

H_∞ FILTERING FOR SPEECH ENHANCEMENT

Xuemin Shen, Li Deng and Anisa Yasmin

Dept. of Elec. & Comp. Eng., Univ. of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ABSTRACT

In this paper, a new approach based on the H_∞ 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_∞ filtering algorithm. Experimented 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_∞ 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_∞ 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_∞ filtering algorithm is straightforward to implement. Our experimental results have demonstrated that the filtering performance of the H_∞ 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 format for the speech enhancement problem. Based on this format, Section 3 presents a H_∞ filtering algorithm. In section 4, we investigate the performance of the H_∞ 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

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

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

$$s_k = x_k + v_k \quad (2)$$

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

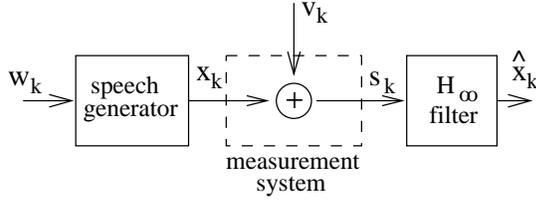


Figure 1: The noisy speech generating and filtering mechanism.

model [2]-[3]

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

$$s_k = CX_k + v_k \quad (4)$$

with $X_k = [x_{k-n+1} \ x_{k-n+2} \ \dots \ x_k]$

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

$$B^T = C = [0 \ 0 \ \dots \ 0 \ 1]$$

The objective of speech enhancement is to estimate x_k given $\{s_i\}$ ($0 \leq i \leq k$). Previous studies have assumed that both w_k and v_k are white, uncorrelated Gaussian processes in order to developing enhancement algorithms, i.e. to estimate x_k [1]-[4]. However neither the speech nor the noise 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 was checked by experiment [3]. In the following, we present a new approach based on the H_∞ filtering algorithm for speech enhancement, where both w_k and v_k may not be white or colored Gaussian processes. For comparison, the Kalman filtering algorithm is briefly reviewed.

3. KALMAN AND H_∞ 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

$$\begin{aligned} E\{w_k w_k^T\} &= Q \\ E\{v_k v_k^T\} &= R \\ E\{w_k v_j^T\} &= 0. \end{aligned}$$

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

$$P_k = E\{e_k e_k^T\} \quad (5)$$

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

$$e_k = x_k - \hat{x}_k \quad (6)$$

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

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

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

$$K_k = P_{k|k-1}C[R + CP_{k|k-1}C^T]^{-1} \quad (8)$$

$$P_{k|k-1} = AP_{k-1}A^T + BQB^T \quad (9)$$

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

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 a priori error covariance matrix, $P_k = E[(X_k - \hat{X}_k)^T(X_k - \hat{X}_k)]$ 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

$$\hat{x}_k = C\hat{X}_k \quad (11)$$

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_∞ 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

$$z_k = CX_k \quad (12)$$

Different from that of the Kalman filter which minimizes the variance of the estimation error, the design criterion for the H_∞ 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 $X_0 \in \mathcal{R}^n$. Note that z_k is equal to x_k . The measure of performance is then given by

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

where $((X_0 - \hat{X}_0), w_k, v_k) \neq 0$, \hat{X}_0 is an *a priori* estimate of X_0 , $Q \geq 0$, $P_0^{-1} > 0$, $W > 0$ and $V > 0$ are the weighting matrices, and are left to the choice of the designer and depend on performance requirements. The notation $|z_k|_Q^2$ 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_∞ filter will search \hat{z}_k such that the optimal estimate z_k among all possible \hat{z}_k (i.e. the worse-case performance measure) should satisfy

$$\sup J \leq \gamma^2 \quad (14)$$

where ‘‘sup’’ stands for supremum and $\gamma > 0$ is a prescribed level of noise attenuation. The H_∞ filtering can be interpreted as a *minimax* problem where the estimator strategy \hat{z}_k play against the exogenous inputs w_k, v_k and the initial state X_0 . The performance criterion becomes

$$\begin{aligned} \min_{\hat{z}_k} \max_{(v_k, w_k, X_0)} J &= -\frac{1}{2}\gamma^2 |X_0 - \hat{X}_0|_{P_0}^2 \\ &+ \frac{1}{2} \sum_{k=0}^N [|z_k - \hat{z}_k|_Q^2 - \gamma^2 (|w_k|_{W^{-1}}^2 + |v_k|_{V^{-1}}^2)] \end{aligned} \quad (15)$$

where “min” stands for minimization and “max” maximization. Note that unlike the traditional minimum variance filtering approach (Wiener and/or Kalman filtering), the H_∞ 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, and $z_k = CX_k$, $\hat{z}_k = C\hat{X}_k$, we can rewrite the performance criterion (15) as

