Audio De-noising Analysis Using Diagonal and Non-Diagonal Estimation Techniques

Sugat R. Pawar[1], Vaishali U. Gaderao[2], and Rahul N. Jadhav[3]
[1, 3] AISSMS, IOIT, Pune, India
Email: sugatpawar@gmail.com
[2] Govt Polytechnique, Pune, India
Email: vaishali.gaderao@gmail.com
Email: jadhavrn@gmail.com

***Abstract* — Audio signals are often contaminated by background environment noise and buzzing or humming noise from audio equipment. Audio denoising aims at attenuating the noise while retaining the underlying signals. Major audio signal denoising techniques are based on attenuation in time-frequency signal representations. The time-frequency audio denoising algorithms of MMSE-LSA estimator of Diagonal Estimation technique and Adaptive time-frequency block Thresholding of non-diagonal estimation techniques are considered. Removing noise from audio signals requires a non-diagonal processing of time-frequency coefficients to avoid producing "musical noise". Time-frequency audio denoising procedures compute a Short-Time Fourier Transform (STFT) of the noisy signal and process the resulting coefficients to attenuate the noise. Simple time-frequency denoising algorithms compute each attenuation factor only from the corresponding noisy coefficients and are thus called diagonal estimators.**

***Index Terms* — Audio Denoising, Block Thresholding, STFT Transform, Spectrogram.**

I. Introduction

The main aim is to reduce the musical noise by performing diagonal and non-diagonal estimation procedures in time-frequency domain to achieve better SNR of the musical noise signal by suppressing the noise and to achieve best audio perception of the musical signal. Diagonal estimators of the SNR are computed from the posterior SNR. The attenuation factor of these diagonal estimators only depends upon corresponding noise coefficient with no time-frequency regularization. The resulting attenuated coefficients thus lack of time-frequency regularity. It produces isolated time-frequency coefficients which restore isolated time-frequency structures that are perceived as a musical noise. This noise is a sum of localized time-frequency structures corresponding to isolated spectrogram coefficients. This superposition of musical noise contaminates the denoised sound and degrades the audio perception. An adaptive block thresholding non-diagonal estimation procedure described, adjusts all parameters adaptively to signal property by minimizing a Stein estimation of the risk. The adaptation is performed by minimizing a Stein unbiased risk estimator calculated from the data. The adaptive block thresholding procedure gives best signal properties. The block attenuation eliminates the residual noise and provides good approximation of the attenuation.

II State Of The Problem

A. Diagonal Estimation

Simple time-frequency denoising algorithms compute each attenuation factor only from the corresponding noisy coefficient and are thus called diagonal estimators. These algorithms have a limited performance and produce a musical noise. In Diagonal Estimation the Posterior SNR is considered. Posterior SNR is the SNR of the Audio Noisy Signal. Diagonal estimators of the SNR are computed from the a posteriori SNR. The attenuation factor of these diagonal estimators only depends upon noisy coefficients with no time-frequency regularization. The resulting attenuated coefficients thus lack of time-frequency regularity. It produces isolated time-frequency coefficients which restore isolated time-frequency structures that are perceived as a musical noise. Some of the Diagonal estimation techniques are Power subtraction (PS), Ephraim and Malah log-spectrogram amplitude (LSA) and decision directed SNR estimator.

B. Non- Diagonal Estimation

To reduce musical noise as well as the estimation risk, several authors have proposed to estimate a priori SNR with a time-frequency regularization of the posteriori SNR. Resulting attenuation factors thus depend upon the data values in a whole neighborhood of and the resulting estimator is said to be non-diagonal. In non-diagonal Estimation prior SNR is considered. Prior SNR is the SNR of original audio signal. .Non-diagonal estimators clearly outperform diagonal estimators but depend upon regularization filtering parameters. Large regularization filters reduce the noise energy but introduce more signal distortion. It is desirable that filter parameters are adjusted depending upon the nature of audio signals. In practice, however, they are selected empirically. Some of the Non-Diagonal estimation techniques are p-point uncertainty model and Block thresholding (BT).

