International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 04 | Apr -2017 www.irjet.net p-ISSN: 2395-0072

SPECTRAL SUBTRACTION AND MMSE: A HYBRID APPROACH FOR SPEECH ENHANCEMENT

Brijesh Anilbhai Soni[1], Kinnar Vaghela[2]

[1]Final year PG student,EC Department,L.D.C.E,Gujarat,INDIA
[2]Associate professor,EC Department,L.D.C.E,Gujarat,INDIA

Abstract - *Speech is one of the fundamental means of communication. However clear speech is never possible in the real world. It is always accompanied by the background noise and thus speech enhancement has been a long standing problem in signal processing .Speech enhancement algorithms are important components in many systems where speech plays a part, including telephony, hearing aids, voice over IP, and automatic speech recognizers. Speech enhancement is generally concerned with the problem of enhancing the quality of speech signals.*

In this paper we propose a hybrid approach for speech enhancement which takes the advantage of spectral subtraction and MMSE. We divide entire setup in two stages. First being the spectral subtraction stage and second MMSE.

We implement the spectral subtraction algorithm and MMSE algorithm individually. Next we hybrid both. We observe that SNR is improved to much extent and results are quite promising as compared to individual stages.

Key Words: **Spectral subtraction, MMSE, SNR, Noizeous.**

1.INTRODUCTION

In this paper, we report our work on suppression of acoustic noise or speech enhancement. This problem has received considerable attention, since it is relevant to many important applications like speech recognition and compression, speaker recognition, restoration of analog audio recordings, etc

This paper is structured as follows. In Section 2.1 and 2.2 we will discuss Boll and berouti algorithm which are based on spectral subtraction method.

In section 2.3, we will discuss about the minimum mean square estimator (MMSE) as proposed in[3].
In section 2.4, we will discuss about proposed hybrid approach.
In section 3 we will discuss about experimental results and evaluation.
And in the last section we conclude this paper

2. SPECTRAL SUBTRACTION

One of the widely used technique for speech enhancement is Spectral subtraction. It is mainly due to ease of its implementation. It was introduced in the late '70s by Boll, then generalized and improved by Berouti. Here both are discussed individually

2.1 Boll's Algorithm

The first step in the application of the spectral subtraction method is to compute the short-time Fourier transform of the noisy signal using the fast Fourier transform (FFT) and windowing the input signal with a Hanning window. For this, we set the length of the window and the FFT to 256, with a shift in steps of 128 points.

In the simplest form of spectral subtraction, the estimated magnitude spectrum of the noise $N(\omega)$ is subtracted from that of the noisy speech to obtain the estimated magnitude spectrum of the clean speech while the phase of each spectral component is left unaltered.

Usually, phase is kept unaltered as has not got much importance as per [4], however as per [5], phase plays a role in reconstruction.

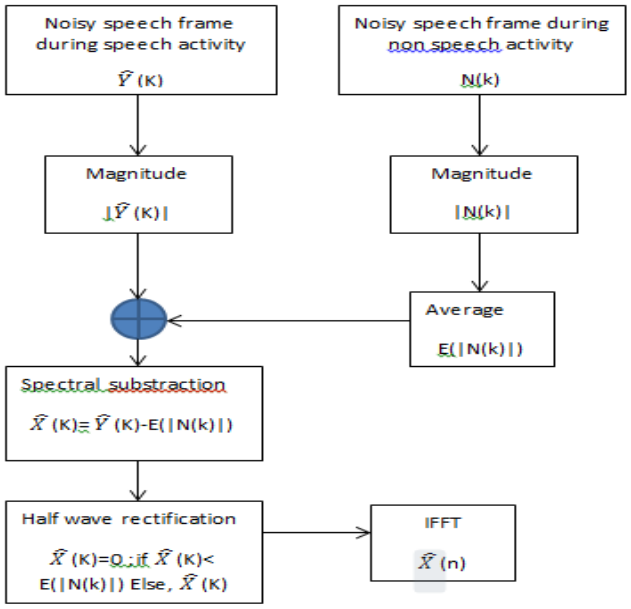

Figure 1: Block diagram of Boll's algorithm. [1]

By first applying the spectral subtraction method as presented in [1], Two problems immediately appears: a clear narrowband of noise still remains in the spectrum, even if our estimate of noise is correct, and listening to the enhanced signal, we can notice an undesirable new noise appearing. As explained by Berouti, peaks and valleys exist in the noise spectrum, and once the estimate is subtracted, peaks remain as randomly occurring peaks, while valleys are set to zero. The peaks are "perceived as time varying tones which we refer to as musical noise."

