RELEVANCE OF H_∞ FILTERING FOR SPEECH ENHANCEMENT

D. Labarre[1], E. Grivel[1], M. Najim[1] and N. Christov[2]

[1]Equipe Signal & Image, UMR LAPS 5131, ENSEIRB BP 99, 33402, Talence, France
[2]Department of Automatics, Higher Institute of Mechanical and Electrical Engineering, 1156, Sophia, Bulgaria.

ABSTRACT

Among parametric methods for speech enhancement, one consists in combining an autoregressive model for speech and a Kalman filter. This filtering is optimal in the H_2 sense providing the initial state vector, the input and the observation vectors in the state space representation of the system are independent, white and Gaussian. However, these assumptions do not necessarily hold when processing speech. In this paper, we propose to investigate an alternative approach, which is based on H_∞ filtering and hence does not depend on these restrictive assumptions. In that setting, the purpose is to minimize the worst possible effects of the noises and system uncertainties on the estimation error. A comparative study between Kalman and H_∞ filtering is carried out, when the additive colored noise can be modeled by a Moving Average (MA) process.

1. INTRODUCTION

When a speech signal s_k is corrupted by an additive background noise n_k, various methods have been developed to retrieve speech from a single sequence of noisy observations y_k:

$$y_k = s_k + n_k . \quad (1)$$

Indeed, in addition to short term spectral attenuation [4], parametric approaches can be considered to enhance speech. On the one hand various subspace methods, listed in [3], have been proposed for the last 10 years. The underlying assumption is that speech is a sum of complex exponentials. On the other hand, a Kalman filter can be considered [12] [8] [6] [5] [11]. In that case, the existing methods are all based on:

- an AutoRegressive (AR) model for speech:

$$s_k = -\sum_{i=1}^{p} a_i s_{k-i} + \alpha_k \quad (2)$$

where $\{a_i\}_{i=1,p}$ are the AR parameters and α_k is the driving process, which is a zero-mean white Gaussian sequence with variance σ_α^2.

- a state space representation of the system of the form:

$$\begin{cases} \underline{x}_k = \Phi \underline{x}_{k-1} + \Gamma \underline{u}_k \\ y_k = H \underline{x}_k + v_k \end{cases} \quad (3)$$

where $\underline{u}_k$ is the input vector with covariance matrix $\sigma_{\underline{u}}^2$, v_k the observation noise with variance σ_v^2, Φ the transition matrix, Γ the input matrix and H the observation vector. In addition, the state vector $\underline{x}_k$, defined from at least p samples of speech, satisfies:

$$s_k = L \underline{x}_k \quad (4)$$

where L is a row vector.

In an H_2 setting, the estimation of the speech signal can be done by using the finite-horizon *a posteriori* Kalman filtering. This method is a recursive way to estimate $\underline{x}_k$ by minimising the H_2 norm of the estimation error defined as follows:

$$J_2 = \sum_{k=1}^{N} \left\| \underline{s}_k - \underline{\hat{s}}_k \right\|_2^2 \quad (5)$$

where $\|.\|_2$ denotes the standard L_2 norm and $\hat{s}_k = L\hat{x}_k$ is the estimate of the speech signal s_k.

The state vector is thus recursively estimated as follows:

$$\underline{\hat{x}}_k = \Phi \underline{\hat{x}}_{k-1} + K_k \left(y_k - H\Phi \underline{\hat{x}}_{k-1} \right) \quad (6)$$

where $K_k = P_k H \left(H P_k H^T + \sigma_v^2 \right)^{-1}$ and the symmetric matrix $P_k > 0$ satisfies the following Riccati recursion [1]:

$$\begin{aligned} P_k &= \Phi_{k-1} P_{k-1} \Phi_{k-1}^T + \Gamma \sigma_u^2 \Gamma^T \\ &\quad - \Phi_{k-1} P_{k-1} H^T \left(H P_{k-1} H^T + \sigma_v^2 \right)^{-1} H P_{k-1} \Phi_{k-1}^T \end{aligned} . \quad (7)$$

Kalman filter is optimal in the H_2 sense providing the initial state vector, $\underline{u}_k$ and v_k are independent zero-mean white and Gaussian [1]. However, these restrictive assumptions do not always hold in real case, due to the two following model uncertainties.

