v^2 > \sigma_w^2$ . Defining the filtered prediction error as the expectation of (44) conditioned on  $\{s(1), s(2), \dots, s(n)\}$ , we have

$$\begin{aligned} \varepsilon(n) &= E\{x(n)|s(1), s(2), \dots, s(n)\} \\ &\quad - \mathbf{a}^T E\{\mathbf{x}(n-1)|s(1), s(2), \dots, s(n)\} \\ &= \tilde{x}(n|n-1) - \mathbf{a}^T \tilde{\mathbf{x}}(n-1). \end{aligned} \quad (46)$$

Thus,  $\varepsilon(n)$  is the optimal estimate of  $w(n)$ . As a consequence,  $\sigma_w^2 > \sigma_\varepsilon^2 = E\{\varepsilon(n)^2\}$ . The above results indicate that the true

$\sigma_w^2$  lies somewhere between  $\sigma_\xi^2 - \sigma_v^2$  and  $\sigma_\varepsilon^2$ . Thus, we can write  $\sigma_w^2$  as a linear combination of these two values.

$$\sigma_w^2 = \alpha\sigma_\varepsilon^2 + (1 - \alpha)(\sigma_\xi^2 - \sigma_v^2) \quad (47)$$

where  $\alpha$  is a constant between 0 and 1. Since  $\xi(n)$  contains unprocessed noise, the variance of the estimated  $\sigma_\xi^2$  is usually larger than that of  $\sigma_\varepsilon^2$ . Thus, the choice of  $\alpha$  should favor  $\sigma_\varepsilon^2$ . A proper choice of  $\alpha$  depends on the signal-to-noise ratio (SNR). When the SNR is high, the estimate of  $\sigma_\xi^2$  is more reliable and  $\alpha$  can be smaller. By contrast, when the SNR is low,  $\alpha$  can be larger. Through simulations, we found that a good  $\alpha$  is 0.9–0.7 for input SNR's 0–20 dB. As in (36) and (42), we still use a fading-memory average to recursively estimate  $\sigma_w^2(n+1)$ :

$$\hat{\sigma}_w^2(n+1) = \beta\hat{\sigma}_w^2(n) + (1 - \beta) \cdot \{\alpha\varepsilon^2(n) + (1 - \alpha)[\xi^2(n) - \sigma_v^2]\}. \quad (48)$$

Note that  $\xi(n)$  and  $\varepsilon(n)$  are as in (43) and (46), with  $\mathbf{a}$  being replaced by  $\hat{\mathbf{a}}(n)$ .

#### D. Computational Complexity

In this subsection, we discuss the computational complexities of Kalman filtering in the fullband and subband domains using the proposed schemes. First, we define three terms for measuring complexity: MPU, multiplications per unit of time; DVU, divisions per unit of time; and APU, additions per unit of time. According to (1) and (9), speech is modeled as AR( $p$ ) and noise AR( $q$ ). Note that AR(0) means 0th-order AR modeling, i.e., white noise. For the cases in which  $p \geq 1$  and  $q = 0$  (white measurement noise), the Kalman filter described in (5)–(8) requires  $3p^2 + 2p$  MPU,  $p$  DVU, and  $3p^2 + 2$  APU. Note that  $\mathbf{F}$  (or  $\mathbf{F}_v$ ) is a shift matrix. Thus, the number of multiplications in the Kalman filter is reduced from order  $O(p^3)$  to  $O(p^2)$ .

For  $p \geq 1$  and  $q \geq 1$  (colored measurement noise), the two types of Kalman filters have different complexities. For the state-augmentation type, the Kalman filter requires  $3(p+q)^2 + 2(p+q)$  MPU,  $(p+q)$  DVU, and  $3(p+q)^2 + 3$  APU. For the measurement-difference type, the Kalman filter requires  $6p^2 + 4p + q + 2$  MPU,  $p+1$  DVU, and  $6p^2 - p + 2q + 3$  ADU. Specifically, when  $p = 0$  and  $q = 0$ , the Kalman filter only requires 1 MPU, 1 DVU, and 1 ADU. The output of the Kalman filter in this case is just a product of scalar gain and noisy speech. As to parameter identification, the LMS prediction-error filter in (31) requires  $2p + 1$  MPU, 1 DVU, and  $2p + 2$  ADU. The identification of driving-noise variance using (36) requires  $p^2 + p + 4$  MPU and  $p^2 + 1$  ADU, using (42) requires  $(p+q)^2 + p + 4$  MPU and  $(p+q)^2 - q + 1$  ADU, and using (48) requires  $2p + 6$  MPU and  $2p + 3$  ADU.

For subband processing, we first consider the overhead complexity of the QMF bank. The QMF bank can be efficiently implemented by using the so-called polyphase structure [23]. If the order of the QMF filter is  $N$ , then the complexity of an  $M$ -channel QMF bank is  $2N/M + M$  MPU and  $2N/M + (M - 1)^2/M$  APU. The overall complexity of subband processing is equal to the summation of the QMF bank, parameter identification, and the Kalman filtering complexities. Note that the order of the Kalman filter in the subband domain is smaller

TABLE I  
OVERALL COMPLEXITIES OF SPEECH ENHANCEMENT SYSTEMS (WHITE NOISE)

Operations	Fullband (4,0)			Subband (1,0)		
	MPU	DVU	ADU	MPU	DVU	ADU
Kalman F.	56	4	50	5	1	5
QMF	-	-	-	29	0	27
LMS	9	1	10	3	1	4
$\hat{\sigma}_w^2$	24	0	17	6	0	2
Total	89	5	77	43	2	38

TABLE II  
OVERALL COMPLEXITIES OF SPEECH ENHANCEMENT SYSTEMS (COLORED NOISE; STATE AUGMENTATION)

Operations	Fullband (4,4)			Subband (1,1)		
	MPU	DVU	ADU	MPU	DVU	ADU
Kalman F.	208	8	195	16	2	15
QMF	-	-	-	29	0	27
LMS	9	1	10	3	1	4
$\hat{\sigma}_w^2$	72	0	61	9	0	4
Total	289	9	266	57	3	50