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

where $\bar{Q} = C^TQC$. In [9]-[10], it has been proved that the following theorem presents a complete solution to the H_∞ 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_∞ filter for z_k if and only if there exists a stabilizing symmetric solution $P_k > 0$ to the following discrete-time Riccati equation

$$\begin{aligned} P_{k+1} = & AP_kA^T - AP_kC^T(V + CP_kC^T)^{-1}CP_kA^T \\ & + BWB^T - \gamma^{-2}P_{k+1}L^T(Q^{-1} + \gamma^{-2}LP_{k+1}L^T)^{-1}LP_{k+1} \\ P_0 = & (p_0^{-1} + \gamma^{-2}LQL^T)^{-1}. \end{aligned} \quad (17)$$

If this is the case, then an H_∞ filter can be given by

$$\hat{z}_k = C\hat{X}_k, \quad k = 1, 2, \dots, N \quad (18)$$

where

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

H_k is the gain of the H_∞ filter and given by

$$H_k = AP_kC^T(V + CP_kC^T)^{-1} \quad (20)$$

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

$$A - AB(C + B^TAB)^{-1}B^TA = (A^{-1} + BC^{-1}B^T)^{-1}, \quad (21)$$

(17) can be written as

$$P_{k+1}^{-1} = [A(P_k^{-1} + C^TV^{-1}C)^{-1}A^T + BWB^T]^{-1} - \gamma^{-2}C^TQC \quad (22)$$

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

Comparing the Kalman filtering algorithm (7)-(10) and the H_∞ 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_∞ 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_∞ 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 , $j = 1, \dots, n$, then filter the noisy speech with the parameters $\{a_j\}$. The Kalman and H_∞ 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 $\{a_j, i = 1, 2, \dots, n\}$ 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 $p_0 = 1000$. The SNR values throughout this paper are the global signal to noise ratios calculated by

$$SNR = 10 \log_{10} \frac{\sum_{k=1}^N x_k^2}{\sum_{k=1}^N [x_k - \hat{z}_k]^2} \quad (23)$$

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. Table 1 gives the performance comparison between the Kalman and H_∞ filtering algorithms for the situations that the speech signal is contaminated by white Gaussian noise and helicopter noise respectively. In both cases, the H_∞ filtering algorithm has over 1.0 dB advantage in the output SNR values over the Kalman filtering algorithm [3]. Figure 2 shows the clean speech signal. Figures 3 and 4 show the original noisy speech signals (contaminated by white Gaussian noise and by helicopter noise respectively) as well as the corresponding noise suppressed signals in the time domain by using the H_∞ filtering algorithm. Informal subjective tests confirm the improvement of SNR values, intelligibility and voice quality. The figures clearly show the effectiveness of the H_∞ filtering for speech enhancement.

Table 1: Performance Comparison with Input SNR=5 dB

Filtering Algorithm	Output SNR (dB)	
	White Noise	Helicopter Noise
Kalman	8.7276	8.9119
H_∞	9.8781	10.0693

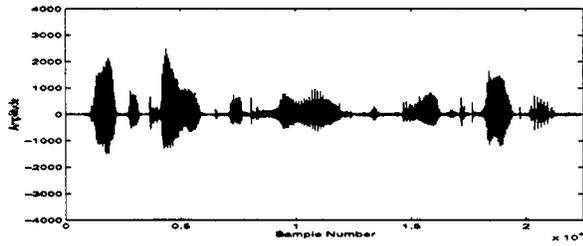


Figure 2: The clean speech signal.

5. CONCLUSIONS

A new speech enhancement method based on the H_∞ 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_∞ filtering approach is less sensitive to uncertainty in the exogenous signal statistics and system model dynamics.