- *The limits of the AR model for speech*

In our field of interest, it is difficult to satisfy the modeling constraints as the driving process α_k is not available and not perfectly known. When α_k is assumed to be white and Gaussian, the AR model (2) is well suited for noise-like speech such as unvoiced consonants /p/, /s/ or /f/.

However, it cannot really capture the quasi periodic nature of voiced speech frames, such as vowels. For this reason, in the framework of speech coding [2], the so-called long-term predictor is often considered to model the driving process: each sample is expressed from the sample which is a pitch period before. In [9], Goh *et al.* take advantage of this idea and propose a Kalman filter based speech enhancement method in which the excitation α_k is adjusted to the analyzed speech frame. This approach has the advantage to satisfy the whiteness assumption made on the input vector $\underline{u}_k$ for both voiced and unvoiced segment. Nevertheless, the sizes of the matrices in the state space representation of the system get much larger, leading to a higher computational cost. In addition, the Voiced/UnVoiced (V/UV) decision and the estimation of the pitch are all the more difficult to complete as only noisy observations are available.

- *The additive noise modelling issue*

Since only one microphone is used, one has to detect non speech segments with a Voice Activity Detector (VAD) to find a well-suited model for the additive quasi-stationary noise n_k.

Stoica's whiteness tests [15] can be used to define whether n_k is white or colored. However, they are not always reliable, particularly when the noise is "close to" a white Gaussian noise and few samples are available.

When the noise is colored and a q^{th} order AR model is considered, the solution consists in defining the state vector as the concatenation of the p last samples of signal and the q last samples of the additive noise. However, this choice leads to the problem of the so-called noise free state-space representation of the system because $v_k = 0$ [1]. For this reason, a coordinate transformation must be introduced to reduce the dimension of the filter [8].

When choosing a Moving Average (MA) model, n_k satisfies:

$$n_k = \sum_{i=0}^{q} b_i \beta_{k-i} \tag{8}$$

where $\{b_i\}_{i=0,q}$ are the MA parameters and β_k a zero-mean white Gaussian noise with variance σ_β^2. It should be noted that depending on the spectral properties of n_k, the order q can be more or less high. In that case, the state vector related to the state space representation (3) can be defined as follows:

$$\underline{x}_k = \begin{bmatrix} s_k & \cdots & s_{k-p+1} & w_k^1 & \cdots & w_k^q \end{bmatrix}^T \tag{9}$$

where $w_k^j = \sum_{i=1}^{q+1-j} b_{i+j-1}\beta_{k-i}$, $j = 1,...,q$.

Hence, the matrices involved in the state space representation (3) are given by:

$$\Gamma = \begin{bmatrix} H_p^T & O(p,1) \\ O(q,1) & \begin{bmatrix} b_1 & \cdots & b_q \end{bmatrix}^T \end{bmatrix},\ H = \begin{bmatrix} H_p & H_q \end{bmatrix},\ L = H_{p+q},$$

and $\Phi = \begin{bmatrix} -a_1 & \cdots & \cdots & -a_p & & & & \\ 1 & 0 & 0 & 0 & & O(p,q) & & \\ 0 & \ddots & 0 & \vdots & & & & \\ 0 & 0 & 1 & 0 & & & & \\ & & & & 0 & 1 & 0 & 0 \\ & O(q,p) & & & 0 & \ddots & \ddots & 0 \\ & & & & \vdots & & \ddots & 1 \\ & & & & 0 & \cdots & \cdots & 0 \end{bmatrix}$,

where $O(p,q)$ denotes the $p \times q$ null matrix and $H_r = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$ with r-1 zeros.

Defining $v_k = b_0\beta_k$, i.e. a white Gaussian process with variance $\sigma_v^2 = b_0^2\sigma_\beta^2$, the measurement equation in (3) is:

$$y_k = s_k + w_k^1 + b_0\beta_k = H\underline{x}_k + v_k. \tag{10}$$

Besides, the driving process vector $\underline{u}(k)$ is given by:

$$\underline{u}_k = \begin{bmatrix} \alpha_k \\ \beta_{k-1} \end{bmatrix} \tag{11}$$

with covariance matrix $\sigma_{\underline{u}}^2 = \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$.