TABLE III  
OVERALL COMPLEXITIES OF SPEECH ENHANCEMENT SYSTEMS (COLORED NOISE; MEASUREMENT DIFFERENCE)

Operations	Fullband (4,4)			Subband (1,1)		
	MPU	DVU	ADU	MPU	DVU	ADU
Kalman F.	118	5	103	13	2	10
QMF	-	-	-	29	0	27
LMS	9	1	10	3	1	4
$\hat{\sigma}_w^2$	14	0	11	8	0	5
Total	141	6	124	53	3	46

than that in the fullband domain. Thus, the complexity of the Kalman filter can be reduced in the subband domain. We now use an example to illustrate this. We first consider the white noise case. Let  $N = 50$ ,  $M = 4$ ,  $(p, q) = (4, 0)$  in the fullband domain, and  $(p, q) = (1, 0)$  in the subband domain. Table I summarizes the computational complexities of the whole fullband and subband enhancement systems. As we can see, the overall amount of computation required for subband processing is less than half of that required for fullband processing. When colored noise is considered, the computational saving is even more significant. Table II shows speech enhancement system complexities using a state-augmentation type Kalman filter. Here, we let  $(p, q) = (4, 4)$  for fullband and  $(p, q) = (1, 1)$  for subband. The complexity of subband processing is less than one-fifth of that required for fullband processing. The high complexity required for fullband processing is due to its high-order augmented state. This can be remedied by using a measurement-difference type Kalman filter. The results shown in Table III indicate that subband processing is still far less complex than fullband processing.

#### IV. PERFORMANCE ANALYSIS

In this section, we analyze the performance of a Kalman filter with biased parameters. Specifically, we consider only the case in which speech is modeled as AR(1) and the measurement noise is modeled as white. Consider the subband signal described in (23) and (24). For simplicity, we will ignore the subscript  $i$ . Let these signals be stationary and denote estimates of  $c$  and  $\sigma_w^2$  as  $\hat{c}$  and  $\hat{\sigma}_w^2$ . These parameters are then

used in the Kalman filter. Thus, the filtering equations are

$$P(n|n-1) = P(n-1)\hat{c}^2 + \hat{\sigma}_w^2 \tag{49}$$

$$k(n) = P(n|n-1)/[\sigma_v^2 + P(n|n-1)] \tag{50}$$

$$\hat{x}(n) = \hat{c}\hat{x}(n-1) + k(n)[s(n) - \hat{c}\hat{x}(n-1)] \tag{51}$$

$$P(n) = (1 - k(n))P(n|n-1). \tag{52}$$

If the steady-state is reached,  $P(n)$  will be equal to  $P(n-1)$ . Substituting (50) and (49) into (52), the steady-state  $P(n)$  can be found as the positive root of the following equation:

$$\hat{c}^2 P(n)^2 + [\hat{\sigma}_w^2 + (1 - \hat{c}^2)\sigma_v^2] - \hat{\sigma}_w^2 \sigma_v^2 = 0. \tag{53}$$

From (49) and (50), we can obtain the steady-state Kalman gain as

$$k = \frac{\hat{c}^2 P(n) + \hat{\sigma}_w^2}{\hat{c}^2 P(n) + \hat{\sigma}_w^2 + \sigma_v^2}. \tag{54}$$

Thus, from (51), the steady-state filtering equation can be written as

$$\hat{x}(n) - (1 - k)\hat{c}\hat{x}(n-1) = ks(n). \tag{55}$$

Taking the  $z$ -transform of (55), we obtain the transfer function of the steady-state Kalman filter:

$$H(z) = \frac{\hat{X}(z)}{S(z)} = \frac{k}{1 - (1 - k)\hat{c}z^{-1}} \tag{56}$$

where  $\hat{X}(z)$  and  $S(z)$  denote the  $z$ -transforms of  $\hat{x}$  and  $s(n)$ , respectively. Let  $h(n)$  be the impulse response of the Kalman filter. The steady-state output of the Kalman filter can be written as the convolution of  $h(n)$  and  $s(n)$ :

$$\hat{x}(n) = s(n) * h(n) = \sum_{\alpha=-\infty}^{\infty} s(n - \alpha)h(\alpha). \tag{57}$$

Define the filtering error based on biased parameters as

$$\varepsilon(n) = x(n) - \hat{x}(n). \tag{58}$$

The autocorrelation function of  $\varepsilon(n)$  is then

$$\begin{aligned} R_\varepsilon(l) &= E\{\varepsilon(n)\varepsilon(n-l)\} \\ &= E\{[x(n) - \hat{x}(n)][x(n-l) - \hat{x}(n-l)]\} \\ &= E\{x(n)x(n-l)\} + E\{x(n)\hat{x}(n-l)\} \\ &\quad + E\{\hat{x}(n)x(n-l)\} + E\{\hat{x}(n)\hat{x}(n-l)\} \\ &= R_x(l) - E\left\{x(n) \sum_{\alpha=-\infty}^{\infty} s(n-l-\alpha)h(\alpha)\right\} \\ &\quad - E\left\{\sum_{\alpha=-\infty}^{\infty} s(n-\alpha)h(\alpha)x(n-l)\right\} \\ &\quad + \sum_{\alpha=-\infty}^{\infty} \sum_{\beta=-\infty}^{\infty} s(n-\alpha)s(n-l-\beta)h(\alpha)h(\beta) \\ &= R_x(l) - h(-l) * R_x(l) - h(l) * R_x(l) + h(l) \\ &\quad * h(-l) * R_s(l) \end{aligned} \tag{59}$$

where  $R_x(l) = E\{x(n)x(n-l)\}$  and  $R_s(l) = E\{s(n)s(n-l)\}$ . The complex PSD of  $\varepsilon(n)$  is obtained as the  $z$ -transform of  $R_\varepsilon(l)$ :

$$\begin{aligned} S_\varepsilon(z) &= S_x(z) - H(z^{-1})S_x(z) - H(z)S_x(z) \\ &\quad + H(z)H(z^{-1})S_s(z) \end{aligned} \tag{60}$$

where

$$S_x(z) = \frac{\sigma_w^2}{(1 - cz)(1 - cz^{-1})} \tag{61}$$

is the complex PSD of  $x(n)$ , and

$$S_s(z) = S_x(z) + \sigma_v^2 \tag{62}$$

is the complex PSD of  $s(n)$ . Substituting (56), (61), and (62) into (60), (60) can be rewritten as

$$\begin{aligned} S_\varepsilon(z) &= \frac{[(1-k)^2(1-\hat{c}z^{-1})(1-\hat{c}z)\sigma_w^2 + k^2(1-cz^{-1})(1-cz)\sigma_v^2]z^2}{(1-k)\hat{c}c[z - (1-k)\hat{c}]\{z - [(1-k)\hat{c}]^{-1}\}(z-c)(z-c^{-1})}. \end{aligned} \tag{63}$$

