# A Dynamic System Approach to Speech Enhancement Using the $H_\infty$ Filtering Algorithm

Xuemin Shen, *Member, IEEE,* and Li Deng, *Senior Member, IEEE*

***Abstract*— This paper presents a new approach to speech enhancement based on the $H_\infty$ filtering. This approach differs from the traditional modified Wiener/Kalman filtering approach in the following two aspects: 1) no *a priori* knowledge of the noise source statistics is required, the only assumption made is that noise signals have a finite energy; 2) the estimation criterion for the filter design is to minimize the worst possible amplification of the estimation error signals in terms of the modeling errors and additive noises. Since most additive noises in speech are non-Gaussian, this estimation approach is highly robust and more appropriate in practical speech enhancement. The proposed approach is straightforward to implement, as detailed in this paper. Experimental results show consistently superior enhancement performance of the $H_\infty$ filtering algorithm over the Kalman filtering counterpart, measured by the global signal-to-noise ratio (SNR). Examination of the spectrogram displays for the enhanced speech shows that the $H_\infty$ filtering approach tends to be more effective where the assumptions on the noise statistics are less valid.**

***Index Terms*—$H_\infty$ filtering, speech enhancement.**

## I. Introduction

NOISE contaminated speech results in various degrees of reduction of speech discrimination. For example, background acoustic noise degrades speech signal quality of mobile telephone systems; airplane engine noise affects the conversation between a pilot and an air traffic controller. With the objective of enhancing the quality and intelligibility of speech, speech enhancement involves manipulation of the contaminated speech signal to mitigate noise effects. There have been numerous studies on this subject [1]–[5]. Based on stochastic speech models, the previous studies have focused on the minimization of the variance of the estimation errors of speech signals, i.e., the celebrated Wiener and/or Kalman filtering approach. With suitable assumptions on noise variances, the Kalman filtering has certain desirable optimality properties, namely, it minimizes the expected estimation error energy and yields maximum-likelihood estimates. The robustness of the Kalman filter in various situations where the statistics are not completely known has been studied by many researchers. However, the question is what the performance of such an estimator will be if the assumptions on the statistics of noise are violated or if there are modeling errors in speech model? In other words, is it possible that small noise and modeling errors may lead to large estimation errors? In this paper, a new approach based on the $H_\infty$ filtering is presented for speech enhancement. This approach differs from the traditional modified Wiener/Kalman filtering approach in the following two aspects. 1) No *a priori* knowledge of the noise source statistics is required. The only assumption is that the noise signals have a finite energy. 2) The estimation criterion in the $H_\infty$ filter design is to minimize the worst possible effects of the disturbances (modeling errors and additive noises) on the signal estimation errors. This will guarantee that if the disturbances are small (in energy), then the estimation errors will be as small as possible (in energy). These two aspects make the $H_\infty$ filtering approach to be more appropriate in practical speech enhancement where there is significant uncertainty in the statistics of noises and speech signal systems. The implementation of the $H_\infty$ filtering algorithm is straightforward. Our experimental results have shown that the filtering performance of the $H_\infty$ estimation has noticeably superior to that of the Kalman estimation. The remainder of this paper is organized as follows. In Section II, the speech source model/vocal tract is characterized by an all-pole filter. Such a speech source model and an observation model (taking into account the additive noise) are then combined to create a canonical state-space model. Section III presents the $H_\infty$ filtering algorithm for estimating speech signal from noisy speech. Since the filter algorithm needs the knowledge of tap-gain parameters of the all-pole filter, an identification algorithm based on the $H_\infty$ filtering theory is introduced in Section IV. In Section V, the performance of the $H_\infty$ filter for speech enhancement is evaluated. The performance is analyzed for both stationary and nonstationary noise, based on the following criteria: 1) the global signal-to-noise ratio (SNR), and 2) the speech spectrogram representation for the enhanced signal. Conclusions of this work are given in Section VI

## II. Problem Formulation

Short segments of speech can be represented by the response of an all pole filter which models the vocal tract [1]. The filter is excited by a pulse train separated by the pitch period for voice sounds, or pseudorandom noise for unvoiced sounds. Thus the speech $x_k$ within a segment (clean speech) is assumed to satisfy a difference equation of the form

$$x_k = \sum_{j=1}^{n} a_j x_{k-j} + w_k \tag{1}$$

Manuscript received August 16, 1996; revised November 20, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Picone.

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1 (e-mail: xshen@bbcr.uwaterloo.ca; ldeng@crg4.uwaterloo.ca).

Publisher Item Identifier S 1063-6676(99)04627-1.

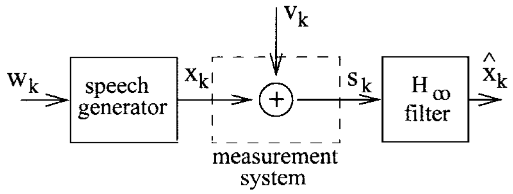


Fig. 1. Noisy speech generating and filtering mechanism.

where $n$ is the number of modeled poles, $a_j^{\prime}s$ are the tap-gain parameters characterizing the filter and $w_k$ is an excitation. If the speech signal $x_k$ is corrupted with background noise signal $v_k$, the observed (measured) noisy speech signal $s_k$ is described as follows:

$$s_k = x_k + v_k. \tag{2}$$

The speech generating mechanism is illustrated in Fig. 1.

Equations (1) and (2) can be represented by the following state-space model

$$X_k = AX_{k-1} + Bw_k \quad \text{(state equation)} \tag{3}$$

$$s_k = CX_k + v_k \quad \text{(measurement equation)} \tag{4}$$

where

$$X_k = \left[x_{k-n+1}^T x_{k-n+2}^T \cdots x_k^T\right]^T$$

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_n & a_{n-1} & a_{n-2} & \cdots & a_2 & a_1 \end{pmatrix}$$

$$B^T = C = [0 \quad 0 \quad \cdots \quad 0 \quad 1]_{1\times n}.$$

Since the noise contaminated speech results in various degrees of reduction of speech discrimination, one needs to enhance the quality and intelligibility of speech from the noisy speech, i.e., to estimate $x_k$ (the last component of $X_k$) given $\{s_i\}$, $i \le k$. The current enhancement algorithms have assumed that both $w_k$ and $v_k$ are white or color Gaussian processes [1]–[5]. However, neither the speech nor the noises may be Gaussian. This is because $w_k$ could be a pulse train for voiced speech, random noise for unvoiced speech or the modeling error, $v_k$ could be any kind of noise. The Gaussian assumptions may provide an estimate which is highly vulnerable to statistical outliers, i.e., a small number of large measurement errors would have a large influence on the resulting estimate, so that the viability of the algorithms has to be checked by experiment [3]. In the following, we present an new approach based on the $H_\infty$ filtering algorithm for speech enhancement, where both $w_k$ and $v_k$ are not necessary to be white or colored Gaussian processes. For comparison, the Kalman filtering algorithm is briefly reviewed.

