A novel Noise estimation by histogram equalization and optimal filtering for robust speech enhancement

M.Shravani[1] , P.Chandra Sekhar[2], Ch. Ganapathy Reddy[3]
[1]Post graduate student, ECE Dept, G.Narayanamma institute of technology and science, Hyderabad, A.P, India,
[2]Assistant Professor, ECE Dept, G.Narayanamma institute of technology and science, Hyderabad, A.P, India,
[3]Professor,ECE Dept,G.Narayanamma institute of technology and science,Hyderabad,A.P,India
Email:srv05mng@yahoo.com, chandrasekharpaseddula@gmail.com, ganapathi7898@yahoo.co.in

Abstract: Generally, Speech enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in overall perceptual quality of degraded speech signal using audio signal processing techniques. In earlier, there are so many algorithms proposed for speech enhancement. But they are not able to enhance the speech effectively by reducing the noise components. Recently a new mathematical algorithm called Empirical mode decomposition (EMDF) method was proposed. Though this algorithm enhances the speech effectively the time taken to process this enhancement is too high. Mainly this is because of the IMF evaluations for the complete speech samples. To overcome this issue this paper proposes a Histogram based speech enhancement technique. The histogram proposed in this work estimates the noise components contaminated with the clean speech samples, an optimal filtering is proposed to filter those estimated noise samples.

I. Introduction

Speech enhancement plays an important role in numerous applications such as hearing aids¸ speech coding¸ cell phones¸ automatic recognition of speech signals by machines and many more. Speech signals from the uncontrolled environment may contain degradation components along with the required speech components. Degradation components include back ground noise¸ reverberation and speech from other speakers. Therefore the degraded speech components need to be processed for the enhancement. Speech enhancement algorithms improve the quality and intelligibility of speech by reducing or eliminating the noise component from the speech signals. Improving quality and intelligibility of speech signals reduce listener's exhaustion; improve the performance of hearing aids¸ speech coders and many other speech processing systems. In most speech enhancement algorithms it is assumed that an estimate of noise spectrum is available. Noise estimate is critical part and it is important for speech enhancement algorithms. Performance of speech enhancement algorithms depends on correct estimation of noise. Simple approach to estimate the noise spectrum of the signal using a Voice Activity Detector (VAD) [1,2,3,4]another approach to estimate the noise using different noise estimation algorithms Noise estimation algorithms that continuously track the noise spectrum. If the VAD approach is conservative, then it will attempt to reduce false alarms for silence detection, which results in less frequent noise power updates. In highly non-stationary environments, the noise power must be tracked even during speech activity. Noise estimation techniques which operate in the short-time Fourier transform (STFT) domain are very popular, including newer noise estimation systems such as the minimum statistics (MS) [5] and the improved minima controlled recursive averaging (IMCRA) [6]. These techniques estimate the noise spectrum based on the observation that the noisy signal power decays to values characteristic of the contaminating noise during speech pauses. The main challenge faced by these techniques is tracking the noise power during speech segments. This would result in poor estimates during long speech segments with few pauses. Speech enhancement systems such as the optimally modified log-spectral amplitude (OMLSA) estimator [7] require a noise estimate to suppress noise and enhance the noisy speech. . In [11], speech enhancement in car interior noise is achieved by using a speech analysis–synthesis approach, based on a harmonic noise model, as post processing after a traditional log-spectral amplitude speech estimation system. This system is sensitive to accurate pitch estimation and voiced/unvoiced speech frame classification.

Recently a new method for analyzing nonlinear and non-stationary data has been developed. The key part of the method is the empirical mode decomposition [8-10] method with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode functions. This decomposition method is adaptive, and, therefore, highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it is applicable to nonlinear and non-stationary processes. The main problem associated with the implementation of EMDF to enhance the speech is it takes too much time, as well as it is also not applicable to those noisy speech signals which are contaminated with the noise having same power

spectral density at low frequencies with highly non-stationary environments. To overcome this issue this paper proposes a histogram [11] based noise estimation which gives an effective PSD characteristics and optimal filtering based speech enhancement.