The variance of  $\varepsilon(n)$  (output mean square error), which is equal to  $R_\varepsilon(0)$ , can be computed from the integral of  $S_\varepsilon(z)z^{-1}$  around the unit circle:

$$\begin{aligned} E\{\varepsilon(n)^2\} &= R_\varepsilon(0) = \frac{1}{2\pi j} \oint S_\varepsilon(z)z^{-1} dz \\ &= \left\{ \frac{(1-k)^2 \hat{c}k [1 - (1-k)\hat{c}^2]}{[1 - (1-k)^2 \hat{c}^2][c - (1-k)\hat{c}][1 - (1-k)\hat{c}c]} \right. \\ &\quad \left. + \frac{(1-k)^2 (c - \hat{c})(1 - \hat{c}c)}{(1 - c^2)[c - (1-k)\hat{c}][1 - (1-k)\hat{c}c]} \right\} \sigma_w^2 \\ &\quad + \frac{k^2 \sigma_v^2}{[1 - (1-k)^2 \hat{c}^2]}. \end{aligned} \tag{64}$$

Note that  $k$  is a function of  $\hat{c}$  and  $\hat{\sigma}_w^2$  in (54). If we assume  $c$  and  $\sigma_w^2$  are fixed, then the output mean square error (MSE)  $R_\varepsilon(0)$  can be rewritten as a function of  $\hat{c}$  and  $\hat{\sigma}_w^2$ . We denote this function as  $\phi(\hat{c}, \hat{\sigma}_w^2)$ . Thus

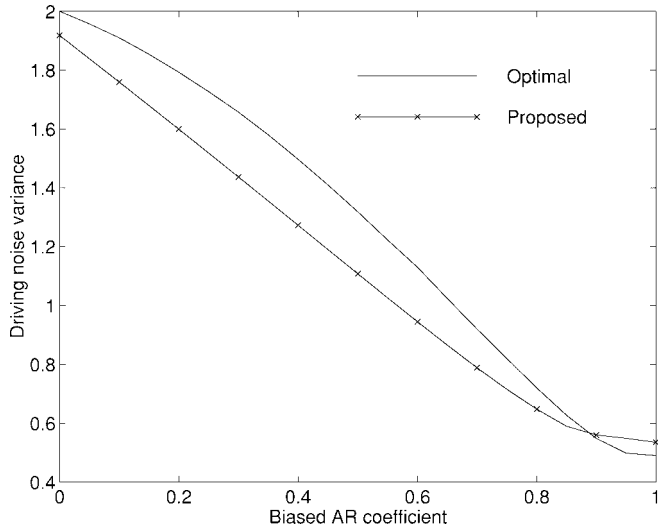
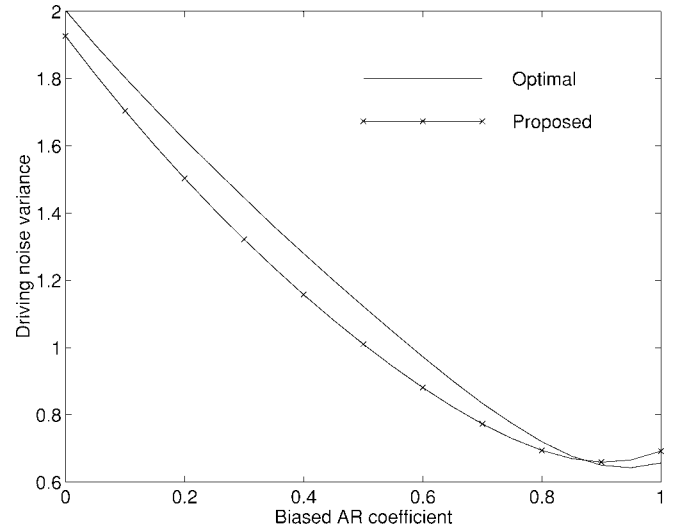
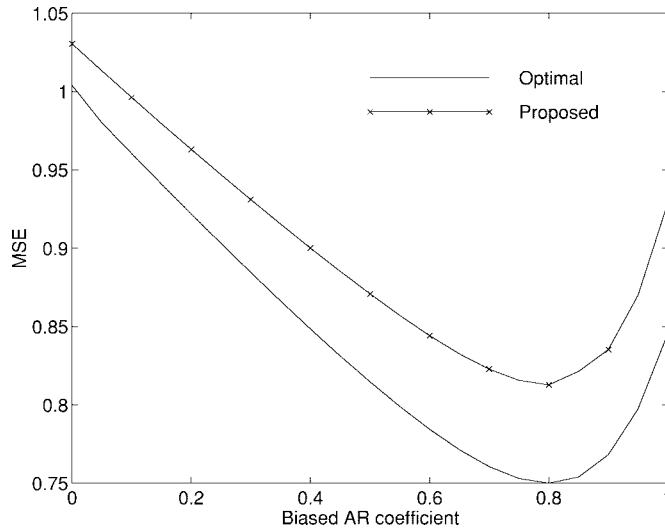
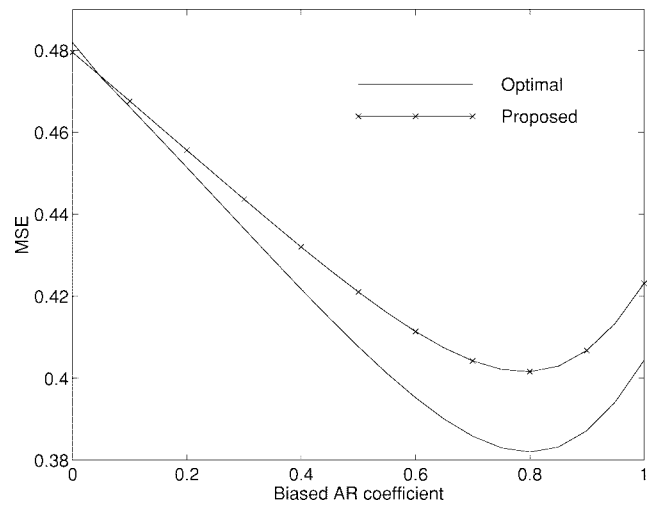
$$E\{\varepsilon(n)^2\} = \phi(\hat{c}, \hat{\sigma}_w^2). \tag{65}$$