III. Methodology

The basic steps followed to denoise the musical noise signal are as shown in the following block diagram.

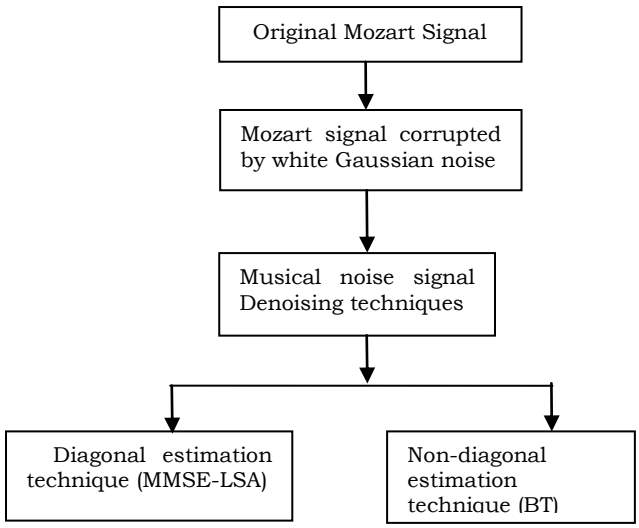

Figure 3.1: Block Diagram of Denoising Musical noise signal.

A. Mmse-LSA Estimator Method

The main aim of the MMSE-LSA Estimator is to minimize the error of the log spectra's of clean signal and the enhanced signal. There were no specific assumptions made about the distribution of the spectral components of either of the signal or noise.

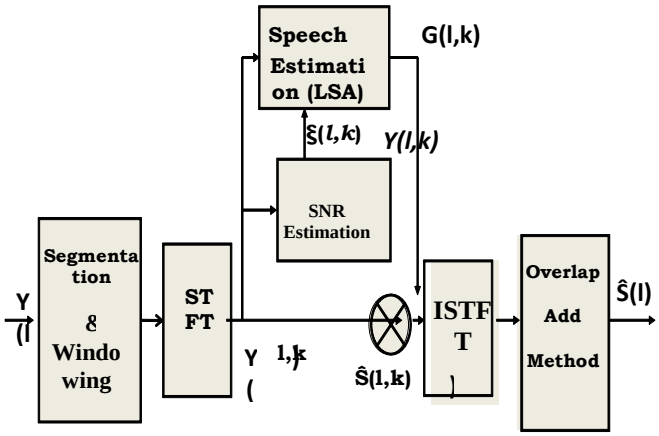

Figure 3.2: Block Diagram of MMSE-LSA Method

This MMSE-LSA technique mainly consists of two steps to be determined .They are,

- Decision-Directed method estimating the 'priori' SNR of the signal spectrum.
- MMSE Short-Time Spectral Amplitude (STSA) estimator.

The MMSE-LSA estimator was implemented in the audio signal enhancement system, operating with the decision- directed a priori SNR estimator. The reduction in the residual noise level is obtained. This method is mainly use for the speech enhancement applications.

In this method the assumptions are made that Stationary additive Gaussian noise with known PSD and An estimate of the speech spectrum is available.Spectral components are statistically independent and each follows Gaussian distribution.

The main aim of the MMSE-LSA Estimator is to minimize the error of the log spectra's of clean signal and the enhanced signal. There were no specific assumptions made about the distribution of the spectral components of either of the signal or noise. In this method, the distribution of the spectral components of the audio signal and the noisy signals are defined. This serves to improve the quality of the resulting enhanced audio signal.

By this method Better output SNR , less musical noise and less distortion to the signal are acheived. The decision-directed approach, proposed by Ephraim and Malah, provides a useful estimation method for the a priori SNR. STFT is applied to determine time and frequency coefficients. The attenuation factor and spectral gain functions are determined by applying MMSE-LSE Estimator to attenuate the noise and to get the denoised signal.