Assuming an additive model of noise, and given the linearity property of the Fourier transform, we get:

$$Y(\omega) = X(\omega) + N(\omega)$$

Where Y(ω), X(ω) and N(ω), are the Fourier transform of the noisy signal, clean signal, and noise (respectively). Boll describes an algorithm where the short-time noise spectrum N(ω)is first estimated with spectra measured within a noise-only segment resulting in the expected amplitude E{| N(ω)|} of N(ω). The estimate of the clean signal spectrum is obtained as follows:

$$|\hat{X}(\omega)| = |Y(\omega)| - E\{|N(\omega)|\}$$

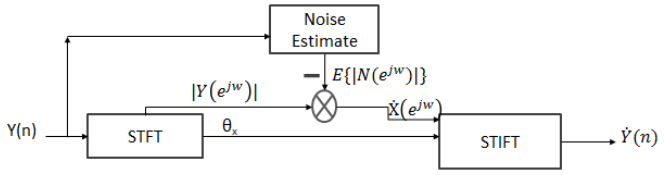

Figure 2: Spectral subtraction method.

The second step is to compute the spectral subtraction estimator $\hat{X}$ using the above amplitude | X(ω) |and the phase of the noisy signal $\theta_Y(e^{j\omega})$ It is widely accepted that the short-time phase is of relative unimportance to estimate $\hat{X}\, e^{j\omega}$. The estimator is then computed:

$$\hat{X}(\omega) = |\hat{X}(\omega)|e^{j\theta}$$

2.2 Berouti Algorithm

Berouti generalizes the spectral subtraction technique by not only considering subtraction of amplitude spectra, but also power spectra, or more generally any power of the short-time amplitude spectrum. Given P_x, P_y, P_n, the power spectra of the estimated clean signal, the noisy signal, and the noise (respectively), Berouti introduced two parameters in the spectral subtractor estimator, which is expressed as follows

$$\hat{P}_x = (P_y^\gamma - \alpha P_n^\gamma)^{1/\gamma}$$

The parameter'α'allows overestimating the power spectrum of noise, and 'γ' raises the power of the power spectrum before subtraction. [2]

Berouti suggested 'α' to be in the range of 3 to 6. Also when γ=1, it is power subtraction and when γ=2, it is magnitude subtraction.

Berouti investigated over-subtraction of both amplitude and power, but it seemed that only over-subtraction of power works well (he concluded that power subtraction performs in general better than amplitude subtraction).

2.3 MMSE (Minimum Mean Square Estimator)

In one of the paper, Ephraim and Malah [1984] proposed an optimal MMSE estimation of the short time spectral amplitude (STSA) [3]; its structure is the same as that of spectral subtraction but, in contrast to the Wiener filtering motivation of spectral subtraction, it optimizes the estimate of the real rather than complex spectral amplitudes.

They proposed two algorithms: a maximum likelihood approach and a decision directed approach which they found

performed better. The maximum likelihood (ML) approach estimates the SNR (or a priori SNR) by subtracting unity from the low-pass filtered ratio of noisy-signal to noise power (the a posteriori or instantaneous SNR) and half-wave rectifying the result so that it is non-negative.