Therefore, assuming that the additive noise can be modelled by a MA process makes it possible to satisfy the whiteness assumptions on v_k and $\underline{u}(k)$.

In any case, fulfilling the whiteness constraints leads to high computation cost method that requires a VAD, V/UV decision and robust pitch estimation. To avoid a Kalman filter based solution, the H_∞ filtering can be an appealing alternative approach (see part 2). Indeed, in the literature dedicated to signal processing, many papers deal with the H_∞ estimation. They often propose theoretical results but few real applications are addressed. Therefore, our purpose is here to evaluate the relevance of H_∞ filtering for speech enhancement. According to our investigations, only Shen *et al.* have proposed a H_∞ filtering based speech enhancement algorithm [14], where the AR parameters are also estimated by using a H_∞ filter. However, no comparison with previously developed Kalman based methods is completed. In addition, the AR parameter estimation errors must be avoided as much as possible, to analyze the relevance of H_∞ to enhance speech. For this reason, we here propose to replace Kalman filter by H_∞ filter in the well-known methods presented in [12] and [8]. A comparative study is then completed in part 3 with various coloured additive noises.

2. THE H_∞ FILTERING PROBLEM

2.1. State of the art

The H_∞ theory historically appeared in 1981 [17] as an alternative to the H_2 theory in the framework of automatic control. The idea is to minimize the worst possible effects of the disturbances (noises or system uncertainties) through a system on the estimation error. The noises of the state-space representation (3) are only assumed to have bounded energies.

Polynomial approaches [10] have been the first solution proposed to the H_∞ estimation problem. However, they lead to complicated formulas and are not used in practice. Another kind of approaches is based on the game theory [16]. In that case, a statistician plays against nature. Its strategy is to consider the worst possible perturbation of the nature and to minimize the cost of that perturbation. Besides, two families of state-space approaches have emerged:

- the first one is based on the resolution of a convex optimization problem under Linear Matrix Inequality (LMI) constraints [7]. However, the computational cost is quite high.
- the second state-space approach consists in solving a suboptimal H_∞ problem leading to the so-called Algebraic Riccati Equation (ARE) [13]. The resulting equations are easy to implement and the computational cost is lower than the previous approaches. In the following, we will more particularly focus on them.

2.2. The ARE based-H_∞ filtering

When dealing with H_∞ filtering, the additive colored noise α_k and n_k are not modelled but are only assumed to have finite energies, respectively denoted Q and R. The state vector is hence defined as follows:

$$\underline{x}_k = \begin{bmatrix} s_k & \cdots & s_{k-p} \end{bmatrix}^T \tag{12}$$

and we also have :

$\underline{u}_k = u_k = \alpha_k$ and $v_k = n_k$, with finite energies,

$$\Phi = \begin{bmatrix} -a_1 & \cdots & \cdots & -a_p \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix}, \ \Gamma^T = H = H_p \text{ and } L = H .$$

In an H_∞ setting, the problem is to minimise the estimation error $s_k - \hat{s}_k$ for any u_k, v_k and uncertainty of the initial state. The problem is then equivalent to minimize the following criteria (the H_∞ norm):

$$J_\infty = \sup_{u_k, v_k, s_1} \frac{\sum_{k=0}^{N-1} \|s_k - \hat{s}_k\|_2^2}{\sum_{k=1}^{N} \left[\|v_k\|_2^2 + \|u_k\|_2^2 \right] + \|s_1 - \hat{s}_1\|_2^2} \tag{13}$$

It should be noted that in relation (13) $\|s_1 - \hat{s}_1\|_2^2$ is only considered when initial conditions are unknown.

Calculating the supremum J_∞ is a hard task. For this reason, the following suboptimal H_∞ problem is usually considered:

$$J_\infty < \gamma^2 \tag{14}$$

where γ is a prescribed noise attenuation level.