The estimation problem of the STSA [9] is formulated as that of estimating the amplitude of each Fourier expansion coefficient of the speech signal *{ x (t) , 0 < t < T}*, given the noisy process *{ y (t) ,0 < t < T}*.

$$\text{Let, } X_k = A_k e^{j\alpha k}, D_k \quad \text{and}$$

$$Y_k = R_k e^{j\upsilon k}, \quad \text{.....Eq. (3.1)}$$

Denote the k^{th} Fourier expansion coefficient of the speech signal, the noise process, and the noisy observations, respectively, in the analysis interval [0, T].

ξ_k and γ_k are interpreted as the *a priori* and *a posteriori* signal to- noise ratio (SNR), respectively. On considering (3.1) , we get the desired amplitude estimator

$$\hat{A}_k = \frac{\xi_k}{1+\xi_k} \exp\{\frac{1}{2}\int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\} R_k \quad \text{.....Eq. (3.2)}$$

It is useful to consider $\hat{A}_k$ as being obtained from R_k , by a multiplicative nonlinear gain function which depends only on the *a* priori and the a posteriori SNR ξ_k and γ_k , respectively. This gain function is defined by,

$$G(\xi_k, \gamma_k) \triangleq \frac{\hat{A}_k}{R_k} \quad \text{.....Eq. (3.3)}$$

The MMSE-STLSA estimator was implemented in the audio signal enhancement system, operating with the decision- directed *a priori* SNR estimator. The reduction in the residual noise level obtained used is probably a result of the lower gain, particularly in regions of low instantaneous SNR values.

B. Block Thresholding Method

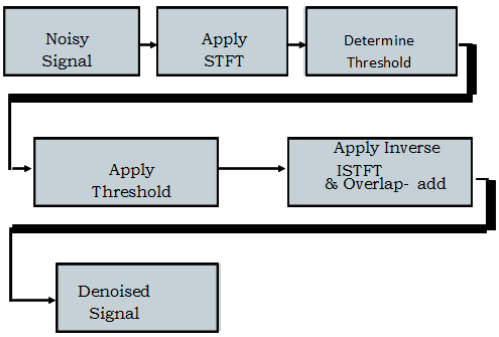

Figure.3.3: BT Block Diagram

To reduce musical noise as well as the estimation risk, several authors have proposed to estimate a priori SNR $\xi[l,k]$ with a time-frequency regularization of the *posteriori* SNR. $\gamma[l,k]$. Resulting attenuation factors $a[l,k]$ thus depend upon the data values $Y[l',k']$ for $[l',k']$ in a whole neighborhood of and the resulting estimator is said to be non-diagonal and is given by,

$\hat{f}[n] = (1/A)\sum_{l,k} a[l,k]Y[l,k]g_{l,k}[n]$ .....Eq. (3.4)

Ephraim and Malah have introduced a decision-directed SNR estimator obtained with a first order recursive time filtering:

$\hat{\xi}[l,k] = \alpha\,\hat{\xi}[l-1,k] + (1-\alpha)(\gamma[l,k]-1)$ .... Eq. (3.5)

In non-diagonal Estimation we consider priori SNR. Priori SNR is the SNR of Audio signal. Non-diagonal estimators clearly outperform diagonal estimators but depend upon regularization filtering parameters.

A block thresholding segments the time-frequency plane in disjoint rectangular blocks of length L_i in time and width W_i in frequency. In the following by "block size" we mean a choice of block shapes and sizes among a collection of possibilities. The adaptive block thresholding chooses the sizes by minimizing an estimate of the risk.

The risk E $\{||f-\hat{f}||^2\}$ cannot be calculated since f is unknown, but it can be estimated with a Stein risk estimate. Best block sizes are computed by minimizing this estimated risk.