The rest of the paper is organized as follows: Section II gives the basic details about the background of empirical mode decomposition and histogram extraction. The proposed histogram based noise estimation and the kalman filtering for speech enhancement is illustrated in sectionIII. The performance evaluation of the proposed approach is illustrated in sectionIV; finally the conclusions are illustrated in sectionV.

II.Back ground

This section gives the basic details about the Empirical mode decomposition and the histogram evaluation. This section is organized under two parts. The first part gives the details about the Empirical mode decomposition which is used recently for speech enhancement. On the other hand the second part gives the basic details about the histogram equalization, applied on the speech signal contaminated with various types of noises.

A.EMDF

EMD is a method of breaking down a signal without leaving the time domain. It can be compared to other analysis methods like Fourier Transforms and wavelet decomposition. The process is useful for analyzing natural signals, which are most often non-linear and non-stationary. This parts from the assumptions of the methods we have thus far learned (namely that the systems in question be LTI, at least in approximation). The EMD method is a necessary step to reduce any given data into a collection of intrinsic mode functions (IMF) to which the Hilbert spectral analysis can be applied. An IMF is defined as a function that satisfies the following requirements:

1. In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

Therefore, an IMF represents a simple oscillatory mode as a counterpart to the simple harmonic function, but it is much more general: instead of constant amplitude and frequency in a simple harmonic component, an IMF can have variable amplitude and frequency along the time axis. The procedure of extracting an IMF is called sifting. The sifting process is as follows:

1. Identify all the local extrema in the test data.
2. Connect all the local maxima by a cubic spline line as the upper envelope.
3. Repeat the procedure for the local minima to produce the lower envelope.

The upper and lower envelopes should cover all the data between them. Their mean is m_1. The difference between the data and m_1 is the first component h_1:

$$X(t) - m_1 = h_1$$

Ideally, h_1 should satisfy the definition of an IMF, for the construction of h_1 described above should have made it symmetric and having all maxima positive and all minima negative. After the first round of sifting, a crest may become a local maximum. New extrema generated in this way actually reveal the proper modes lost in the initial examination. In the subsequent sifting process, h_1 can only be treated as a proto-IMF. In the next step, it is treated as the data, then

$$h_1 - m_{11} = h_{11}$$

After repeated sifting up to k times, h_1 becomes an IMF, that is

$$h_{1(k-1)} - m_{1k} = h_{1k}$$

Then, it is designated as the first IMF component from the data:

$$c_1 = h_{1k}$$

At the end of the decomposition, the data $s(t)$ will be represented as a sum of n IMF signals plus a residue signal,

$$s(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)$$

The finally obtained s(t) gives the completely denoised sample of original speech, though it is efficient to enhance the speech sample it takes too much time to enhance as well as it is not applicable all types of noise contaminated signals. To over this problem a novel speech enhancement technique is proposed in this paper and the complete details are provided in further sections.

B.Histogram

In general, a histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable. A histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval. The height of a rectangle is also equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval.

Consider a speech signal $\{x\}$ and let n_i be the number of occurrences of pauses i. The probability of an occurrence of a pause of level i in the speech is

$$p_x(i) = p(x = i) = \frac{n_i}{n} \; 0 \le i < L$$

L being the total number of pauses in the image, n being the total number of occurrences in the speech, and $p_x(i)$being in fact the histogram for occurrence value i, normalized to [0,1]. Let us also define the cumulative distribution function corresponding to p_x as

$$cdf_x(i) = \sum_{j=0}^{i} p_x(i)$$

which is also the speech's accumulated normalized histogram.

We would like to create a transformation of the form $y = T(x)$ to produce a new speech $\{y\}$, such that its CDF will be linearized across the value range, i.e.

