**ABSTRACT**

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 statistics is required; instead the noise signals are only assumed to have finite energy; 2) the estimation criterion for the filter design is to minimize the worst possible amplification of the estimation error signal in terms of the modeling errors and additive noises. Since most additive noises in speech are not Gaussian, this approach is highly robust and is more appropriate in practical speech enhancement. The global signal-to-noise ratio (SNR), time domain speech representation and listening evaluations are used to verify the performance of the $H_\infty$ filtering algorithm. Experimental results show that the filtering performance is better than other speech enhancement approaches in the literature under similar experimental conditions.

1. 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. In order to enhance the quality and intelligibility of speech, speech enhancement involves processing of speech signal to improve these aspects. There have been numerous studies involving this subject [1]-[4].

Based on stochastic speech models, the previous studies have been focused on the minimization of the variance of the estimation error of the speech, i.e., the celebrated Wiener and/or Kalman filtering approach. This type of estimation assumes that both dynamics of signal generating processes and the statistical properties of noise sources are known in advance. However, this unrealistic assumption naturally limit the application of the estimators since in many situations only approximate signal models are available and/or the statistics of the noise sources are not fully known or unavailable. Furthermore, both Wiener and Kalman estimators may not be sufficiently robust to the signal model errors. In this paper, a new approach based on the recently developed optimal filtering [5]-[10] — $H_\infty$ filtering is presented to 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 statistics is required. The only assumption is that the noise signals have finite energy; and 2) The estimation criterion in the $H_\infty$ filter design is to minimize the worst possible effects of the modeling errors and additive noise on the signal estimation errors. Since the noise added to speech is not Gaussian in general, this filtering approach appears highly robust and more appropriate in practical speech enhancement. Furthermore, the $H_\infty$ filtering algorithm is straightforward to implement. Our experimental results have demonstrated that the filtering performance of the $H_\infty$ estimation noticeably superior to that of the Kalman estimation. The remainder of this paper is organized as follows. In Section 2, the speech source model (vocal tract) is characterized by an all pole filter. This speech source model and the observation model are then combined into a canonical state space formalat for the speech enhancement problem. Based on this formalat, Section 3 presents a $H_\infty$ filtering algorithm. In section 4, we investigate the performance of the $H_\infty$ filter for the speech enhancement, and compare with that of the Kalman filter. Conclusions and discussions are followed.

2. 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 $x_k$ (clean speech) is given by

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

(1)

where $n$ is the number of modeled poles, $a_j$'s are the tap-gain parameters characterizing the filter and $w_k$ is an excitation. If the speech signal is corrupted with a noise signal $v_k$, the observed noisy speech signal $s_k$ is described as follows

$$s_k = x_k + v_k$$

(2)

The speech generating mechanism (1) and observation process (2) are illustrated in Figure 1. Equations (1)-(2) can be represented by the following state-space equations:

$$x_{k+1} = A x_k + B w_k$$

$$z_k = C x_k + D w_k$$

where $A$, $B$, $C$, and $D$ are matrices depending on the model parameters and the observation process. The system is assumed to be non-causal, i.e., the future values of the input are known. The state-space representation of the system is given by

$$\begin{align*}
    x_{k+1} &= A x_k + B w_k \\
    z_k &= C x_k + D w_k
\end{align*}$$

The aim is to design a filter that estimates $x_k$ given $z_k$ and $v_k$. This is achieved by minimizing the worst possible amplification of the estimation error signal in terms of the modeling errors and additive noises. Since the noise added to speech is not Gaussian in general, this filtering approach appears highly robust and more appropriate in practical speech enhancement. Furthermore, the $H_\infty$ filtering algorithm is straightforward to implement. Our experimental results have demonstrated that the filtering performance of the $H_\infty$ estimation noticeably superior to that of the Kalman estimation. The remainder of this paper is organized as follows. In Section 2, the speech source model (vocal tract) is characterized by an all pole filter. This speech source model and the observation model are then combined into a canonical state space formalat for the speech enhancement problem. Based on this formalat, Section 3 presents a $H_\infty$ filtering algorithm. In section 4, we investigate the performance of the $H_\infty$ filter for the speech enhancement, and compare with that of the Kalman filter. Conclusions and discussions are followed.
Kalman Filtering Algorithm