A time-frequency block thresholding estimator regularizes estimation by calculating a single attenuation factor over time-frequency blocks. The signal estimator $\hat{f}$ is calculated from the noisy data y with a constant attenuation factor a_i over each block B_i

$\hat{f}$ [n] $= \sum_{i=1}^{I}\sum_{(l,k)\in B_i} a_i Y[l,k]g_{[l,k]}[n]$.....Eq. (3.6)

To understand how to compute each a_i one relates the Stein estimation risk, r =$E\{||f-\hat{f}||^2\}$ to the frame energy conversion and given by,

$\gamma = E\{\|f-\hat{f}\| \le \frac{1}{A}\sum_{i=1}^{I}\sum_{(l,k)\in B_i} E\{|E\{a_i Y[l,k] - F[l,k]|^2\}$Eq. (3.7)

Where f *is* the original signal.

And $\hat{f}$ is the signal estimator calculated from the noisy data with a constant attenuation factor a_i over each block B_i.

A is the redundant factor .If *A=1* then a tight frame is an orthogonal basis. A frame representation provides an energy control. The redundancy implies that a signal f has a non-unique way to be reconstructed from a tight frame.

Since $Y[l,k]=F[l,k]+\varepsilon[l,k]$ one can verify that the upper bound is minimized by choosing

$$a_i = 1-\lambda/(\xi_i+1) \text{.....Eq. (3.8)}$$

A block thresholding estimator can be interpreted as a non-diagonal estimator derived from averaged SNR estimations over blocks. Each attenuation factor is calculated from all coefficients in each block, which regularizes the time-frequency coefficient estimation. To estimate the block thresholding risk Cai used the stein estimator of the risk when computing the mean.

Estimation to the risk $E\{\|f-\hat{f}\|^2\}$ is derived by computing an Estimator $\hat{R}_i$ of the risk in each block B_i

$R_i = \sum_{(l,k)\in B_i} E\{F[l,k]-a_i Y[l,k]^2\}$ .....Eq. (3.9)

Under the hypothesis that the noise variance remains constant on each block B_i, the resulting Stein estimator of the risk is given by,

$$\hat{R}_i = \left\{ \frac{\left[B_i^{\#}+\lambda^2 B_i^{\#2}\right]-2\lambda(B_i^{\#}-2)}{\bar{Y}_i^2/\bar{\sigma}_i^2}\Big|_{(\bar{Y}_i^2<\lambda\bar{\sigma}_i^2)} + (\bar{Y}_i^2/\bar{\sigma}_i^2) + 2B_i^{\#}\Big|_{(\bar{Y}_i^2<\lambda\bar{\sigma}_i^2)} \right\} \cdot ()$$

Here, $B_i^{\#} = l_i \times w_i$

λ is the fixed wave length at each block size. The parameter λ is set depending upon B_i by adjusting residual noise probability.

Given a choice of block size and the residual noise probability level δ that one tolerates, the thresholding level λ .For each block width and length, λ is estimated using "Monte Carlo simulation" .The below Table shows the resulting λ with δ = 0.1%. Let us remark that for a block width W > 1, blocks that contain same number of coefficients, $B^{\#} = L_i \times W$, have close λ values.

Table.3.1: Thresholding level λ calculated with different block size B# = L X W and with δ = 0.1%.

λ value	W= 16	W = 8	W = 4	W = 2	W = 1
L = 8	1.5	1.6	1.9	2.3	2.5
L = 4	1.7	1.9	2.4	3.0	3.4
L = 2	1.9	2.5	3.4	3.2	4.8

The partition of macro blocks in to blocks of different sizes is as shown below:

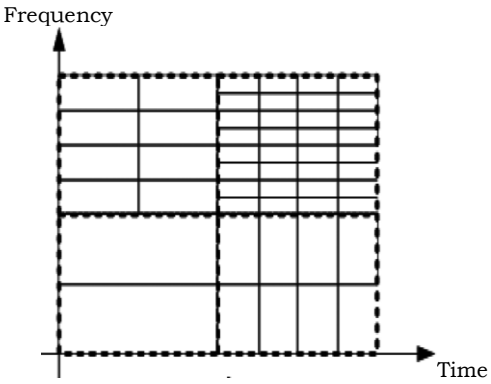

Figure 3.4: Partition of macro blocks

IV. Performance Comparison

The below table compares the performance of Minimum Mean Square Error Log Spectral Amplitude Estimation algorithm by using Decision Direct method (MMSE-LSA-DD) and Block Thresholding (BT) algorithm in terms of SNR.