$$cdf_x(i) = iK$$

for some constant K. The properties of the CDF allow us to perform such a transform (see Inverse distribution function); it is defined as

$$y = T(x) = cdf_x(x)$$

Notice that the T maps the levels into the range [0,1]. In order to map the values back into their original range, the following simple transformation needs to be applied on the result:

$$y' = y.(\max\{x\} - \min\{x\}) + \min\{x\}$$

III. Proposed Approach

A. Estimation of noise

In a general mathematical sense, a histogram is a function that counts the number of observations that fall into each of the disjoint categories known as bins, whereas the graph of a histogram is merely one way to represent a histogram as shown in figure-1. Histogram based noise estimation algorithms are motivated by the observation that the Most frequent value (that is the Histogram maximum) of energy values in individual frequency bands corresponds to the noise level of the specified frequency band, that is the noise level corresponds to the maximum of the histogram of energy values. In some cases, the histogram of spectral energy values may contain two modes lst a low energy mode corresponding to the speech absent and low energy segments of speech and 2nd a high energy mode corresponding to the (noisy) voiced segments of speech. The noise estimate is obtained based on the histogram of part power spectrum values [12] that is for each in coming frame, 1st construct the histogram of power spectrum

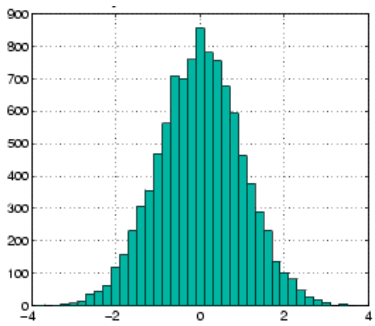

Figure1: Histogram plot

values spanning a window of several hundreds of milliseconds and take as an estimate of the noise spectrum the value corresponding to the Maximum of the histogram values. This is done separately for each individual frequency bin. The histogram based noise estimation is summarized [13] as follows

1. Compute the power spectrum of a noisy speech
$$|y(\lambda, k)|^2.$$
2. Smooth the noisy psd using 1_{st} order recursion.
$$P(\lambda,k) = \alpha P(\lambda-1, k) + (1-\alpha)|y(\lambda, k)|^2$$
Where α is smoothing constant.
3. Compute the histogram of D part PSD estimates
$$P(\lambda, k) \& P(\lambda-1, k)\ P(\lambda-2, k), \text{---------}\ P(\lambda-D,k)$$
using say l bins
4. Let $C = [C_1, C_2, \text{---- } C_l]$ be the counts in each of the 40 bins in the histogram and $S = [S_1, S_2, \text{------ } S_l]$ denote the corresponding centers of the histogram bins.
5. Let C_{max} be the index of the Maximum Count C_{max} = arg Max (C_i) for $I < i < l$. Then take an estimate of the noise psd denoted by $N_{max}(\lambda, k)$ the value corresponding to the maximum of the histogram $N_{max}(\lambda, k) = P(C_{max})$.
6. Smooth the noise estimate $N_{max}(\lambda, k)$ using 1st order recursion
$$r^2(\lambda, k) = \alpha m d^2(\lambda-1, k) + (1-\alpha m) N_{max}(\lambda, k)$$
Where $r^2(\lambda, k)$ is the smoothed estimate of the noise psd and αm is a smoothing constant.

B. Optimal Filtering

If we assume that the speech and noise signals are independent and stationary (even though it is only approximately true), we can use the non-causal optimal filter ([16],[15],[14]) to find a gain factor g_w.

$$g_w = \frac{r_{sx}}{r_x} = \frac{r_x - r_n}{r_x}$$

The equality $r_{sx}=r_s$ comes from the independence of original y_s and y_n from $y_x=y_s+y_n$. Using the certainty equivalence principle ([17]) $y=f(x) \Longrightarrow \hat{y}=f(\hat{x})$.

$$\text{We get: } \hat{g}_w = {}^{\hat{r}_s - \hat{r}_n}/_{\hat{r}_s}$$

In practice we want to assure the non-negativity of g_w. I.e.halfwave rectification

Finally we get

$$r_s = g_w^2 r_x = \frac{\max\{0, r_x - r_n\}}{r_x}$$

Finally the obtained r_s gives the denoised speech signal. The performance of the proposed approach is illustrated in next section.

IV.Performance Evaluation

This section gives the complete details about the performance of the proposed approach. The performance under this section is evaluated for various types of noises. To test the proposed approach we have considered the three types of noisy speech signals. Those are babble noisy speech, restaurant noisy speech and car noisy speech. Now each source signal has 10,000 samples and having the sampling frequency of 16,000 Hz.