Apparently, the minimum value of  $\phi(\hat{c}, \hat{\sigma}_w^2)$  is obtained at the optimal solution  $(\hat{c}, \hat{\sigma}_w^2) = (c, \sigma_w^2)$ , i.e., the true AR parameters.

In practice, estimates of AR coefficients from noisy speech are always biased. Thus, the minimum value is never obtained. Given a biased AR coefficient  $\hat{c}$ , however, we can find a  $\hat{\sigma}_w^2$  that minimizes  $\phi(\hat{c}, \hat{\sigma}_w^2)$ . This corresponds to the optimal result in the biased environment. Note that as we show below, this optimal  $\hat{\sigma}_w^2$  is not equal to  $\sigma_w^2$ . In other words, some biased driving-noise variance can compensate for degradation due to biased AR coefficients. For a given  $\hat{c}$ , we denote the value of  $\hat{\sigma}_w^2$  that minimizes  $R_\varepsilon(0)$  as  $\rho(\hat{c})$ :

$$\rho(\hat{c}) = \underset{\hat{\sigma}_w^2}{\text{Minimize}} \phi(\hat{c}, \hat{\sigma}_w^2). \tag{66}$$

This is to say that for a biased AR coefficient  $\hat{c}$ , the optimal driving-noise variance is  $\rho(\hat{c})$ . The solution can be found by solving  $\partial\phi(\hat{c}, \hat{\sigma}_w^2)/\partial\hat{\sigma}_w^2 = 0$ . Once  $\rho(\hat{c})$  has been found, we


 Fig. 4.  $\rho(\hat{c})$  and  $\theta(\hat{c})$  for  $c = 0.8$ ,  $\sigma_w^2 = 0.72$  (SNR = 0 dB).

 Fig. 6.  $\rho(\hat{c})$  and  $\theta(\hat{c})$  for  $c = 0.8$ ,  $\sigma_w^2 = 0.72$  (SNR = 5 dB).

 Fig. 5.  $\phi[\hat{c}, \rho(\hat{c})]$ , and  $\psi[\hat{c}, \theta(\hat{c})]$  for  $c = 0.8$ ,  $\sigma_w^2 = 0.72$  (SNR = 0 dB).

 Fig. 7.  $\phi[\hat{c}, \rho(\hat{c})]$  and  $\psi[\hat{c}, \theta(\hat{c})]$  for  $c = 0.8$ ,  $\sigma_w^2 = 0.72$  (SNR = 5 dB).

can substitute back to find  $\phi[\hat{c}, \rho(\hat{c})]$ . The function  $\phi[\hat{c}, \rho(\hat{c})]$  is informative. By comparing it to  $\phi(c, \sigma_w^2)$ , we can assess the effects of biased AR coefficients.

Using the analysis above, we now evaluate the performance of our estimate of the driving-noise variance in (36). For the AR(1) process, (36) can be rewritten as follows:

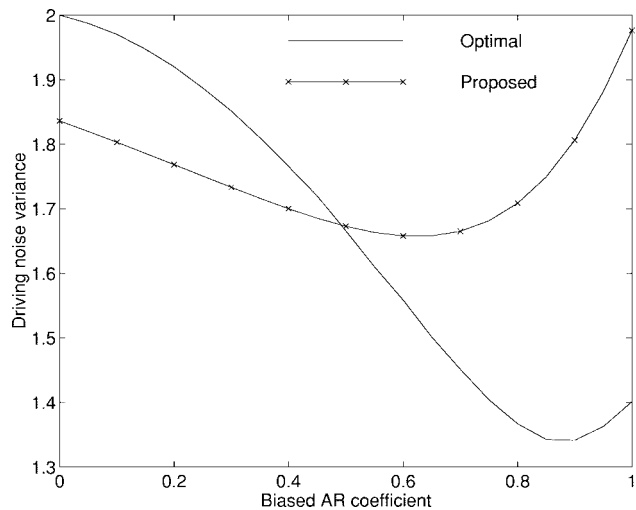
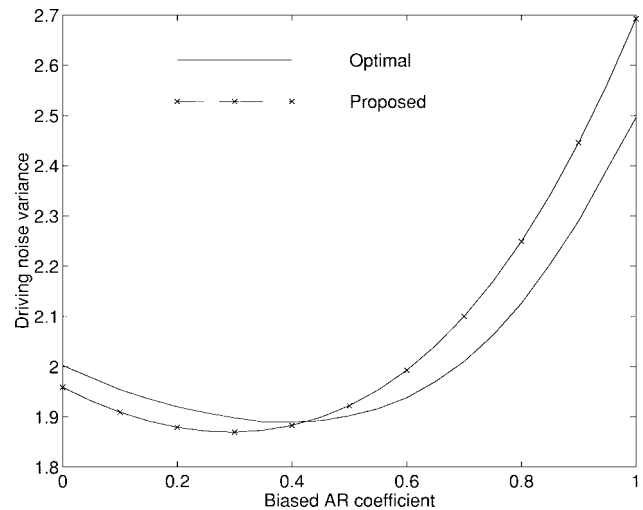
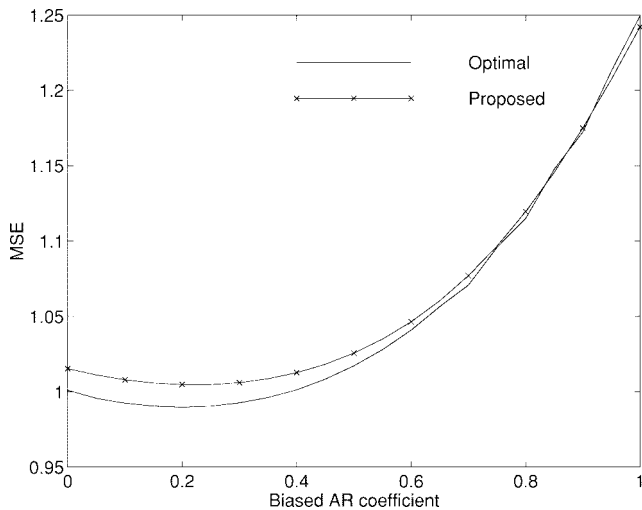
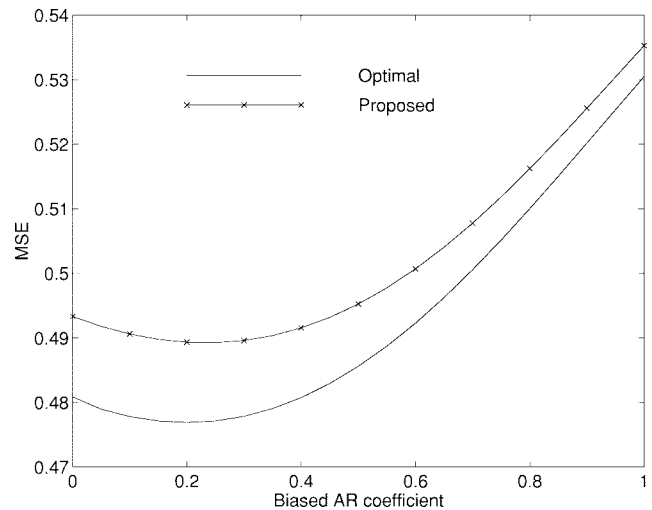
$$\begin{aligned} \hat{\sigma}_w^2(n+1) = & \beta \hat{\sigma}_w^2(n) + (1-\beta) \\ & \cdot \{k(n)[s(n) - \hat{c}\hat{x}(n-1)]^2 - P(n-1)\hat{c}^2\}. \end{aligned} \quad (67)$$