Given the level γ finite horizon *a posteriori* H_∞ filter solving the problem (14) exists if and only if:

$$P_k^{-1} + H^T H - \gamma^{-2} L^T L > 0 \tag{15}$$

where P_k satisfies the following Riccati recursion:

$$\begin{aligned} P_k &= \Phi_{k-1} P_{k-1} \Phi_{k-1}^T + \Gamma \sigma_u^2 \Gamma^T \\ &\quad - \Phi_{k-1} P_{k-1} \begin{bmatrix} H^T & L^T \end{bmatrix} M^{-1} \begin{bmatrix} H \\ L \end{bmatrix} P_{k-1} \Phi_{k-1}^T \end{aligned} \tag{16}$$

with $M = \begin{bmatrix} R & 0 \\ 0 & -\gamma^2 \end{bmatrix} + \begin{bmatrix} H \\ L \end{bmatrix} P_{k-1} \begin{bmatrix} H^T & L^T \end{bmatrix}$.

In that case, the estimations of the state vector and the speech signal are updated as follows:

$$\hat{\underline{x}}_k = \Phi \hat{\underline{x}}_{k-1} + K_k \left(y_k - H \Phi \hat{\underline{x}}_{k-1} \right) \tag{17}$$

where $K_k = P_k H \left(R + H P_k H^T \right)^{-1}$.

Although H_∞ ARE (16) reduces to the conventional Kalman ARE (7) when $\gamma \to \infty$, two differences remain between H_∞ filter and Kalman filter :

- unlike the H_2 problem, a solution to the problem (14) is not always guaranteed since the H_∞ ARE has indefinite quadratic terms. This hence leads to the condition (15).
- the row vector L plays a role in the ARE (16), which is not the case in (7). Indeed, H_∞ filter deals with the estimation of the state but also aims at estimating an arbitrary linear combination of the state vector components. This property is all the more appealing in the framework of speech enhancement as $s_k = L \underline{x}_k$.

3. SPEECH ENHANCEMENT SIMULATIONS

3.1. Introduction

The theoretical approach leads to the belief that H_∞ filter may be a relevant approach for speech enhancement. In this section, we propose to replace H_∞ filter by Kalman filter in the two following standard approaches:

- In Paliwal's method [12], the AR parameters and the noise features are respectively estimated from the clean speech signal and the noise sequence, both assumed available. The noisy observations are then filtered to retrieve the speech signal. Although this method cannot be used in real cases, it makes it possible to focus on the filtering step as no parameter estimation error disturbs the enhancement.

- The second one is based on Gibson's method [8] based on a simplified Expectation-Maximisation (EM) algorithm. The model parameters and the variance σ_α^2 are first estimated from the noisy observations. The noisy observations are then filtered. Subsequently, the parameters are alternately estimated from the enhanced signal and then used to filter anew the noisy signal. The estimation of the additive noise characteristics is done or updated during silent periods.

3.2. Protocol, results and comments

The comparative study we complete is based on Signal-to-Noise Ratio (SNR) improvements and informal subjective tests. We use the utterance /WAZIWAZA/, sampled at 16 kHz. In addition, three colored additive noises at three SNR are here considered:

- Noise MA1: a 4th order MA noise is generated with the corresponding zeros $0.2e^{\pm j0.1\pi}$ and $0.2e^{\pm j0.9\pi}$.
- Noise MA2: a 6th order MA noise is generated with the corresponding zeros $0.9e^{\pm j0.4\pi}$, $0.8e^{\pm j0.5\pi}$ and $0.9e^{\pm j0.6\pi}$.
- Real car noise: the noise is recorded in a running car.

For the simulations on synthetic data, results are based on 100 realizations of the noise. The AR model parameters are estimated with the Levinson algorithm [9]. For the H_∞ based method, the optimal value of γ is estimated with an optimization procedure based on the SNR and is used for the whole sentence.

For the first noise MA1, the optimal value for the estimation of γ is "close to" infinity. The results obtained with both filters are very similar. See tables 1 and 2.

In the MA2 noise case, the colored Kalman filter outperforms the H_∞ filter. The H_∞ based algorithm provides a significant enhancement of the speech without any modeling assumption on the noise. Besides, the optimum value for γ depends on the input SNR.