## III. Kalman and $H_\infty$ Filtering Algorithms

### A. Kalman Filtering Algorithm

In the Kalman filtering, the clean speech signal $\{x_k\}$ is considered to be a random process. Assuming that both exciting term $w_k$ and observation noise $v_k$ (additive noise) are white Gaussian processes with zero mean and uncorrelated variances $Q$ and $R$

$$E\{w_k w_k^T\} = W$$
$$E\{v_k v_k^T\} = V$$
$$E\{w_k v_j^T\} = 0.$$

$E$ means *expectation*. The design objective of Kalman filter is to determine the optimal estimate $\hat{x}_k$ based on the $\{s_i\}$ $(0 \le i \le k)$ such that

$$P_k = E\{e_k e_k^T\} \tag{5}$$

is minimum. The estimation error $e_k$ is defined by the equation

$$e_k = x_k - \hat{x}_k. \tag{6}$$

For the state-space model (3)–(4), the Kalman filtering algorithm is given by

$$\hat{X}_k = A\hat{X}_{k-1} + K_k[s_k - CA\hat{X}_{k-1}] \tag{7}$$

with the initial condition $\hat{X}_0 = [0]_{n\times 1}$. The filter gain and error variance equations are

$$K_k = P_{k|k-1}C[V + CP_{k|k-1}C^T]^{-1} \tag{8}$$

$$P_{k|k-1} = AP_{k-1}A^T + BWB^T \tag{9}$$

$$P_k = [I - K_k C]P_{k|k-1} \tag{10}$$

where $K_k$ is a Kalman gain vector, $P_{k|k-1} = E[(X_k - \hat{X}_{k|k-1})^T(X_k - \hat{X}_{k|k-1})]$ is an *a priori* error covariance matrix, $P_k = E[(X_k - \hat{X}_k)^T(X_k - \hat{X}_k)$ is an *a posteriori* error covariance matrix. The initial condition $P_0 = [0]_{n\times n}$. $I$ is an $n\times n$ identity matrix. The estimated speech sample $\hat{x}_k$ can be obtained by

$$\hat{x}_k = C\hat{X}_k. \tag{11}$$

If the additive noise $\{v_k\}$ is a colored Gaussian process, the Kalman filter algorithm for such speech estimation is given in [3].

### B. $H_\infty$ Filtering Algorithm

Consider the state space model (3)–(4). We make no assumption on the nature of unknown quantities $w_k$ and $v_k$, and are interested not necessarily in the estimation of $X_k$ but in the estimation of some arbitrary linear combination of $X_k$ using the observations $\{s_i, i \le k\}$, i.e.,

$$z_k = LX_k \tag{12}$$

where $L \in \mathcal{R}^{1\times n}$. Different from that of the modified Wiener/Kalman filter which minimizes the variance of the estimation error, the design criterion of the $H_\infty$ filter is to provide a uniformly small estimation error, $e_k = z_k - \hat{z}_k$, for any $w_k$, $v_k \in l_2$ and $X_0 \in \mathcal{R}^n$. The measure of performance is then given by

$$J = \frac{\sum_{k=0}^{N} |z_k - \hat{z}_k|_Q^2}{|X_0 - \hat{X}_0|^2_{P_0^{-1}} + \sum_{k=0}^{N}\left\{|w_k|^2_{W^{-1}} + |v_k|^2_{V^{-1}}\right\}} \tag{13}$$

where $((X_0 - \hat{X}_0), w_k, v_k) \neq 0$, $\hat{X}_0$ is an *a priori* estimate of $X_0$ and $(X_0 - \hat{X}_0)$ represents unknown initial condition error, $Q \geq 0$, $p_0^{-1} > 0$, $W > 0$ and $V > 0$ are the weighting matrices. $p_0^{-1} > 0$ denotes a positive definite matrix that reflects *a priori* knowledge on how close the initial guess $\hat{X}_0$ is to $X_0$. The notation $|z_k|_Q^2$ is defined as the square of the weighted (by $Q$) $L_2$ norm of $z_k$, i.e., $|z_k|_Q^2 = z_k^T Q z_k$. The $H_\infty$ filter will search $\hat{z}_k$ such that the optimal estimate of $z_k$ among all possible $\hat{z}_k$ (i.e., the worse-case performance measure) should satisfy

$$\sup J \leq \gamma^2 \tag{14}$$

where "sup" stands for supremum and $\gamma > 0$ is a prescribed level of noise attenuation. The matrices $Q$, $W$, $V$ and $p_0$ are left to the choice of the designer and depend on performance requirements. The above problem formulation shows that $H_\infty$ optimal estimators guarantee the smallest estimation error energy over all possible disturbances of finite energy. They are, therefore, overly conservative, which results in a better robust behavior to disturbance variations. The discrete $H_\infty$ filtering can be interpreted as a *minimax* problem where the estimator strategy $\hat{z}_k$ plays against the exogenous inputs $w_k, v_k$ and the uncertainty of the initial state $X_0$, so the performance criterion is equivalent to

$$\begin{aligned}\min_{\hat{z}_k} \max_{(v_k, w_k, X_0)} J = -\frac{1}{2}\gamma^2 |X_0 - \hat{X}_0|_{p_0^{-1}}^2 + \frac{1}{2}\sum_{k=0}^{N} \big[|z_k - \hat{z}_k|_Q^2 \\ - \gamma^2\big(|w_k|_{W^{-1}}^2 + |v_k|_{V^{-1}}^2\big)\big]\end{aligned} \tag{15}$$

where "min" stands for minimization and "max" maximization. Note that unlike the traditional minimum variance filtering approach (Wiener and/or Kalman filtering), the $H_\infty$ filtering deals with deterministic disturbances and no *a priori* knowledge of the noise statistics is required. Since the observation $s_k$ is given, $v_k$ can be uniquely determined by (2) once the optimal values of $w_k$ and $X_0$ are found. Using $z_k = LX_k$, $\hat{z}_k = L\hat{X}_k$, we can rewrite the performance criterion (15) as

$$\begin{aligned}\min_{\hat{X}_k} \max_{(s_k, w_k, X_0)} J = -\frac{1}{2}\gamma^2 |X_0 - \hat{X}_0|_{p_0^{-1}}^2 + \frac{1}{2}\sum_{k=0}^{N} \big[|X_k - \hat{X}_k|_{\bar{Q}}^2 \\ - \gamma^2\big(|w_k|_{W^{-1}}^2 + |s_k - CX_k|_{V^{-1}}^2\big)\big]\end{aligned} \tag{16}$$

where $\bar{Q} = L^T QL$.

Extensive research work for $H_\infty$ filter design has been done in the past years [6]–[15]. The following theorem presents a complete solution to the $H_\infty$ estimation problem for the state-space model (3)–(4) with the performance criterion (16).