Two sets of simulations were conducted: one for a narrow-band AR signal with  $c = 0.8$  and  $\sigma_w^2 = 0.72$ ; and the other for a wide-band AR signal with  $c = 0.2$  and  $\sigma_w^2 = 1.92$ . In both cases, white Gaussian noise was added to the AR signals yielding SNR's of 0 and 5 dB. Since computing a theoretical  $\rho(\hat{c})$  using its closed-form solution is tedious, we used numerical methods instead. For filtering simulations, a given biased AR coefficient was applied to the Kalman filter described in (49)–(52), and (67) was used to estimate the

driving-noise variance of speech. For a given  $\hat{c}$ , we recorded the steady-state values of the driving-noise variances estimated by (67) denoted as  $\theta(\hat{c})$ , and the output MSE's denoted as  $\psi[\hat{c}, \theta(\hat{c})]$ . These values were compared with  $\rho(\hat{c})$  and  $\phi[\hat{c}, \rho(\hat{c})]$ .

Figs. 4–7 show the results for the narrow-band signal, and Figs. 8–11 those for the wide-band signal. As these figures show, output MSE curves obtained using our estimates were always close to the optimal ones, especially for the wide-band signal. This can be seen in Figs. 9 and 11. The maximum difference in MSE from the optimal was less than 3% (appearing when  $\hat{c} = 0$ ). Even in the narrow-band cases, as Figs. 5 and 7 show, the maximum MSE difference was below 10% (appearing when  $\hat{c} = 0.8$ ). It is interesting to note that in Fig. 8, the difference between  $\rho(\hat{c})$  and  $\theta(\hat{c})$  is large in many places, however,  $\phi[\hat{c}, \rho(\hat{c})]$  and  $\psi[\hat{c}, \theta(\hat{c})]$  in Fig. 9 are still very close. We can also observe the effect of biased AR coefficients. Consider the worst case shown in Fig. 5. In this case, the true AR coefficient was 0.8. Suppose that we had



Fig. 8.  $\rho(\hat{c})$  and  $\theta(\hat{c})$  for  $c = 0.2$ ,  $\sigma_w^2 = 1.92$  (SNR = 0 dB).Fig. 10.  $\rho(\hat{c})$  and  $\theta(\hat{c})$  for  $c = 0.2$ ,  $\sigma_w^2 = 1.92$  (SNR = 5 dB).Fig. 9.  $\phi[\hat{c}, \rho(\hat{c})]$  and  $\psi[\hat{c}, \theta(\hat{c})]$  for  $c = 0.2$ ,  $\sigma_w^2 = 1.92$  (SNR = 0 dB).Fig. 11.  $\phi[\hat{c}, \rho(\hat{c})]$  and  $\psi[\hat{c}, \theta(\hat{c})]$  for  $c = 0.2$ ,  $\sigma_w^2 = 1.92$  (SNR = 5 dB).

applied a biased value, say, 0.5. The coefficient would have deviated 37.5%. The figures show that the optimal MSE only increased by 8.6%. Consider another case, say Fig. 9, in which the true coefficient was 0.2. If we had used a biased value of 0, the coefficient would have deviated 100%, but the optimal MSE only increased by 1.1%. Similar effects can be observed if a biased value of 0.4 is used. From the discussion above, we can conclude that accurate estimates of AR coefficients are not necessary, and our estimate of driving-noise variance can yield good performance in biased environments. Using the proposed scheme, degradation due to biased coefficients can be ignored.

## V. SIMULATIONS

To evaluate the performance of our approach, we carried out some simulations. Four “speeches” were used, as follows.

- 1) She had your dark suit in greasy wash water all year.
- 2) Don't ask me to carry an oily rag like that.
- 3) What is England's estimated time of arrival at Townsville.
- 4) Draw a chart centered around fox using stereo graphic projection.