For real car observation noise, results are very similar. However, the computation cost is higher when using Kalman filtering since the order of the MA model for the additive noise is high.

SNR (dB)	Noise MA1		Noise MA2		Real car noise	
	Kalman	H_∞	Kalman	H_∞	Kalman	H_∞
15	4.01	4.01	6.60	6.66	0.62	0.59
10	5.42	5.43	8.87	8.71	1.14	1.15
5	6.94	6.97	11.51	11.00	2.06	2.11

Table 1: SNR improvements in dB; Paliwal's method

4. CONCLUSION

In this paper, we complete a comparative study between Kalman and H_∞ based speech enhancement algorithms. The results exhibit quite similar results. However, H_∞ based approach have the advantage to avoid restrictive assumption and the computational cost is therefore lower.

SNR (dB)	Noise MA1		Noise MA2		Real car noise	
	Kalman	H_∞	Kalman	H_∞	Kalman	H_∞
15	3.73	3.72	4.33	3.41	0.54	0.59
10	5.09	5.06	5.59	4.39	1.05	1.02
5	6.58	6.49	6.46	6.36	1.89	1.75

Table 2: SNR improvements in dB; Gibson's method

REFERENCES

[1] B. D. O. Anderson, J. B. Moore, Optimal filtering, Ed. T. Kailath, Prentice Hall, 1979.

[2] "Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction", CCITT Recommendation G.728, 1992.

[3] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement", IEEE Trans. on Signal Processing, vol. 50, n°9, 2002, pp. 2230-2244.

[4] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 32, n°6, 1984, pp. 1109-1121.

[5] M. Gabrea, E. Grivel and M. Najim, "A Single Microphone Kalman Filter-Based Noise Canceller", IEEE Signal Processing Letters, vol. 6, n°3, 1999, pp. 55-57.

[6] S. Gannot, D. Burchtein, E. Weinstein, "Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms", IEEE Trans. on Speech and Audio Processing, vol. 6, n°4, 1998, pp. 373-385.

[7] J. C. Geromel, "Optimal Linear Filtering Under Parameter Uncertainty", IEEE Trans. on Signal Processing, vol. 47, n°1, 1999, pp. 168-175.

[8] J. D. Gibson, B. Koo, S. D. Gray, "Filtering of colored noise for speech enhancement and coding", IEEE Trans. on Signal Processing, vol. 39, n°8, 1991, pp. 1732-1742.

[9] Z. Goh, K. C. Tan, B. T. G. Tan, "Kalma_filtering speech enhancement method based on a voiced-unvoiced speech model", IEEE Trans. on Speech and Audio Processing, vol. 7, n°5, 1999, pp. 510-524.

[10] M. J. Grimble and A. E. Sayed "Solution of the H_∞ Optimal Linear Filtering Problem for Discrete-Time Systems", IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 38, n°7, 1990, pp. 1092-1104.

[11] E. Grivel, M. Gabrea, M. Najim, "Speech Enhancement as a realization issue", Signal Processing, vol. 82, 2002, pp. 963-1978.

[12] K. K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering", ICASSP '87, vol. 1, pp. 177-180.

[13] U. Shaked and Y. Theodor, "H_∞-Optimal Estimation: a Tutorial", Proc. of the IEEE Conf. on Decision and Control, Tucson, Arizona USA, 1992, pp. 2278-2286.

[14] X. Shen, L. Deng, "A Dynamic System Approach to Speech Enhancement Using the H_∞ Filtering Algorithm", IEEE Trans. on Speech and Audio Processing, vol. 7, n°4, 1999, pp. 391-399.

[15] P. Stoica, "A test for whiteness", IEEE Trans. on Automatic Control, vol. AC-22, n°6, Decembre 1977, pp. 992-993.

[16] I. Yaesh and U. Shaked, "Game Theory Approach to Optimal Linear State Estimation and Its Relation to the Minimum H_∞-Norm Estimation", IEEE Trans. on Automatic Control, vol. 37, n°6, 1992, pp. 828-831.

[17] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses", IEEE Trans. on Automatic Control, vol. AC-26, n°2, 1981, pp. 301-320.