The decision-directed approach forms the SNR estimate by taking a weighted average of this ML estimate and an estimate of the previous frame's SNR determined from the enhanced speech; the weights used were 0.02 and 0.98 respectively. Both algorithms assume that the mean noise power spectrum is known in advance. [3]

The *a priori* SNR $\xi_{n,k}$ can be considered the true SNR of the k^{th} spectral bin at time n and is given by the ratio of the power of the clean signal and of the noise power:

$$\xi_{n,k} = \frac{E[X_{n,k}^2]}{E[D_{n,k}^2]}$$

The *a posteriori* SNR $\gamma_{n,k}$ can be considered the observed and measured SNR of the k^{th} spectral bin at time n after noise is added which is given by the ratio of the squared magnitude of the observed noisy signal and the noise power.[3] Also, Furthermore $v_{n,k}$ is given by:

$$v_{n,k} = \frac{\xi_{n,k}}{1+\xi_{n,k}} \cdot \gamma_{n,k}$$

Since the clean speech signal is not available, the *a priori* SNR $\xi_{n,k}$ is approximated by the use of the decision-directed approach. The decision-directed *a priori* estimator is defined by the following recursive equation[3]

$$\hat{\xi}_{n,k} = a\,\frac{\hat{X}_{n,k}^2(m-1)}{E[D_{n,k}(m-1)^2]} + (1-a)\max[\gamma_{n,k}(m)-1,0]$$

where $0 < a < 1$ is the weighting factor, where it has been chosen to use $a = 0.98$.

The following initial condition for the first frame (i.e. for m=0) is as:

$$\hat{\xi}_{n,k}(0) = a + (1-a)\mathrm{MAX}[\gamma_{n,k}(0)-1,0]$$

2.4Hybrid Approach(spectral subtraction + MMSE)

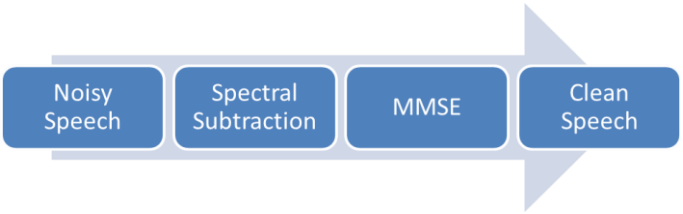

Figure 3: Hybrid Approach (SS + MMSE)

By first applying the spectral subtraction method, two problems immediately appears: a clear narrowband of noise still remains in the spectrum, even if our estimate of noise is correct, and listening to the enhanced signal, we can notice an undesirable new noise appearing. As explained by Berouti [2], peaks and valleys exist in the noise spectrum, and once the estimate is subtracted, peaks remain as randomly occurring peaks, while valleys are set to zero. The peaks are "perceived as time varying tones which we refer to as musical noise."

3 EXPERIMENTAL RESULTS AND EVALUATION

We have used NOIZEUS database for evaluation. It is a noisy speech corpus recorded in Center for Robust Speech Systems, Department of Electrical Engineering, University of Texas at Dallas.

Noise signals were taken from the AURORA database and included the following recordings from different places: babble (crowd of people), car, exhibition hall, restaurant, street, airport, train station, and train. The noise signals were added to the speech signals at SNRs of 0dB, 5dB, 10dB and 15dB. [7]

However we have used only restaurant, street and airport noise all at 10dB.

We have used sp_03.wav as a test speech ('her purse was full of useless trash').

Performance parameter used for evaluation is SNR (signal to noise ratio)

$$SNR = 10log\frac{signal\,power}{Noise\,power}$$

$$SNR = 10log\frac{\sum_{i=1}^{i=n} x^2(t)}{\sum_{i=1}^{i=n}(x(t)-\hat{x}(t))^2}$$