These speeches were obtained from the DARPA speech database. They were uttered by two female speakers and one male speaker and digitized at an 8-kHz sampling frequency with 16-bit quantization. Three noises were used to contaminate the speech: additive white Gaussian noise, automobile engine noise, and motorcycle exhaust-pipe noise. The last two were obtained by recording a 1600cc sedan and a 125cc motorcycle; their spectra are shown in Fig. 12. Two objective performance criteria were used to evaluate filtering results, namely SNR and segmental SNR (SSNR) improvements. The input SNR was defined as:

$$\text{SNR}_{\text{in}} \triangleq 10 \log_{10} \frac{\frac{1}{N} \sum_{n=1}^{n=N} x^2(n)}{\frac{1}{N} \sum_{n=1}^{n=N} [s(n) - x(n)]^2} \quad (68)$$

where  $N$  was the length of speech. The output SNR,  $\text{SNR}_{\text{out}}$ , was defined as that in (68) except that  $s(n)$  was replaced by  $\hat{x}(n)$ . The SNR improvement was then  $\text{SNR}_{\text{out}} - \text{SNR}_{\text{in}}$ .

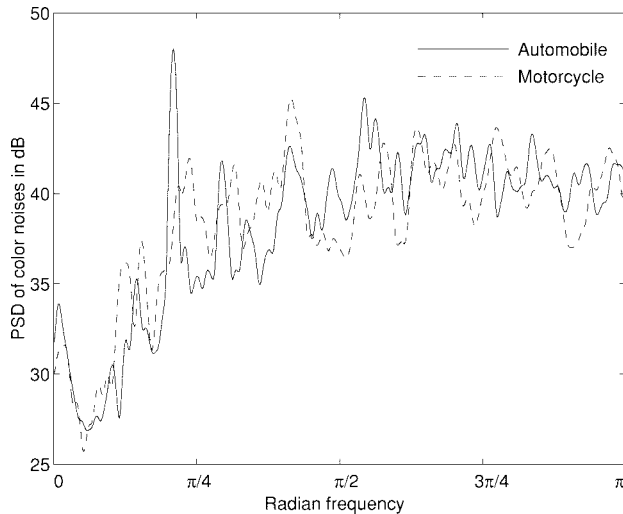


Fig. 12. PSD's of color noises.

and SNR's was represented in dB. To calculate SSSNR, we first divided a speech signal into consecutive segments and then averaged SNR's obtained from those segments. The improvement in SSSNR was defined as that in SNR.

Noise sequences were added to speeches yielding 0 and 5 dB  $\text{SNR}_{\text{in}}$ . The forgetting factor  $\beta$  in (36) was set to 0.95. The QMF bank was implemented using a tree structure with a prototype filter of length 16 [24]. For comparison, we also considered ideal cases in which noise-free speeches were assumed to be available. In this circumstance, AR parameters could be accurately estimated using some complex methods. We applied the recursive maximum-likelihood algorithm (RMLE) [27]. These results provided references for performance bounds. Table IV shows averaged SNR improvements or all speeches in white noise environments. The term "BN" in the tables means "band number." The AR orders of speech and noise are represented by  $(p, q)$ ; the first element is the order of speech, and the second is that of noise. To determine performance degradation due to model simplification, we also carried out simulations in which colored noise was modeled as white noise, and in which even speech was modeled as white signal. We look first at the results for the ideal cases. We can easily see that improvement increased as band number was increased. However, for actual cases, the best performance was achieved when the band number was four. In Table IV (white noise), we find that the SNR improvement for four-band decomposition was about 0.8 dB higher than that of fullband processing when speech signals were modeled as colored signal. The difference was 1.3 dB when speech was modeled as white signal. Note that for subband processing, filtering performance was almost undegraded for white signal modeling of speech. The SNR improvement for subband processing using four-band decomposition was about 1 dB lower than that for ideal cases.

For simulations of colored noise, we set  $\beta = 0.95$  in (42) and (48), and  $\alpha = 0.9$  in (48). Tables V and VI show the SNR improvement when automobile noise was used. Two types of Kalman filters were used here, state-augmentation and measurement-difference. We found that subband processing

TABLE IV  
SNR IMPROVEMENT OF SPEECH ENHANCEMENT FOR WHITE NOISE

BN	$\text{SNR}_{\text{in}} = 0$ dB			$\text{SNR}_{\text{in}} = 5$ dB		
	Ideal	Actual		Ideal	Actual	
1	7.278 (4,0)	6.070 (4,0)	5.550 (0,0)	5.425 (4,0)	4.468 (4,0)	3.960 (0,0)
2	7.292 (1,0)	6.766 (1,0)	6.642 (0,0)	5.655 (1,0)	5.158 (1,0)	5.006 (0,0)
4	7.851 (1,0)	6.889 (1,0)	6.825 (0,0)	6.144 (1,0)	5.257 (1,0)	5.205 (0,0)
8	8.259 (1,0)	6.672 (1,0)	6.545 (0,0)	6.430 (1,0)	4.942 (1,0)	4.846 (0,0)

TABLE V  
SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR AUTOMOBILE NOISE ( $\text{SNR}_{\text{in}} = 0$  dB)

BN	State Augmentation			Measurement Diff.		
	Ideal	Actual		Ideal	Actual	
1	7.862 (4,4)	6.576 (4,4)	5.915 (4,0)	5.446 (0,0)	9.478 (4,4)	6.890 (4,4)
2	8.675 (1,1)	7.968 (1,1)	7.678 (1,0)	7.556 (0,0)	9.360 (1,1)	8.220 (1,1)
4	9.437 (1,1)	8.303 (1,1)	7.993 (1,0)	8.009 (0,0)	9.772 (1,1)	8.557 (1,1)
8	9.878 (1,1)	8.100 (1,1)	7.913 (1,0)	7.853 (0,0)	10.20 (1,1)	8.386 (1,1)

TABLE VI  
SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR AUTOMOBILE NOISE ( $\text{SNR}_{\text{in}} = 5$  dB)