The accuracy of the recovered signal ***r(n)*** compared to the desired speech signal can be measured by the signal to noise ratio which is given by

$$SNR = 10log10\left(\frac{r_n}{r_e}\right) = \frac{r_n}{r_x - r_s}$$

Where r_e is the error signal. The following figures (2-16) give the performance evaluation of the proposed approach.

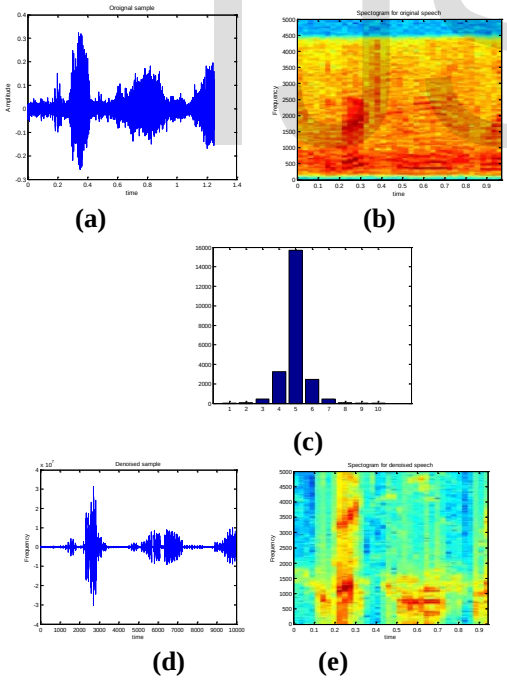

Figure2: (a) Babble noise signal at 5db, (b): Histogram of babble at 5db, (c): Spectrogram of babble at 5db, (d): Denoised sample, (e): Spectrogram of denoised sample

The above figures from figure2 are the test samples belonging to babble noisy speech. The histogram of this sample is shown in figure 3 with 10 bins.

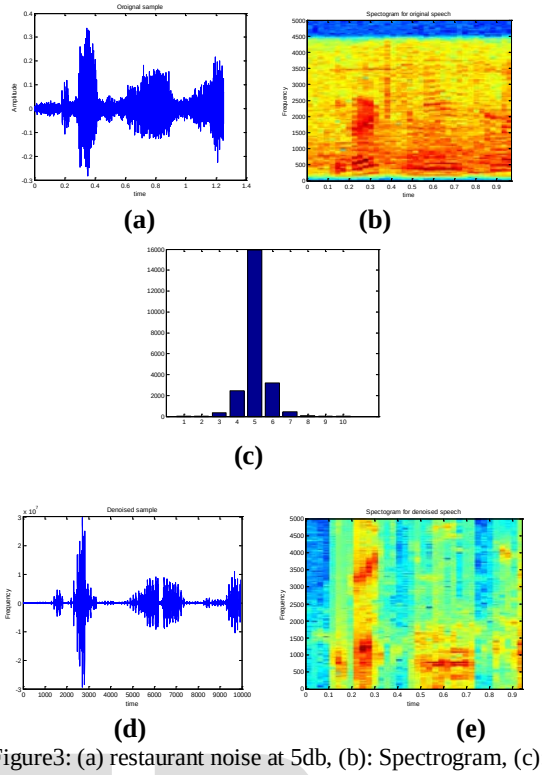

Figure3: (a) restaurant noise at 5db, (b): Spectrogram, (c): Histogram of original sample, (d): Denoised sample, (e): Spectrogram of denoised sample