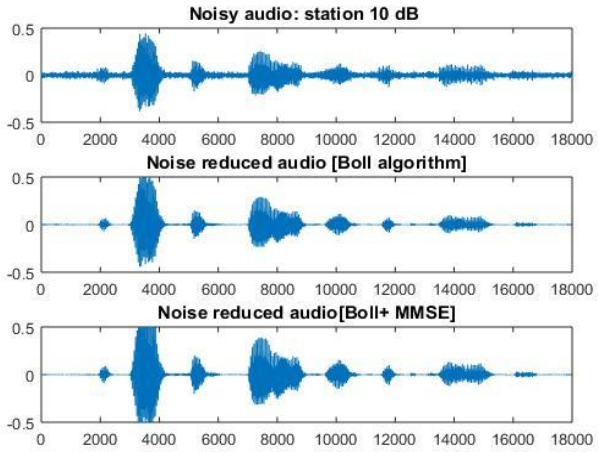

Figure 4: Time domain representation at 10dB station noise

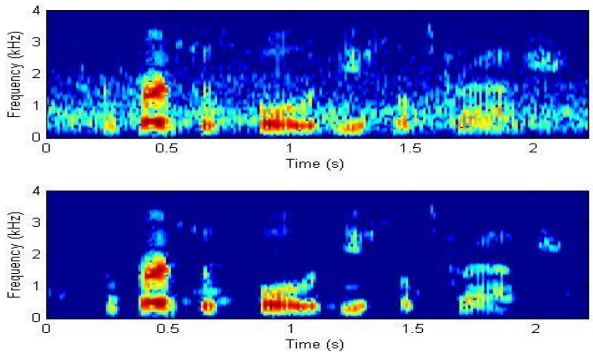

Figure 5: Spectrogram Output at 10db station noise

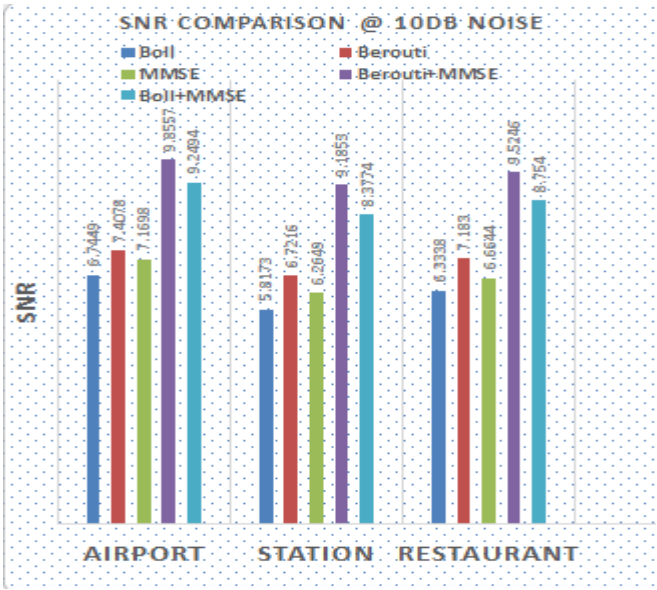

Figure 6: Bar graph of SNR comparison at 10dB noise

Table 1: SNR(dB) comparison of different methods

Speech Sample	Boll	Berouti	MMSE	Berouti + MMSE	Boll + MMSE
Airport_10db	6.74	7.40	7.16	9.85	9.24
Station_10db	5.81	6.72	6.26	9.18	8.37
Restaurant_10db	6.33	7.18	6.66	9.52	8.75

4. CONCLUSIONS

From this work we conclude that Berouti outperforms Boll's algorithm output with SNR as evaluation parameter. However musical noise is still present to some extent.
We also implemented MMSE spectral amplitude algorithm. In this state of art, we did hybridization of spectral subtraction algorithms(Boll and Berouti both) with MMSE. From the experimental evaluation of the proposed approach, we can infer that results of hybrid approach are very promising.

REFERENCES

[1] Boll, S.F. Suppression of Acoustic Noise in Speech using Spectral Subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing (27), pp. 113-120, 1979

[2] Berouti, M., Schwartz, R., and Makoul J. Enhancement of speech corrupted by additive noise. IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 208-211, 1979

[3] Y. Ephraim and D. Malah. Speech Enhancement Using a Minimum-mean Square ErrorShort-time Spectral Amplitude Estimator. Acoustics, Speech and Signal Processing, IEEETransactions on, 32(6):1109–1121, Dec 1984

[4] D. L.Wang and J. S. Lim, "The unimportance of phase in speechenhancement," IEEE Trans. Acoust., Speech, Signal Process.,no. 4, pp. 679–681, 1982.

[5] Stft phase improvement for single chanel speech enhancement, Martin Krawczyk and Timo Gerkmann, Speech Signal Processing Group, Institute of Physics, University of Oldenburg, Germany

[6] The NOIZEUS database Available: http://ecs.utdallas.edu/loizou/speech/noizeus

[7] Hans gunter Hirsch, David Pearce, Aurora experimental framework for the performance evaluation of speech Recognition system under noisy conditions, Automatic Speech recognition (ASR) challenge for the new millennium,France, 18-20th Sept, 2000.