*Theorem:* Let $\gamma > 0$ be a prescribed level of noise attenuation. Then, there exists an $H_\infty$ filter for $X_k$ if and only if there exists a stabilizing symmetric solution $P_k > 0$ to the following discrete-time Riccati type equation

$$\begin{aligned}P_{k+1} &= AP_k(I - \gamma^{-2}\bar{Q}P_k + C^T V^{-1} CP_k)^{-1} A^T + BWB^T \\ P_0 &= p_0.\end{aligned} \tag{17}$$

If this is the case, then an $H_\infty$ filter can be given by

$$\hat{z}_k = L\hat{X}_k, \quad k = 1, 2, \cdots, N \tag{18}$$

TABLE I

PERFORMANCE COMPARISON OF KALMAN AND $H_\infty$ FILTERING ALGORITHMS

| Filtering Algorithm | Input SNR (dB) | Output SNR (dB) | |
|---|---|---|---|
| | | White Noise | Helicopter Noise |
| Kalman | 0 | 5.6525 | 6.1688 |
| $H_\infty$ | 0 | 6.1930 | 6.6874 |
| Kalman | 5 | 8.7276 | 8.9119 |
| $H_\infty$ | 5 | 9.8781 | 10.0693 |
| Kalman | 10 | 12.2059 | 12.4074 |
| $H_\infty$ | 10 | 13.3735 | 13.5256 |

where

$$\hat{X}_k = A\hat{X}_{k-1} + H_k(s_k - CA\hat{X}_{k-1}), \quad \hat{X}_0 = 0. \tag{19}$$

$H_k$ is the gain of the $H_\infty$ filter and is given by

$$H_k = AP_k(I - \gamma^{-2}\bar{Q}P_k + C^T V^{-1} CP_k)^{-1} C^T V^{-1}. \tag{20}$$

The proof of the theorem is given in the Appendix. Solving Riccati equation (17) for the solution $P_k$ is not trivial due to its nonlinearity. Let $P_k^{-1} = R_k^{-1} + \gamma^{-2}\bar{Q}$, applying the following matrix inversion lemma (MIL)

$$A - AB(C + B^T AB)^{-1} B^T A = (A^{-1} + BC^{-1} B^T)^{-1} \tag{21}$$

equation (17) can be rewritten as

$$\begin{aligned}R_{k+1}^{-1} &= \big[A\big(R_k^{-1} + C^T V^{-1} C\big)^{-1} A^T + BWB^T\big]^{-1} - \gamma^{-2}\bar{Q} \\ R_0 &= \big(p_0^{-1} + \gamma^{-2}\bar{Q}\big)^{-1}, \quad k = 0, 1, 2, \cdots, N\end{aligned} \tag{22}$$

so that we can obtain $P_k$ from (23) recursively.

It should be mentioned that the structure of the $H_\infty$ filter depends, via the Riccati type equation (17), on the linear combination of the states that we intend to estimate $(LX_k)$, and on the weighting matrices $(W, V)$ of the noises and $p_0$ of the initial condition in the performance criterion. In other words, the designer can choose weighting matrices based on the performance requirements. Since the $H_\infty$ filter is designed based on an upper bound of the estimation error, it is more robust.

Comparing the Kalman filtering algorithm (7)–(10) and the $H_\infty$ filtering algorithm (17)–(20), we can observe the following.

1) The Kalman filtering algorithm gives the minimum mean-square-error estimate of the state vector $X_k$ based on the $\{s_i\}$ $(0 \leq i \leq k)$, independent of $L$.
2) The $H_\infty$ filtering algorithm gives the optimal estimate of $LX_k$ based on the $\{s_i\}$ $(0 \leq i \leq k)$ such that the effect of the worst disturbances (noises) on the estimation error is minimized.
3) Kalman and $H_\infty$ filters have similar observer structure. Let weighting matrices $(W, V)$ and $p_0$ of $H_\infty$ filter be same as the variances $(W, V)$ and $P_0$ of Kalman filter. In the limiting case, where the parameter $\gamma \to \infty$, the $H_\infty$ reduces to a Kalman filter.

It is interesting to note that if we choose $L = [1 \ 0 \ \cdots \ 0]_{1\times n}$, the $H_\infty$ filter is designed to minimize the worst possible amplification of the estimation error of the first component of the state vector $X_k$, i.e. $x_{k-n+1}$ in terms of all

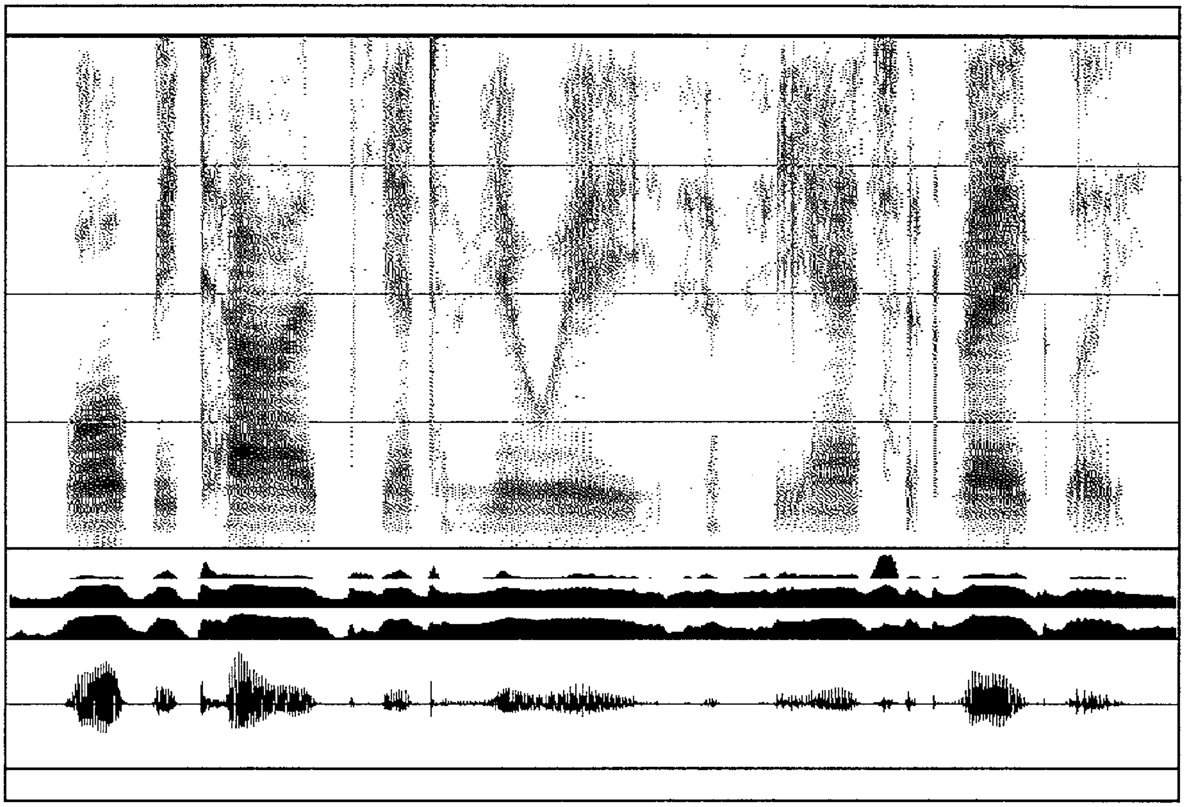


Fig. 2. Spectrogram of clean speech.

exogenous inputs. The estimate $\hat{x}_{k-n+1}$ should give a better estimation of speech signal $x_{k-n+1}$ at the $k$th instant since the estimation is based on the $\{s_i\}$, $0 \leq i \leq k$. This estimation is equivalent to the fixed-lag smoothing problem. The only difference from the traditional fixed-lag smoothing problem is that no additional computation is required in this case.