3. KALMAN AND $H_{\infty}$ FILTERING ALGORITHMS

3.1. 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 $R$

\[\begin{align*}
E\{w_k w_k^T\} &= Q \\
E\{v_k v_k^T\} &= R \\
E\{w_k v_k^T\} &= 0.
\end{align*}\]

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

\[P_k = E\{e_k e_k^T\}\] (5)

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

\[e_k = x_k - \hat{x}_k\] (6)

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

\[
\hat{X}_k = A\hat{X}_{k-1} + K_k(s_k - CA\hat{x}_{k-1})
\] (7)

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

\[
K_k = \frac{P_{k|k-1}C}{1 + CK_kC^T} = \frac{P_{k|k-1}CA^T}{\sigma^2 + CP_{k|k-1}C^T}
\] (8)

\[
P_{k|k-1} = AP_{k|k-1}A^T + BRB^T
\] (9)

\[P_k = |I - KS_kC|P_{k|k-1}
\] (10)

where $K_k$ is a Kalman gain vector, $P_{k|k-1} = E[(X_k - \hat{X}_k)\hat{X}_k^T]$, $P_k = E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T]$ is 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

\[\hat{x}_k = C\hat{X}_k\] (11)

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

3.2. $H_{\infty}$ Filtering Algorithm

Consider the state space model (3)-(4), we make no assumptions on the nature of unknown quantities $w_k$ and $v_k$, and are interested not necessarily in the estimation of $X_k$ but $x_k$ using the observations $\{s_i\}$ $(0 \leq i \leq k)$. Let

\[\sigma_k = CX_k\] (12)

Different from that of the Kalman filter which minimizes the variance of the estimation error, the design criterion for the $H_{\infty}$ filter is to provide an uniformly small estimation error, $e_k = x_k - \hat{x}_k$, for any $w_k$, $v_k \in l_2$ and $X_0 \in \mathcal{R}^n$. Note that $\sigma_k$ is equal to $x_k$. The measure of performance is then given by

\[J = \sum_{i=0}^{N-1}||z_i - \bar{z}_i||_Q^2 + \sum_{i=0}^{N-1}||w_i||_W^2 + ||v_i||_{P_{-1}}^2\] (13)

where $\{X_0 - \hat{X}_0, w_k, v_k\} \neq 0$, $X_0$ is an an a priori estimate of $X_0$, $Q \geq 0, P_{-1} > 0, W > 0$ and $V > 0$ are the weighting matrices, and are left to the choice of the designer and depend on performance requirements. The notation $\bar{z}_k^2$ is defined as the square of the weighted (by $Q$) $L_2$ norm of $z_k$, i.e., $\bar{z}_k^2 = z_k^T Q z_k$. The $H_{\infty}$ filter will search $\bar{z}_k$ such that the optimal estimate $\hat{x}_k$ among all possible $\bar{z}_k$ (i.e. the worse-case performance measure) should satisfy

\[\sup J < \gamma^2\] (14)

where “sup” stands for suprenum and $\gamma > 0$ is a prescribed level of noise attenuation. The $H_{\infty}$ filtering can be interpreted as a minimax problem where the estimator strategy $\hat{x}_k$ plays against the exogenous inputs $w_k, v_k$ and the initial state $X_0$. The performance criterion becomes

\[
\min_{\bar{z}_k} \max_{(s_i, w_i, x_i)} J = \frac{1}{2} \gamma^2||X_0 - \hat{X}_0||_P^2 + \frac{1}{2} \gamma^2 ||x_k||_Q^2
\] (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 $u_k$ and $X_0$ are found, and $\hat{x}_k = C\hat{x}_k$, $\hat{z}_k = C\hat{x}_k$, we can rewrite the performance criterion (15) as

$$\min_{\hat{x}_k} \max_{(s_0, s_1, \ldots, s_k)} J = -\frac{1}{2} \gamma^2 |X_0 - \hat{X}_0|^2_p, \quad (16)$$

$$+ \frac{1}{2} \sum_{k=0}^{N} |X_k - \hat{X}_k|^2_Q - \gamma^2 (|u_k|^2_{V-1} + |s_k - C\hat{x}_k|^2_{V-1})$$