BN	State Augmentation			Measurement Diff.		
	Ideal	Actual		Ideal	Actual	
1	5.984 (4,4)	4.957 (4,4)	4.438 (4,0)	3.902 (0,0)	7.550 (4,4)	5.544 (4,4)
2	6.852 (1,1)	6.145 (1,1)	5.958 (1,0)	5.787 (0,0)	7.423 (1,1)	6.266 (1,1)
4	7.478 (1,1)	6.366 (1,1)	6.149 (1,0)	6.129 (0,0)	7.764 (1,1)	6.460 (1,1)
8	7.837 (1,1)	6.072 (1,1)	5.963 (1,0)	5.903 (0,0)	8.097 (1,1)	6.210 (1,1)

TABLE VII  
SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR MOTORCYCLE NOISE ( $\text{SNR}_{\text{in}} = 0$  dB)

BN	State Augmentation			Measurement Diff.		
	Ideal	Actual		Ideal	Actual	
1	7.889 (4,4)	5.459 (4,4)	4.806 (4,0)	4.565 (0,0)	9.523 (4,4)	5.899 (4,4)
2	8.515 (1,1)	7.395 (1,1)	7.040 (1,0)	6.966 (0,0)	9.201 (1,1)	7.435 (1,1)
4	9.473 (1,1)	8.067 (1,1)	7.868 (1,0)	7.858 (0,0)	9.749 (1,1)	8.108 (1,1)
8	9.855 (1,1)	7.971 (1,1)	7.838 (1,0)	7.805 (0,0)	10.08 (1,1)	8.127 (1,1)

outperformed fullband processing substantially, and that the best performance appeared in four-band decomposition. Note that for (0, 0) modeling, the SNR improvement was close to other cases. For example, it was 0.3 dB lower than that of (1, 1) modeling in four-band decomposition. But, the computational complexity required for (0, 0) modeling was much lower. Tables VII and VIII show the filtering results for motorcycle noise. As we can see, subband processing performed even better than fullband processing. The SNR improvement for (1, 1) modeling in four-band decomposition was 2.6 dB higher than that for (4, 4) modeling in fullband when  $\text{SNR}_{\text{in}} = 0$  dB. For (0, 0) modeling, the performance was still very close to that of (1, 1). For ideal cases, the SNR improvement with four-band decomposition was 1.3–1.6 dB higher than actual. We also observed that the ideal SNR improvement achieved by the measurement-difference type in fullband was much higher than that achieved by the state-augmentation type. This is due to the fact that the bandwidths of speech signals are narrower in fullband, which favors the smoothing results. When the number of bands is increased, this effect is reduced.

The aliasing effect in (25) was also studied. The length of the prototype filter was increased from 16 to 64, making the QMF transition band much sharper, and as a consequence, yielding a much smaller aliasing effect. Motorcycle-noise simulations were performed again, and the results may be

TABLE VIII  
SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR MOTORCYCLE NOISE ( $\text{SNR}_{\text{in}} = 5$  dB)

BN	State Augmentation				Measurement Diff.	
	Ideal	Actual		Ideal	Actual	
1	5.936 (4,4)	4.130 (4,4)	3.534 (4,0)	3.250 (0,0)	7.554 (4,4)	4.764 (4,4)
2	6.683 (1,1)	5.597 (1,1)	5.282 (1,0)	5.155 (0,0)	7.353 (1,1)	5.628 (1,1)
4	7.484 (1,1)	6.187 (1,1)	6.040 (1,0)	6.011 (0,0)	7.770 (1,1)	6.268 (1,1)
8	7.825 (1,1)	6.041 (1,1)	5.950 (1,0)	5.915 (0,0)	8.016 (1,1)	6.072 (1,1)

TABLE IX  
SEGMENTAL SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR MOTORCYCLE NOISE ( $\text{SNR}_{\text{in}} = 0$  dB)

BN	State Augmentation				Measurement Diff.	
	Ideal	Actual		Ideal	Actual	
1	13.05 (4,4)	10.49 (4,4)	10.02 (4,0)	9.908 (0,0)	14.02 (4,4)	9.260 (4,4)
2	12.99 (1,1)	11.09 (1,1)	10.71 (1,0)	10.74 (0,0)	13.77 (1,1)	9.847 (1,1)
4	13.48 (1,1)	10.95 (1,1)	10.57 (1,0)	10.73 (0,0)	13.82 (1,1)	10.40 (1,1)
8	13.54 (1,1)	10.38 (1,1)	10.28 (1,0)	10.23 (0,0)	13.72 (1,1)	10.18 (1,1)

TABLE X  
SEGMENTAL SNR IMPROVEMENT OF SPEECH ENHANCEMENT  
FOR MOTORCYCLE NOISE ( $\text{SNR}_{\text{in}} = 5$  dB)

BN	State Augmentation				Measurement Diff.	
	Ideal	Actual		Ideal	Actual	
1	10.03 (4,4)	7.990 (4,4)	7.487 (4,0)	7.310 (0,0)	11.12 (4,4)	7.404 (4,4)
2	10.05 (1,1)	8.369 (1,1)	8.079 (1,0)	8.015 (0,0)	10.80 (1,1)	7.808 (1,1)
4	10.56 (1,1)	8.456 (1,1)	8.100 (1,0)	8.220 (0,0)	10.89 (1,1)	8.217 (1,1)
8	10.65 (1,1)	7.895 (1,1)	7.805 (1,0)	7.756 (0,0)	10.85 (1,1)	7.901 (1,1)

summarized as follows. When the input SNR was 0 dB (5 dB), averaged SNR improvements were increased by 0.15 (0.18), 0.16 (0.21), and 0.16 (0.24) dB for 2-band, 4-band, and 8-band decomposition, respectively. Note that the SNR improvement resulting from the 5 dB input was higher than that for 0 dB. We also found that the greater the number of band decompositions, the greater the improvement we could obtain. Subjective listening tests indicated no perceptible differences from previous results. We thus conclude that the aliasing effect was indeed smaller for the sharp-transition QMF band. However, at low SNR's, the aliasing effect tends to be masked by other effects.