## IV. Tap-Gain Parameter Estimation

The filtering algorithm for speech enhancement in Section III requires the knowledge of tap-gain parameter vector $a = [a_1 \quad \cdots \quad a_n]^T$ of the all-pole filter. Estimation of the parameter vector from noisy speech has been a long standing research problem with most efforts being focused on an white/color Gaussian noise processes [3]. However, in our case the estimation of the tap-gain parameter vector can not be performed by Wiener or Kalman filtering algorithm since the statistics of both noise excitation $w_k$ and measurement noise $v_k$ are not known. In other words, both $w_k$ and $v_k$ can be non-Gaussian. Here we apply the $H_\infty$ filtering algorithm to identify/estimate the source model parameter vector. To express the model of (1) and (2) in a more suitable form for application of the identification method, we introduce the shifting operator $z$ defined by

$$z^{-1}s_k = s_{k-1} \tag{23}$$

so that (2) can be written as

$$A(z^{-1})s_k = u_k \tag{24}$$

where

$$A(z^{-1}) = 1 - a_1 z^{-1} - \cdots - a_p z^{-p}$$
$$u_k = w_k + A(z^{-1})v_k.$$

Let

$$\alpha_k^T = [s_{k-1} \quad \cdots \quad s_{k-n}], \quad \theta^T = [a_1 \quad \cdots \quad a_n]. \tag{25}$$

Equation (23) becomes

$$s_k = \alpha_k^T \theta + u_k. \tag{26}$$

In speech enhancement, the noisy speech is usually divided into number of frames and the length of each frame is within 10 to 30 ms. In each frame interval, it is assumed the *AR* model (26) is time-invariant. In the parameter estimation problem, $\theta_k$ is updated recursively by equating $\theta_k$ to $\theta_{k+1}$, i.e.,

$$\theta_{k+1} = \theta_k \tag{27}$$
$$s_k = \alpha_k^T \theta + u_k. \tag{28}$$

The state model (27)–(28) can be identified with the similar estimation algorithm given in Section III. The $H_\infty$ identifier should be chosen for the worst possible $u_k$, i.e.

$$\min_{\hat{\theta}_k} \max_{(u_k, \theta_0)} J = -\frac{1}{2}\gamma_l^2 |\theta - \theta_0|^2_{\mu_0^{-1}} + \frac{1}{2}\sum_{k=0}^{N} \left[ |\theta_k - \hat{\theta}_k|^2 - \gamma_l^2 |u_k|^2 \right]. \tag{29}$$

Following the similar analysis given in Section III, the $H_\infty$ identification algorithm to compute the optimal $\theta_k$ can be obtained as

$$R_k = M_k \alpha_k \big(1 + \alpha_k^T M_k \alpha_k\big)^{-1} \tag{30}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + R_k \big(s_{k+1} - \alpha_k^T \hat{\theta}_k\big), \quad \hat{\theta}_0 = 0 \tag{31}$$

$$M_{k+1}^{-1} = M_k^{-1} + \alpha_k \alpha_k^T - \gamma_l^{-2} I, \quad M_0^{-1} = \mu_0^{-1} - \gamma_l^{-2} I. \tag{32}$$

Note that the above algorithm is not suitable for on-line recursive parameter estimation unless we can obtain an online $\gamma_l$ such that the matrix $M_k$ must be positive definite. In the following, we propose an algorithm for adaptively adjusting $\gamma_l$ value to its minimum at each iteration for $M_k$. In order for $M_{k+1}$ to be positive definite, it requires

$$\begin{aligned} &M_k^{-1} + \alpha_k \alpha_k^T - \gamma_{l_{k+1}}^{-2} I > 0 \\ \Longrightarrow &\gamma_{l\,k+1} > \max\big\{\mathrm{eig}\big(M_k^{-1} + \alpha_k \alpha_k^T\big)^{-1}\big]\big\}^{0.5} \\ \Longrightarrow &\gamma_{l\,k+1} = \xi \max\big\{\mathrm{eig}\big(M_k^{-1} + \alpha_k \alpha_k^T\big)^{-1}\big]\big\}^{0.5} \end{aligned} \tag{33}$$

where $\max\{\mathrm{eig}(A)\}$ indicates the maximum eigenvalue of the matrix $A$, and $\xi$ is a constant very close to one but larger than one to ensure that $\gamma_l$ is always greater than the minimum value but is very close to it.

The $H_\infty$ adaptive approach computes the speech model coefficients (3)–(4) from noisy speech, and these coefficients are then used in the $H_\infty$ filter. This has the advantage of adapting the coefficients over the utterance, at the cost of using coefficients calculated from corrupted speech [3].

## V. Speech Enhancement Experiments

The $H_\infty$ and Kalman filtering algorithms described in Sections III and IV are applied to speech enhancement by first dividing the noisy speech into equal-length segments. Within each segment, the parameter $\theta$ is first estimated according to (30)–(33) for both $H_\infty$ and Kalman filtering algorithms, and is then used to filter the noisy speech. The $H_\infty$ filter algorithm is initialized only for the first segment with all the remaining segments utilizing the filtering results obtained from the previous segments. In the experiments, we choose the initial state vector $\hat{X}_0 = 0$, and weight matrix $p_0 \gg 0$. In the subsequent segments, $\hat{X}_0$ and $p_0$ are initialized using the corresponding last values from the previous segment. There exists a tradeoff in the choice of the length of the segments. Large segments improve the accuracy of the prediction parameters for stationary sounds (e.g., vowels), but short segments improve the accuracy for nonstationary sounds. In our experiment, the segment length used for calculating the parameters $\{a_i, i = 1, 2, \cdots, n\}$ is set to be 128 samples, which corresponds to 16 ms (with a sampling frequence of 8 kHz). The order of the all-pole filter $n$ is ten, which is a commonly used value in linear predictive analysis of speech signal, and the order of state space model is set to be equal to the order of the all-pole filter. The input SNR varies from 0 dB to 15 dB. The parameter $\gamma$ is chosen to be 1.05. The expectation and maximization (EM) algorithm [17] is used to calculate $W$ and $V$, which are the weighting matrices for the $H_\infty$ filter and the variances of $w_k$ and $v_k$ for the Kalman filter, respectively. Two types of noise are used: white noise (stationary) and helicopter noise (nonstationary). The performance of both the $H_\infty$ filtering and Kalman filtering algorithms is measured in terms of SNR and speech spectrogram representation. Three sentences are tested and the outcomes are similar. The sentences are

- "Woe betide the interviewee if he answered vaguely;"
- "Drop five forms in the box before you go out;"
- "Sometime, he coincided with my father's being at home."

Experimental results obtained for the sentence "Woe betide the interviewee if he answered vaguely" embedded in noise are summarized in Table I.