Signal & SNR	MMSE -LSA	BT
Mozart 2 dB	4.1555dB	10.60dB
Mozart 5 dB	11.1902dB	15.67dB
Mozart 8 dB	14.85dB	18.19dB
Mozart 10 dB	15.67dB	19.53dB
Mozart 15 dB	17.92dB	21.98dB
Mozart 20 dB	18.63dB	23.78dB
Piano 5 dB	13.67 dB	14.35 dB
Recorded 5 dB	13.74 dB	15.61 dB

Table 4.1: Performance Comparison of MMSE-LSA and BT methods for Mozart, Piano and Recorded Signal with different SNR values.

From the above comparison we can conclude that the residual noise masks the musical noise. Block thresholding introduces less signal distortion as reflected by the systematic = 4dB SNR improvement than LSA method.

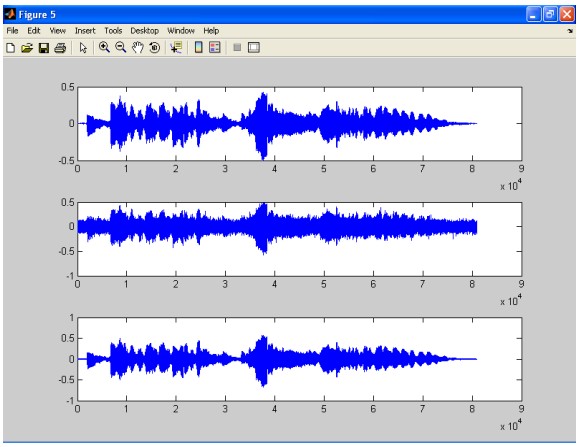

Figure 4.1: Comparison of Original, Noisy and Denoise Mozart Signal Graphs of MMSE-LSA.

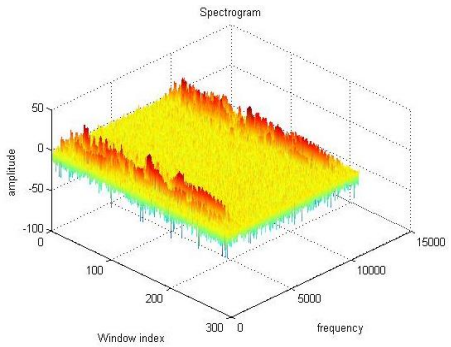

Figure 4.2: Mozart Signal Graph after MMSE-LSA method

The above figure shows the output signal graph after application of MMSE-LSA method. The SNR value obtained after applying MMSE-LSA method for the Mozart signal on application of 5dB white Gaussian noise is 11.1902dB

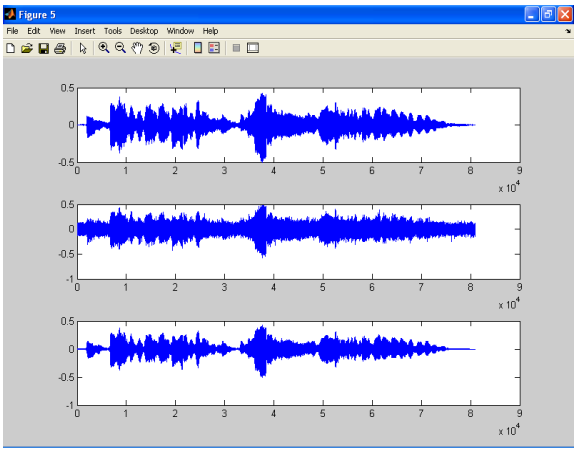

Figure.4.3: Comparison of Original, Noisy and Denoised Mozart Signal Graphs of BT Method.