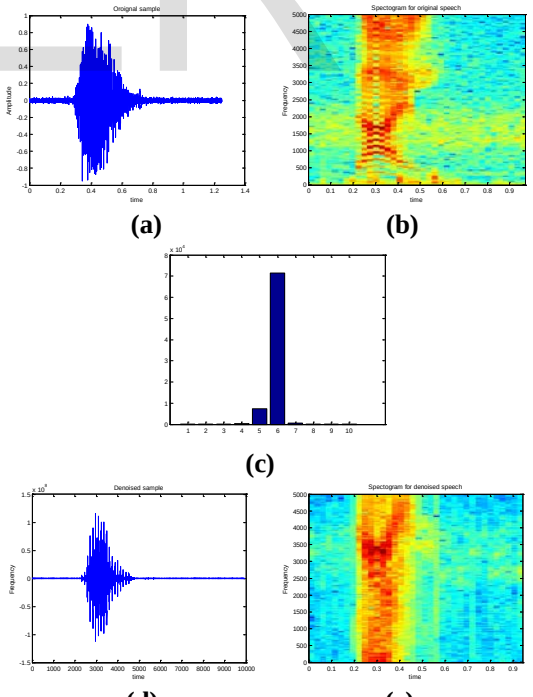

Figure4: (a) car noisy speech at 5db, (b): Spectrogram, (c): Histogram of car noisy speech, (d): Denoised sample, (e): Spectrogram of denoised sample

The table given below shows the SNR values obtained foe the above tested samples.

TABLEI. Segmental SNR for various types of noisy speech samples

Sample	dB	Seg SNR
Car noise	5db	6.234
Babble noise	5db	0.646
Restaurant noise	5db	1.763

Table I shows the average segmental SNR obtained for various noise types and at a 5dB noise level. The proposed approach consistently achieves a higher improvement in the segmental SNR. Its advantage is more significant in non-stationary noise environments.

V. Conclusions

This paper proposes a novel speech enhancement technique to reduce the noise components contaminated in clean speech samples. This proposes a histogram based noise estimation method to estimate the non-stationary noise components, an optimal filtering concept to remove those estimated noise components. The performance of this technique was evaluated using speech contaminated with car interior noise, babble noise, and restaurant noise conditions. When compared to an IMCRA, EMDF systems, this method was shown to give improved performance at suppressing background noise under the presented noisy conditions.

References

[1] Sohn J. and Kim N.(1999), "Statistical Model based voice activity detection", IEEE Signal Proc.Lett.6(1), 1-3.
[2] Tanyer S. and Ozer H. (2000), "Voice Activity Detection in Non stationary Noise", IEEE Speech Audio Procc.8 (4), pp. 478-482.
[3] Shrinivasan K. and Gersho A. (1993), "Voice Activity Detection for Cellular Network", Prec. IEEE Speech Coding Workshop pp. 85-86.
[4] Haigh J. and Mason J.(1993)"Robust Voice Activity Detection using Ceptral Features," Prec. IEEE TENCON
321-324A.
[5] R. Martin, "Noise PSD estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
[6] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Sep. 2003.
[7] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," in *Signal Processing*. Amsterdam, The Netherlands: Elsevier, Nov. 2001, vol. 81, pp. 2403–2418.
[8] P. Flandrin *et al.*, "Detrending and denoising with empirical mode decompositions," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2004, pp. 1581–1584.
[9] K. Khaldi *et al.*, "Speech enhancement via EMD," in *Proc. EURASIP J. Adv. Signal Process.*, 2008, vol. 2008, p. 8.
[10] Y. Kopsinis and S. McLaughlin, "Development of EMD-based denoising methods inspired by wavelet thresholding," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1351–1362, Apr. 2009.
[11] Anuradha R. Fukane, Shashikant L. Sahare, "Noise estimation Algorithms for Speech Enhancement in highly non-stationary Environments", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011
ISSN (Online): 1694-0814
[12] Hirsch, H., Ehrlicher, C. , (1995) "Noise estimation Techniques for robust speech recognition. Proc. IEEE Internat. Conf. on Acoust. Speech Signal Proc.pp 153–156. [13] P. C. Loizou, "Speech Enhancement: Theory and Practice" 1st ed. Boca Raton, FL. CRC, 2007
[14] M. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley and Sons, 1996.
[15] Z. Hrdina, *Statistick´a radiotechnika*, skriptum ˇCVUT, Praha, 1996.
[16] M. Sambur, *Adaptive Noise Canceling for Speech Signals*, IEEE Trans. on Acoustics, Speech and Sig. Proc., October 1978.
[17] J. ˇ Stecha a V. Havlena, *Modern´ı teorie ˇr´izen´ı,* skriptum ˇCVUT, Praha, 1996.