The SNR values used to measure the enhancement performance are the global signal to noise ratios calculated by

$$SNR = 10 \log_{10} \frac{\sum_{k=1}^{N} x_k^2}{\sum_{k=1}^{N} [x_{k-n+1} - \hat{z}_k]^2} \tag{34}$$

where $N$ is the total number of samples of each sentence, $x_k$ is the clean (noise-free) sequence and $\hat{z}_k$ is the enhanced speech. The results of Table I consistantly show moderate performance advantage of the $H_\infty$ filtering algorithm (measured by the output SNR values) over the Kalman filtering algorithm, for both the white and helicopter noise. The performance gain is about 0.5 dB for the input SNR of 0 dB, which increases to slightly over 1 dB when the input SNR is increased to 5 and 10 dB. In order to examine the details of the speech enhancement results, both the waveforms and wideband spectrograms are plotted for the original clean speech signal (Fig. 2), speech embedded in the white noise (Fig. 3), enhanced speech by the Kalman filter (Fig. 4), and the enhanced speech by the $H_\infty$ filter (Fig. 5). By comparing the spectrogram plots of Figs. 4 and 5, it is noted that the $H_\infty$ filter tends to perform better than the Kalman filter in the relatively fast changing regions of the speech. This appears to be accounted for by the fact that in such regions with fast spectra changes, the assumption for the driving noise being white in (3) tends to be grossly invalid. Since the $H_\infty$ filtering approach makes no assumption about the noise statistics, it outperforms the Kalman filtering approach which is based on grossly inaccurate assumption about the statistics. It is also interesting to note that in the estimation of the noise statistics, the EM algorithm has been used based on the maximum-likelihood principle consistent with the Kalman filter formulation. The nevertheless inferior performance with the Kalman filter suggests that when the assumptions on the speech model are invalid, it is better to resort to an approach that is robust to the statistics of the speech model than to the one which attempts to accurately estimate the model parameters.

## VI. Conclusion

A new speech enhancement method based on the $H_\infty$ filtering has been developed. This method exploits a waveform-based speech production model without requiring detailed knowledge of noise statistics. Since the design criterion of the $H_\infty$ filtering algorithm is based on the worst case disturbances, the method is less sensitive to uncertainty in the exogenous

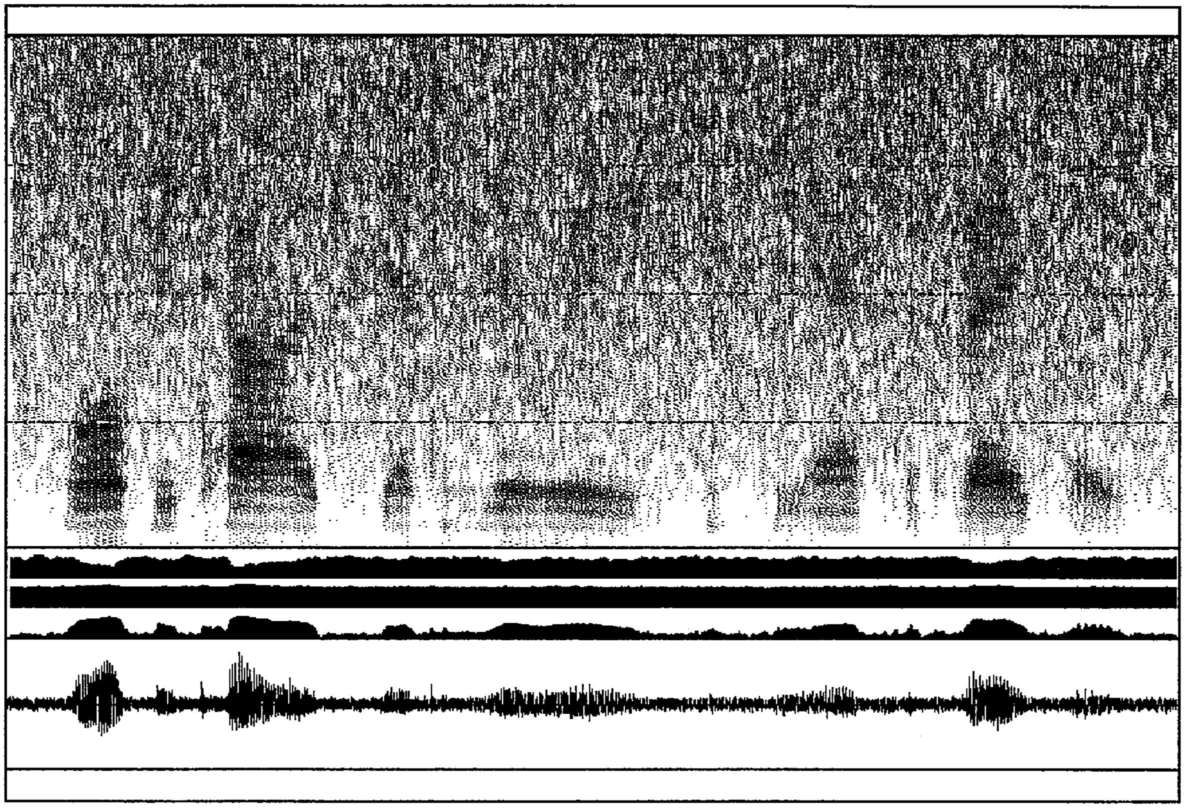


Fig. 3. Spectrogram of white noisy speech ($\mathrm{SNR} = 5$ dB).

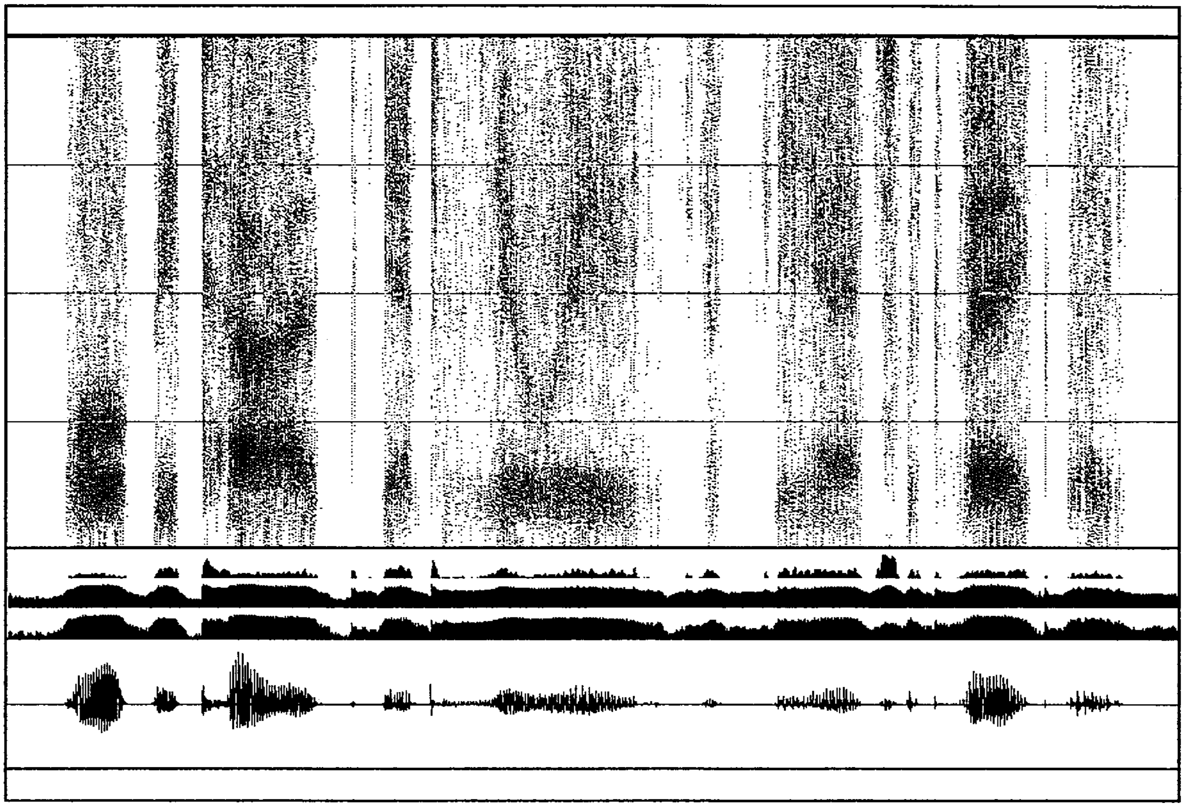


Fig. 4. Spectrogram of Kalman filtered speech.