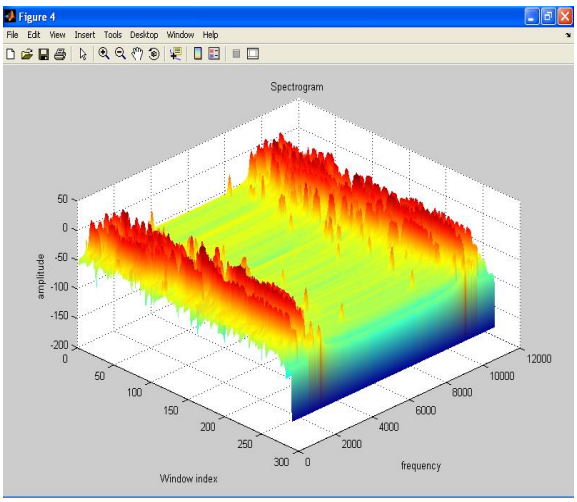

Figure.4.4: Mozart Signal Graph after Block Thresholding

The above figure shows the output signal graph after application of BT method. The SNR value obtained after applying BT method for the Mozart signal on application of 5dB white Gaussian noise is 15.53dB.

V. Conclusions

An adaptive block thresholding non-diagonal estimation procedure is described, which adjusts all parameters adaptively to signal property by minimizing a Stein unbiased risk estimator calculated from the data. Non-diagonal estimators are derived from a time-frequency SNR estimation performed with parameterized filters applied to time-frequency coefficients. The coefficients are grouped into blocks to compute attenuation factors. This block grouping regularizes the estimation which removes musical noises. The objective evaluations have been performed by considering the SNR measures of musical noise signal and denoised signal. The Block Thresholding based algorithm achieved systematically a better SNR improvement than MMSE-LSA method with an average gain of 2dB.The numerical SNR values for different noise levels are tabulated. These results confirm that that Block Thresholding out performed MMSE-LSA based algorithm.

Thus a non-diagonal audio denoising algorithm through adaptive Time-frequency Block Thresholding produces hardly any musical noise signal and improves the SNR compared to Diagonal estimation procedure of MMSE-LSA technique. Block thresholding regularizes the estimate and is thus effective in musical noise reduction Numerical experiments demonstrate improvements of Time-Frequency Block Thresholding with respect to MMSE-LSA procedure.

References

[1] C.Stein, “Estimation of the mean of a multivariate normal distribution,”Ann.Statist.,vol.9,pp.1135–1151,1980.

[2] G. Matz and F. Hlawatsch, “Minimax robust on stationary signal estimation based on a p- point uncertainty model ”, J. Franklin Inst.(Special Issue on Time-Frequency Signal Analysis and Applications), vol.337, no. 4, pp. 403–419, Jul. 2000.

[3] Cohen,“Optimal speech enhancement under signal presence uncertainty using log- spectral amplitude estimator,” IEEE Signal Process. Lett., vol. 9, no. 4, pp. 113–116, Apr.2002.

[4] Cohen, “Relaxed statistical model for speech enhancement and a priori SNR estimation”, IEEE Trans. Speech Audio Process., vol.13, no.5, pp.870–881, Sep.2005.

[5] Ivan Selesnick, "Short Time Fourier Transform”, Connexions, August 9, 2005.

[6] R. M. Gray, A. Buzo, A. H. Gray, Jr., and Y. Matsuyama, “Distortion measures for speech processing,”IEEE Trans.A coust,,Speech, Signalprocessing, vol. ASSP-28, pp. 367-376, Aug. 1980.

[7] T. Cai, “Adaptive wavelet estimation: A block thresholding and oracle inequality approach,” Ann. Statist., vol. 27, pp. 898–924, 1999.

[8] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error Log- spectral amplitude estimator”, IEEE.Trans. Acoust., Speech, Signal Process., vol. 32, no. 6, pp. 1109–1121, Dec.1985.

[9] Zaxel Robel, “Analysis/resynthesis with the short time Fourier transform” Institute of communication science 25th August 2006.