We then evaluated the SSNR filtering results. We used a segment size of 120 and did not take segments around silence into account. This enabled us to observe the filtering behavior during high energy periods. Tables IX and X show the averaged SSNR improvement for the motorcycle noise. Unlike the results in Tables VII and VIII, the gain due to subband processing was not that significant (below 1.2 dB). As we can see, the best results still appeared in four-band decomposition. Note that there were two design parameters ( $\alpha$  and  $\beta$ ) in the measurement-difference type but only one ( $\beta$ ) in the state-augmentation type. The speech signal during high energy periods is usually highly nonstationary. For such periods, filtering with fixed parameter values is not optimal. In other words, during these periods, AR parameters are more difficult to identify. This effect explains the following results. 1) For the ideal cases, the SSNR improvement of the measurement-difference type was higher than that of the state-augmentation type. However, for the actual cases, the SSNR improvement of the measurement-difference type was lower. 2) The difference in SSNR improvement between the ideal and actual cases for four-band decomposition was 2.1–3.4 dB,

which is larger than that for SNR improvement. As to (0, 0) modeling, the SSNR improvement remained close to other cases. We can thus conclude that if computational complexity is the main implementation concern, (0, 0) modeling with subband processing can be used, and that the performance degradation will be small. For other types of noise, the results are similar and details have been omitted.

## VI. CONCLUSIONS

In this paper, we have proposed techniques for speech enhancement in the subband domain. We first split noisy speech into subband signals using a QMF bank. We then modeled the subband signals as AR processes and applied Kalman filters to perform enhancement. For ease of implementation, we only considered AR(1) and AR(0) for subband signals. We used a prediction-error filter with the LMS algorithm for AR coefficient identification and proposed new methods for estimating driving-noise variance. Due to its inherent characteristics, the prediction-error filter converges to biased solutions in noisy environments. Performance of the Kalman filter with biased parameters was also analyzed. We found that when the AR coefficients were biased, we could apply an optimal driving-noise variance to minimize the output MSE. Simulations showed that the MSE yielded by our estimates of driving-noise variance were close to the optimal. We also found through analysis that MSE's for biased parameters do not deviate much from the optimal MSE's with true parameters. This indicates that accurate estimates of AR coefficients are not required provided driving-noise variances are properly estimated. This justifies the use of the prediction-error filtering. Finally, we found that speech enhancement in the subband domain not only had much lower computational complexity, but also gave higher SNR improvement than that in the fullband domain.

## REFERENCES

- [1] J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 471–472, Oct. 1978.
- [2] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 113–120, Oct. 1979.
- [3] R. J. McAulay and M. L. Malpass, "Speech enhancement using soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 137–145, Apr. 1980.
- [4] Y. Ephraim and D. Malah, "Speech enhancement using minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.
- [5] 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.
- [6] J. H. L. Hansen and M. A. Clement, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Trans. Signal Processing*, vol. 39, pp. 795–805, Apr. 1991.
- [7] T. V. Sreenivas and P. Kirnapure, "Codebook constrained Wiener filtering for speech enhancement," *IEEE Trans. Speech, Audio Processing*, vol. 4, pp. 383–389, Sept. 1996.
- [8] Y. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," *IEEE Trans. Signal Processing*, vol. 39, pp. 1943–1954, Sept. 1991.
- [9] J. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 59–71, Jan. 1995.

- [10] Y. Ephraim and H. L. van Tree, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, July 1995.
- [11] S. H. Jensen, P. H. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 439–448, Nov. 1995.
- [12] Y. Ephraim, "A Bayesian estimation approach for speech enhancement using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 40, pp. 725–735, Apr. 1992.
- [13] K. Y. Lee and K. Shirai, "Efficient recursive estimation for speech enhancement in color noise," *IEEE Signal Processing Lett.*, vol. 3, pp. 196–199, July 1996.
- [14] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 177–180, Apr. 1987.
- [15] J. D. Gibson, B. Koo, and S. D. Grey, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1732–1741, Aug. 1991.
- [16] B. Lee, K. Y. Lee, and S. Ann, "An EM-base approach for parameter enhancement with an application to speech signals," *Signal Process.*, vol. 46, no. 1, pp. 1–14, Sept. 1995.
- [17] M. Niedźwiecki and K. Cisowski, "Adaptive scheme for elimination of broadband noise and impulsive disturbance from AR and ARMA signals," *IEEE Trans. Signal Processing*, vol. 44, pp. 528–537, Mar. 1996.
- [18] V. Somayajulu, S. Mitra, and J. Shynk, "Adaptive line enhancement using multirate techniques," in *Proc. ICASSP*, May 1989, pp. 928–931.
- [19] X. Gillore and M. Vetterli, "Adaptive filtering in subbands with critical sampling: Analysis, experiments and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1862–1874, Aug. 1992.
- [20] H. Kiya and S. Yamaguchi, "Frequency sampling filter bank for adaptive system identification," in *Proc. ICASSP*, vol. 4, Mar. 1992, pp. 261–264.
- [21] U. Iyer, M. Nayeri, and H. Ochi, "Polyphase based adaptive structure for adaptive filtering and tracking," *IEEE Trans. Circuits Syst. II*, vol. 43, pp. 220–232, Mar. 1996.
- [22] O. Tanrikulu, B. Baykal, A. G. Constantinides, and J. A. Chambers, "Residual echo signal in critically sampled subband acoustic echo cancellers based on IIR and FIR filter banks," *IEEE Trans. Signal Processing*, vol. 45, pp. 901–911, Apr. 1997.
- [23] P. L. Vaidyanathan, *Multirate Systems and Filter Bank*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [24] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [25] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [26] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [27] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.



**Wen-Rong Wu** (S'87–M'89) was born in Taiwan, R.O.C., in 1958. He received the B.S. degree in mechanical engineering from Tatung Institute of Technology, Taiwan, in 1980, the M.S. degrees in mechanical and electrical engineering, and the Ph.D. degree in electrical engineering from State University of New York at Buffalo in 1985, 1986, and 1989, respectively.

Since August 1989, he has been a faculty member in the Department of Communication Engineering, National Chiao Tung University, Hsin-Chu, Taiwan.

His research interests include statistical signal processing and digital communications.



**Po-Cheng Chen** was born in Taiwan, R.O.C., in 1968. He received the B.S. and M.S. degrees in electrical engineering from Tatung Institute of Technology, Taiwan, in 1990, and 1992. Since 1992, he has been a Ph.D. student in the Department of Communication Engineering in National Chiao Tung University, Hsin-Chu, Taiwan. His research interests include signal processing and adaptive filtering.