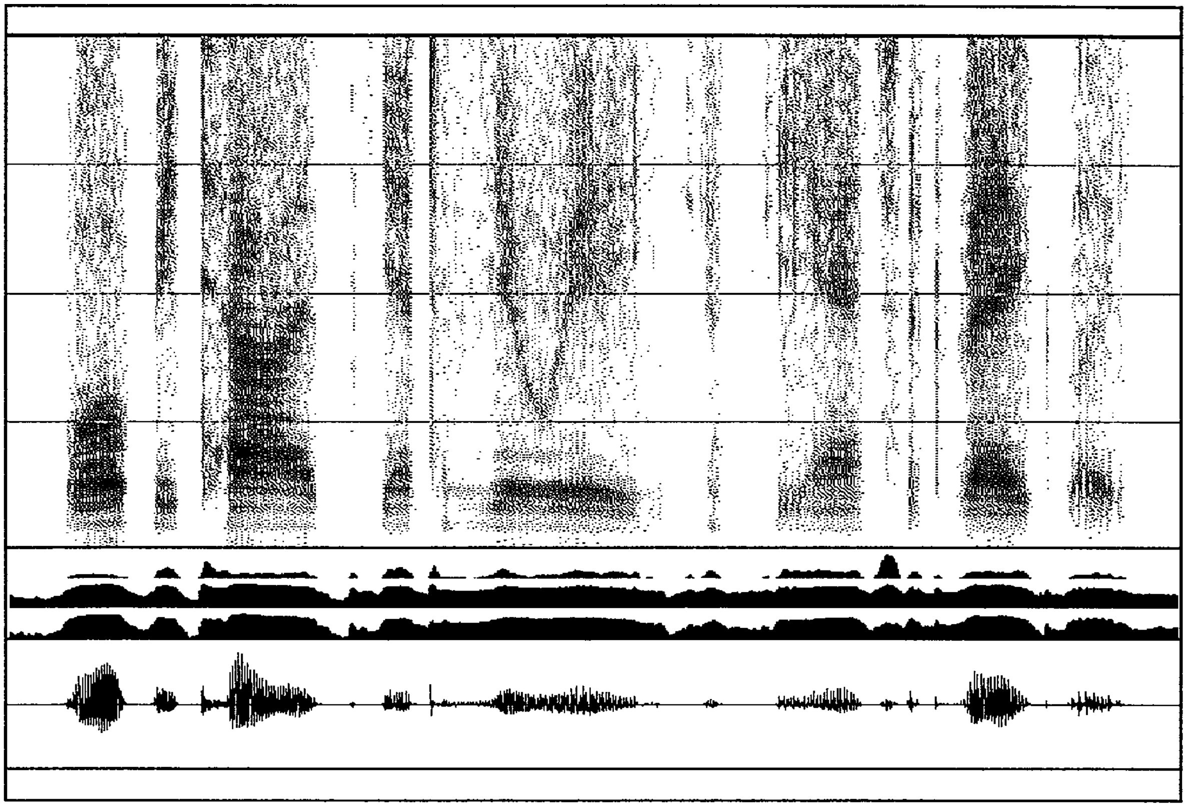


Fig. 5. Spectrogram of $H_\infty$ filtered speech.

signal statistics and system model dynamics. This theoretical advantage has been confirmed in the speech enhancement experiments where the global SNR is used to measure the performance. The experiments have shown consistently the superiority of the $H_\infty$ filtering approach over the Kalman filtering conterpart.

## Appendix
## Proof of Theorem

By using a set of Lagrange multipliers to adjoin the constraint (3)–(4) to the performance criterion (16), the resulting *Hamiltonian* is

$$\begin{aligned} M = \frac{1}{2}\big[|X_k - \hat{X}_k|^2_{\bar{Q}} - \gamma^2\big(|w_k|^2_{W^{-1}} + |s_k - CX_k|^2_{V^{-1}}\big)\big] \\ + \lambda_{k+1}^T\gamma^2[AX_k + Bw_k - X_{k+1}] \\ + [AX_k + Bw_k - X_{k+1}]^T\lambda_{k+1}\gamma^2. \qquad \text{(A.1)}\end{aligned}$$

Taking the first variation, the necessary conditions for a maximum are

$$X_0 = \hat{X}_0 + p_0\lambda_0, \quad \lambda_N = 0 \qquad \text{(A.2)}$$

$$w_k = WB^T\lambda_{k+1} \qquad \text{(A.3)}$$

$$\lambda_k = A^T\lambda_{k+1} + \gamma^{-2}\bar{Q}(X_k - \hat{X}_k) + C^TV^{-1}(s_k - CX_k) \qquad \text{(A.4)}$$

These first order necessary conditions result in a two point boundary value problem

$$\begin{pmatrix} X_{k+1} \\ \lambda_k \end{pmatrix} = \begin{pmatrix} A & BWB^T \\ \gamma^{-2}\bar{Q} - C^TV^{-1}C & A^T \end{pmatrix}\begin{pmatrix} X_k \\ \lambda_{k+1} \end{pmatrix} + \begin{pmatrix} 0 \\ -\gamma^{-2}\bar{Q}\hat{X}_k + C^TV^{-1}s_k \end{pmatrix} \qquad \text{(A.5)}$$

with boundary conditions

$$X_0 = \hat{X}_0 + p_0\lambda_0, \quad \lambda_N = 0. \qquad \text{(A.6)}$$

Since the two-point boundary value problem is linear, the solution is assumed to be of the form

$$X_k^* = \bar{X}_k + P_k\lambda_k^* \qquad \text{(A.7)}$$

where $\bar{X}_k$ and $P_k$ are undetermined variables. $X_k^*$ and $\lambda_k^*$ represent optimal value of $X_k$ and $\lambda_k$, respectively, for any fixed admissible functions of $\bar{X}_k$ and $s_k$. The optimal values for $w_k$ and $X_0$ are

$$w_k^* = WB^T\lambda_{k+1}^*, \quad X_0^* = \hat{X}_0 + p_0\lambda_0^*. \qquad \text{(A.8)}$$