where $Q = C^T QC$. In [9]-[10], it has been proved that the following theorem presents a complete solution to the $H_\infty$ filtering 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 $s_k$ if and only if there exists a stabilizing symmetric solution $P_k > 0$ to the following discrete-time Riccati equation

$$P_{k+1} = AP_k A^T - AP_k C^T (V + CP_k C^T)^{-1} CP_k A^T$$

$$+ BWB^T - \gamma^{-2}P_{k+1}C^T(Q^{-1} + \gamma^{-2}LP_k L^T)^{-1}LP_{k+1}$$

$$R_k = (P_0^{-1} + \gamma^{-2}QLQ^T)^{-1}.$$ (17)

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

$$\hat{z}_k = C\hat{x}_k, \quad k = 1, 2, \ldots, N$$

where

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

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

$$H_k = AP_k C^T(V + CP_k C^T)^{-1}$$

Solving Riccati equation (17) for the solution $P_k$ is not trivial due to its nonlinearity. Applying the following matrix inversion lemma (MIL.)

$$A - AB(C + B^T A B)^{-1}B^T A = (A^{-1} + B C^{-1} B^T)^{-1},$$

(17) can be written as

$$P_{k+1}^{-1} = [A(P_{k+1} + C^T V^{-1} C)^{-1}A^T + BWB^T]^{-1} - \gamma^{-2}C^T QC$$

(21)

so that we can obtain the solution of $P_k$ from (22) recursively.

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

1) The Kalman filtering algorithm gives the minimum mean-square-error estimate of $X_k$ based on the $\{s_i\}$ $0 \leq i \leq k$;

2) The $H_\infty$ filtering algorithm gives the optimal estimate of $x_k$ based on the $\{s_i\}$ $0 \leq i \leq k$ such that the effect of the worst disturbance (noises) on the estimation error is minimized.

### 4. Experimental Results

Both Kalman and $H_\infty$ filtering algorithms described in Section 3 are applied to speech enhancement as follows.

Noisy speech is divided into equal-length segments. Within each segment, we first estimate the tap-gain parameters $a_j$, $j = 1, \ldots, n$, then filter the noisy speech with the parameters $\{a_j\}$. The Kalman and $H_\infty$ filtering algorithms are initialized only for the first segment. In our experiment, we choose the 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 trade off in the choice of the length of the segments. Large segments improve the accuracy of the prediction parameters for stationary sounds (e.g. vowels) and short segments improve the accuracy for nonstationary sounds. In our experiment, the segment length used for calculating the parameters $\{a_j\}$ is 128 samples, which corresponds to 16 ms (with a sampling frequency of 8 KHz). The order of the all pole filter is 10, which is a commonly used value in linear predictive analysis of speech signal and the input SNR is 5 dB. Two types of noise are used: white noise (stationary) and helicopter noise (nonstationary).

The performance of both filtering algorithms is measured in terms of SNR, time domain speech representation, and listening evaluation. In the testing, the order of state space model is set to be equal to the order of the all pole filter. Experiment results obtained for the sentence “Woe betide the interviewee if he answered vaguely” are shown in the following tables. Table 1 shows the output SNR results for the testing where the weightings are $W = 1$, $V = 3$, the attenuation parameter $\gamma = 1.02$, and $\gamma_0 = 1000$. The SNR values throughout this paper are the global signal to noise ratios calculated by

$$SNR = 10 \log_{10} \frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} [x_i - \hat{x}_i]^2}$$

Table 1: Performance Comparison with Input SNR=5 dB

<table>
<thead>
<tr>
<th>Filtering Algorithm</th>
<th>White Noise</th>
<th>Helicopter Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalman</td>
<td>8.7276</td>
<td>8.9119</td>
</tr>
<tr>
<td>$H_\infty$</td>
<td>9.8781</td>
<td>10.0693</td>
</tr>
</tbody>
</table>
5. CONCLUSIONS

A new speech enhancement method based on the $H_\infty$ filtering has been developed. This method exploits a speech production model without requiring the knowledge of noise statistics. The effectiveness of the method has been demonstrated based on measurement data and computer simulations. Since the design criterion is based on the worst case disturbance, the $H_\infty$ filtering approach is less sensitive to uncertainty in the exogenous signal statistics and system model dynamics.

6. REFERENCES