6. REFERENCES

1. J.S. Lim and A.V. Oppenheim, "All-Pole Modeling of Degraded Speech", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, 197-210, 1978.
2. K.K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering", *Proc. IEEE ICASSP*, 177-180, 1987.
3. J.D. Gibson, B. Koo and S.D. Gray, "Filtering of Colored Noise for Speech Enhancement and Coding", *IEEE Trans. Signal Processing*, Vol. SP-39, 1732-1742, 1991.
4. Y. Ephraim, "A Minimum Mean Square Error Approach for Speech Enhancement", *Proc. IEEE ICASSP*, 829-832, 1990.
5. R.N. Banavar and J.L. Speyer, "A Linear Quadratic Game Theory Approach to Estimation and Smoothing", *Proc. IEEE ACC*, 2818-2822, 1991.
6. M.J. Grimble and A. Elsayed, "Solution of the H_∞ Optimal Linear Filtering Problem for Discrete-Time Systems", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-38, 1092-1104, 1990.
7. U. Shaked and Y. Theodor, " H_∞ -Optimal Estimation: A Tutorial", *Proc. 31st IEEE CDC*, 2278-2286, 1992.
8. K.M. Nagpal and P.P. Khargonekar, "Filtering and Smoothing in an H_∞ Setting", *IEEE Trans. Auto. Control*, Vol. AC-36, 152-166, 1991.
9. X. Shen and L. Deng, "Discrete H_∞ Filter Design with Application to Speech Enhancement", *Proc. IEEE ICASSP'95*, Detroit, 1504-1507, 1995.
10. B. Hassibi and T. Kailath, " H_∞ Adaptive Filtering", *Proc. IEEE ICASSP'95*, Detroit, 949-952, 1995.

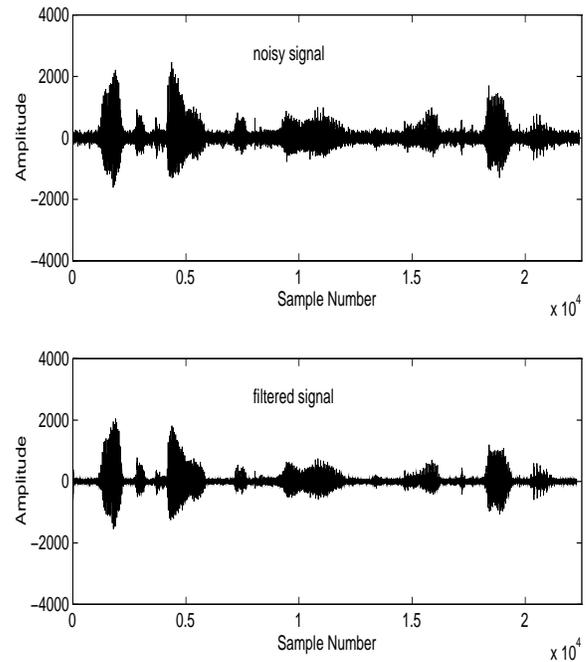


Figure 3: The white Gaussian noise contaminated speech signal and the estimated speech signal using the H_∞ filtering.

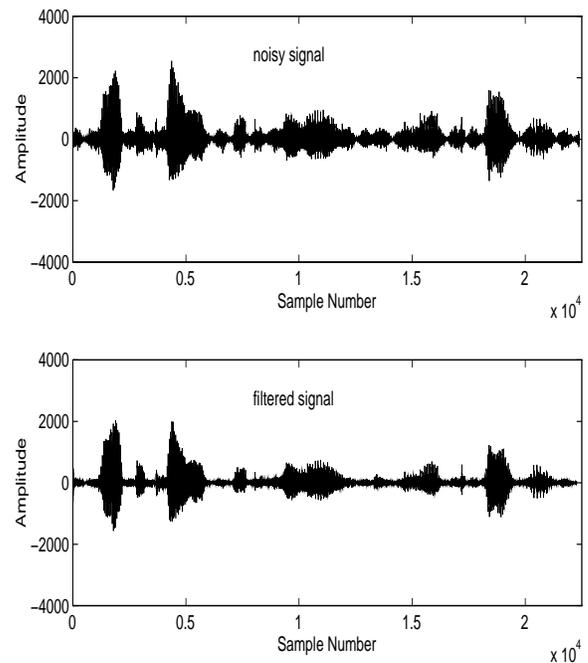


Figure 4: The helicopter noise contaminated speech signal and the estimated speech signal using the H_∞ filtering.