Substituting (A.7) into (A.5) results in

$$\bar{X}_{k+1} + P_{k+1}\lambda_{k+1}^* = A\bar{X}_k + AP_k\lambda_k^* + BWB^T\lambda_{k+1}^* \qquad \text{(A.9)}$$

and

$$\begin{aligned}\lambda_k^* = (I - \gamma^{-2}\bar{Q}P_k + C^TV^{-1}CP_k)^{-1}[\gamma^{-2}\bar{Q}(\bar{X}_k - \hat{X}_k) \\ + C^TV^{-1}(s_k - C\bar{X}_k) + A^T\lambda_{k+1}^*]. \qquad \text{(A.10)}\end{aligned}$$

From (A.9)–(A.10) we have

$$\begin{aligned}\bar{X}_{k+1} &+ P_{k+1}\lambda_{k+1}^* \\ &= A\bar{X}_k + AP_k(I - \gamma^{-2}\bar{Q}P_k + C^TV^{-1}CP_k)^{-1} \\ &\quad \times [\gamma^{-2}\bar{Q}(\bar{X}_k - \hat{X}_k) + C^TV^{-1}(s_k - C\bar{X}_k) \\ &\quad + A^T\lambda_{k+1}^*] + BWB^T\lambda_{k+1}^* \end{aligned} \tag{A.11}$$

i.e.,

$$\begin{aligned}\bar{X}_{k+1} &- A\bar{X}_k - AP_k(I - \gamma^{-2}\bar{Q}P_k + C^TV^{-1}CP_k)^{-1} \\ &\times [\gamma^{-2}\bar{Q}(\bar{X}_k - \hat{X}_k) + C^TV^{-1}(s_k - C\bar{X}_k)] \\ &= [-P_{k+1} + AP_k(I - \gamma^{-2}\bar{Q}P_k \\ &\quad + C^TV^{-1}CP_k)^{-1}A^T + BWB^T]\lambda_{k+1}^*. \end{aligned} \tag{A.12}$$

For (A.12) to hold true for arbitrary $\lambda_k^*$, both sides are set identically to zero, resulting in

$$\begin{aligned}\bar{X}_{k+1} &= A\bar{X}_k + AP_k[(I - (\gamma^{-2}\bar{Q} - C^TV^{-1}C)P_k]^{-1} \\ &\quad \cdot [\gamma^{-2}\bar{Q}(\bar{X}_k - \hat{X}_k) + C^TV^{-1}(s_k - C\bar{X}_k)] \\ \bar{X}_0 &= \hat{X}_0 \end{aligned} \tag{A.13}$$

and

$$\begin{aligned}P_{k+1} &= AP_k(I - \gamma^{-2}\bar{Q}P_k + C^TV^{-1}CP_k)^{-1}A^T + BWB^T \\ P_0 &= p_0. \end{aligned} \tag{A.14}$$

Equation (A.14) is the well-known Riccati difference equation. It has been proofed that if the solution $P_k$ to the Riccati equation (A.14) exists $\forall k \in [0, N-1]$, then $P_k > 0$ $\forall k \in [0, N-1]$.

Now substituting the optimal strategies (A.8) into the performance (16), we obtain

$$\begin{aligned}\min_{\hat{X}_k}\max_{s_k} J = &-\frac{1}{2}\gamma^2|\lambda_0^*|_{P_0}^2 + \frac{1}{2}\sum_{k=0}^{N-1}\big[|\bar{X}_k + P_k\lambda_k^* - \hat{X}_k|_{\bar{Q}}^2 \\ &- \gamma^2\big(|WB^T\lambda_{k+1}^*|_{W^{-1}}^2 \\ &+ |s_k - C\bar{X}_k - CP_k\lambda_k^*|_{V^{-1}}^2\big)\big]. \end{aligned} \tag{A.15}$$

In the sequel we will perform the *min-max* optimization of $J$ with respect to $\hat{X}_k$ and $s_k$, respectively. Adding to (A.15) the identically zero term

$$\frac{1}{2}\gamma^2\Big[|\lambda_0^*|_{P_0}^2 - |\lambda_N^*|_{P_N}^2 + \sum_{k=0}^{N-1}\big(|\lambda_{k+1}^*|_{P_{k+1}}^2 - |\lambda_k^*|_{P_k}^2\big)\Big] = 0 \tag{A.16}$$

after lengthy algebra, results in the following *min-max* problem:

$$\min_{\hat{X}_k}\max_{s_k} J = \frac{1}{2}\sum_{k=0}^{N-1}\big[|\bar{X}_k - \hat{X}_k|_{\bar{Q}}^2 - \gamma^2|s_k - C\bar{X}_k|_{V^{-1}}^2\big] \tag{A.17}$$

subject to the dynamic constraints (A.13) and (A.14).

Let

$$r_k = \bar{X}_k - \hat{X}_k, \quad q_k = s_k - C\bar{X}_k. \tag{A.18}$$

Equation (A.17) becomes

$$\min_{r_k}\max_{q_k} J = \frac{1}{2}\sum_{k=0}^{N-1}\big[|r_k|_{\bar{Q}}^2 - \gamma^2|q_k|_{V^{-1}}^2\big]. \tag{A.19}$$

The two independent players $r_k$ and $q_k$ in (A.19) affect the variables $\bar{X}_k$, but $\bar{X}_k$ does not appear in the performance index, therefore the optimal strategies of $r_k$ and $q_k$ are

$$r_k^* = 0, \quad q_k^* = 0 \tag{A.20}$$

i.e.,

$$\bar{X}_k = \hat{X}_k^*, \quad s_k^* = C\bar{X}_k. \tag{A.21}$$

The value of the game is the value of the cost function (16). When the optimal strategies $\hat{X}_k^*$, $s_k^*$, $w_k^*$ and $x_0^*$ in (A.8) and (A.21) are substituted into the (16)

$$J(\hat{X}_k^*, s_k^*, w_k^*, X_0^*) = 0 \tag{A.22}$$

giving a zero value game.

So far, the strategies of $\hat{X}_k^*, s_k^*, w_k^*$ and $X_0^*$ have been assumed to be optimal, based on satisfying the necessary conditions for optimality. If the strategies can also satisfy a saddle-point inequality, they represent optimal strategies. A saddle point strategy can be obtained by solving two optimization problems

$$\min_{\hat{X}_k}\max_{s_k}\max_{w_k}\max_{X_0} J = J^* \tag{A.23}$$

$$\max_{s_k}\max_{w_k}\max_{X_0}\min_{\hat{X}_k} J = J_* \tag{A.24}$$

When $J^* = J_*$, the solutions to (A.23) and (A.24) produce saddle point strategies. It can be easily shown that if $P_k$ exists $\forall k \in [0, N-1]$, the optimal strategies $\hat{X}_k^*, s_k^*, w_k^*$ and $X_0^*$ satisfy a saddle point inequality

$$\begin{aligned}J(\hat{X}_k^*, s_k, w_k, X_0) &\le J(\hat{X}_k^*, s_k^*, w_k^*, X_0^*) \\ &\le J(\hat{X}_k, s_k^*, w_k^*, X_0^*). \end{aligned} \tag{A.25}$$

Note that the notation $J_1 \ge J_2$ means that $J_1 - J_2$ is positive semidefinite matrix.

The right inequality can be checked by adding the identically zero term

$$\begin{aligned}\frac{1}{2}\gamma^2\Bigg[&|X_0^* - \hat{X}_0|_{P_0^{-1}}^2 - |X_N^* - \hat{X}_N|_{P_N^{-1}}^2 \\ &+ \sum_{k=0}^{N-1}\big(|X_{k+1}^* - \hat{X}_{k+1}|_{P_{k+1}^{-1}}^2 - |X_k^* - \hat{X}_k|_{P_k^{-1}}^2\big)\Bigg] \end{aligned} \tag{A.26}$$

to $J(\hat{X}_k, s_k^*, w_k^*, X_0^*)$, and the left inequality can be checked by adding the identically zero term

$$\begin{aligned}\frac{1}{2}\gamma^2\Bigg[&|X_0 - \hat{X}_0^*|_{P_0^{-1}}^2 - |X_N - \hat{X}_N^*|_{P_N^{-1}}^2 \\ &+ \sum_{k=0}^{N-1}\big(|X_{k+1} - \hat{X}_{k+1}^*|_{P_{k+1}^{-1}}^2 - |X_k - \hat{X}_k^*|_{P_k^{-1}}^2\big)\Bigg] \end{aligned} \tag{A.27}$$

to $J(\hat{X}_k^*, s_k, w_k, X_0)$. The optimal strategy of the measurement noise can be obtained by

$$v_k^* = s_k^* - C\hat{X}_k^* = C\bar{X}_k - C\hat{X}_k^* = 0. \tag{A.28}$$

With (A.13) and (A.21), the optimal $H_\infty$ filter is given by

$$\hat{z}_k^* = L\hat{X}_k^*, \quad k = 0, 1, \cdots, N-1 \tag{A.29}$$

where

$$\hat{X}_{k+1}^* = A\hat{X}_k^* + K_k(s_k - C\hat{X}_k^*), \quad \bar{X}_0 = \hat{X}_0 \tag{A.30}$$

$$K_k = AP_k(I - \gamma^{-2}\bar{Q}P_k + C^TV^{-1}CP_k)^{-1}C^TV^{-1} \tag{A.31}$$

and $P_k$ is given by (17).

## Acknowledgment

The authors would like to thank the Natural Science and Engineering Research Council of Canada for the financial support of this work.

## References

[1] 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, 1978.

[2] K. K. Paliwal and A. Basu, "A speech enhancement method based on kalman filtering," in *Proc. IEEE ICASSP*, 1987, pp. 177–180.

[3] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1732–1742, 1991.

[4] Y. Ephraim, "A minimum mean square error approach for speech enhancement," in *Proc. IEEE ICASSP*, 1990, pp. 829–832.

[5] K. Y. Lee and K. Shirai, "Efficient recursive estimation for speech enhancement in colored noise," *IEEE Signal Processing Lett.*, vol. 3, pp. 196–199, 1996.

[6] R. N. Banavar and J. L. Speyer, "A linear quadratic game theory approach to estimation and smoothing," in *Proc. IEEE ACC*, 1991, pp. 2818–2822.

[7] C. E. de Souza, U. Shaked, and M. Fu, "Robust $H_\infty$ filtering with parametric uncertainty and deterministic signal," in *Proc. IEEE CDC'92*, pp. 2305–2310.

[8] M. J. Grimble and A. Elsayed, "Solution of the $H_\infty$ optimal linear filtering problem for discrete-time systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1092–1104, 1990.

[9] U. Shaked and Y. Theodor, "$H_\infty$-optimal estimation: A tutorial," in *Proc. 31st IEEE CDC*, 1992, pp. 2278–2286.

[10] K. M. Nagpal and P. P. Khargonekar, "Filtering and smoothing in an $H_\infty$ setting," *IEEE Trans. Automat. Contr.*, vol. AC-36, pp. 152–166, 1991.

[11] W. M. Haddad, D. S. Berstein, and D. Mustafa, "Mixed-norm $H_2/H_\infty$-regulation and estimation: The discrete-time case," *Syst. Contr. Lett.*, vol. 16, pp. 235–247, 1991.

[12] B. Hassibi and T. Kailath, "$H_\infty$ adaptive filtering," in *Proc. IEEE ICASSP'95*, Detroit, MI, 1995, pp. 949–952.

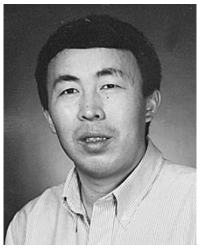
[13] X. Shen and L. Deng, "Discrete $H_\infty$ filter design with application to speech enhancement," in *Proc. IEEE ICASSP'95*, Detroit, MI, pp. 1504–1507.

[14] I. Yaesh and U. Shaked, "A transfer function approach to the problem of discrete-time systems: $H_\infty$-optimal linear control and filtering," *IEEE Tans. Automat. Contr.*, vol. 36, pp. 1264–1271, 1991.

[15] T. Basar, "Optimum performance levels for $H_\infty$ filters, predictors, and smoothers," *Syst. Contr. Lett.*, vol. 16, pp. 309–317, 1991.

[16] C. Moler, J. Little, and S. Bamgert, *PC-MATLAB*. Sherborn, MD: Mathworks, 1987.

[17] R. H. Shumay and D. S. Stoffer, "An approach to the time series smoothing and forecasting using the EM algorithm," *J. Time Ser. Anal.*, vol. 3, pp. 253–264, 1982.



**Xuemin Shen** (M'97) received the B.Sc. degree from Dalian Marine University, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, in 1987 and 1990, respectively, all in electrical engineering.

From September 1990 to September 1993, he was first with Howard University, Washington DC, and then with the University of Alberta, Edmonton, Alta., Canada. Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada, where he was first a Visiting Research Scientist and is currently an Assistant Professor. His current research focuses on control algorithm development for mobility and resource management in interconnected wireless/wireline networks (traffic flow control, connection admission and access control, handoff, end-to-end performance modeling and evaluation); large scale system modeling, simulation and performance analysis; stochastic process and $H_\infty$ filtering. He is the coauthor of *Singular Perturbed and Weakly Coupled Linear Systems—A Recursive Approach* (Berlin, Germany: Springer-Verlag, 1990) and *Parallel Algorithms for Optimal Control of Large Scale Linear Systems* (Berlin, Germany: Springer-Verlag, 1993).

**Li Deng** (S'83–M'86–SM'91) received the B.S. degree in biophysics from University of Science and Technology of China in 1982, and the M.S. and Ph.D. degrees in electrical engineering from University of Wisconsin, Madison, in 1984 and 1986, respectively.

He worked on large vocabulary automatic speech recognition at INRS-Telecommunications, Montreal, P.Q., Canada, from 1986 to 1989. Since 1989, he has been with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada, where he is currently Full Professor. From 1992 to 1993, he conducted sabbatical research at the Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, working on statistical models of speech production and the related speech recognition algorithms. His research interests include acoustic-phonetic modeling of speech, speech recognition, synthesis, enhancement, speech production and perception, statistical methods for signal analysis and modeling, nonlinear signal processing, neural network algorithms, computational phonetics and phonology for the world's languages, and auditory speech processing.