ISSN 1311-4360

SPEECH ENHANCEMENT BASED ON WAVELET DENOISING

S.MANIKANDAN,
Asst Prof,
Department of ECE,
KSR College of tech,
Anna University,
Tiruchengode,
Tamilnadu,
INDIA,
E-mail:smani2k@hotmail.com.

***Abstract**: -* Noise is an unwanted and inevitable interference in any form of communication. It is non-informative and plays the role of sucking the intelligence of the original signal. Any kind of processing of the signal contributes to the noise addition. A signal traveling through the channel also gathers lots of noise**.** It degrades the quality of the information signal**.** The effect of noise could be reduced only at the cost of the bandwidth of the channel which is again undesired as bandwidth is a precious resource. Hence to regenerate original signal, it is tried to reduce the power of the noise signal or in the other way, raise the power level of the informative signal, at the receiver end this leads to improvement in the signal to noise ratio (SNR). There are several ways in doing it and here the focus is on adaptive Signal processing new technique (Grazing Estimation method) to improving the signal to noise ratio.

***Key-words*—**ANC, SNR, MATLAB6.5, LMS, RLS Algorithms, Grazing Estimation Method, Wavelet.

1. INTRODUCTION

This paper is about reducing the content of noise present in the received wireless signal using adaptive techniques. The signal IS corrupted by random additive white Gaussian noise. It is well known that to cancel the noise component present in the received signal using adaptive signal processing technique, a reference signal is needed, which is highly correlated to the noise. Since the noise gets added in the channel and is totally random, hence there is no means of creating a correlated noise, at the receiving end. Only way possible is to somehow extract the noise, from the received signal, itself, as only the received signal can say the story of the noise added to it. The technique used in this paper is a two way process.[1]

As the first step an attempt is made lo estimate a signal correlated to the actual signal i.e. the information bearing component of the received signal. The method used for this generation is presented. Since the signal and the noise are non-coherent to each other, this signal is used to extract noise from the received signal using the interference cancellation technique of the adaptive signal processing, thus giving noise which to a good extent will be correlated lo the noise, in the receded signal.

A technique to generate a signal correlated to the actual signal, which is thus the foremost step in generating correlated noise, is presented next. In this unit, a discussion about a method of generating such a correlated reference noise from the signal received.[2][3]

2. GRAZING ESTIMATION OF SIGNAL

The adaptive techniques to reduce noise are effective when the reference noise is highly correlated to the corrupting noise. But owing to the highly random nature of the corrupting noise, it is difficult to estimate it. Here, it have tried to generate an effective reference noise from the received signal itself, which can *be* then used to reduce the noise content of the same received signal.[4]

The technique used is that of trying to graze through the informative signal and thus trying to find the approximate noise and information content at every instant. This technique is based on having first two samples of the original signal correctly. What is tried to do is, estimate the third sample using the first two samples. This is done by finding the slope between the first two samples for the third sample. This next estimated sample is subtracted with the value at that instant in the received signal. These values give the estimated noise sample at that instant and call it as the estimated noise. Now the second and the third samples are used to estimate the fourth sample in a similar way as was Ihe third sample found. The same method is carried on in generating all the higher samples. The equations governing the above give technique are as follows:[5]

$$m=S^1_{n-1}- S^1_{n-2} \quad (1)$$

$$S^1_n= S^1_{n-1}+m \quad (2)$$

$$N^1_n= X_n+ S^1_n \quad (3)$$

Where S' represents the estimated signal, X represents the received signal, and N' represents the estimated noise of the first stage. The suffices denotes the sample at the instant. Also a threshold for the estimated noise set. This threshold is based on Ihe likely level of noise. The threshold level can be near around 0.5 limes the max absolute value that the noise can take. When ever the absolute value of the estimated noise level Crosses this preset threshold level, the estimated signal value at that infant is reset i.e. when[6][7]

$$N^1_n > \text{threshold, then}$$

$$S^1_n= S^1_{n-1}+ N^1_n/2 \quad (4)$$

This ensures that one don't just keep moving in a single direction. When ever there is more than the expected deviation, attempt is made to bring the estimated signal value in proximity of the signal value. This way the estimated signal sample is kept in close proximity to the original signal through out the course of estimation. The above procedure is efficient for the speech signal has a low bandwidth and which occupy low frequency regions, so That, there will not be many sudden changes in it in the time domain, i.e. the signal will be smooth to the required extent. There are least chances that the estimated sample coincides with the actual value of the original signal at each instant. But it is sufficient to serve the purpose of generating signal which is correlated to information signal to good extent.[8][9]

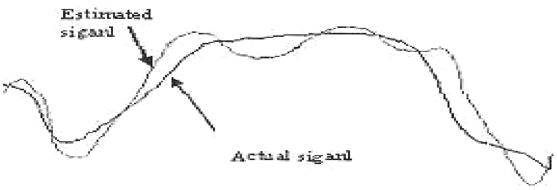

Figure 1: Estimated signal generated by grazing estimation

3. GENERALIZED PROCEDURE

Thus using this scheme, one can generate the well correlated signal to the actual signal indented to transmit. After generating the estimated signal, it is used as the reference signal in the interference cancellation application of the adaptive signal processing. Since the clean signal and noise are uncorrelated, the output of this process is die noise signal itself. The level of correlation of this generated noise and the actual noise depends on the level of the correlation between the estimated signal and the clean signal. Thus our basic aim of generating a correlated noise to the noise corrupting the signal has been now achieved. It is now used this generated noise to cancel the noise In the received signal by using the interference cancellation application again [10][11]

The received signal is passed through the grazing estimation block, the output of which is an estimated signal, denoted by S'(n), which will be correlated to the original signal. Next, this S'{n) is passed through adaptive filtering where it cancels the information content of the received signal, which is totally corrupted with noise. The efficiency of this cancellation depends on the correlation level of the estimated Signal with that of the original signal. With the above suggested method, it was seen that a good cancellation was possible since the noise and the information signal are uncorrelated what is obtained at the output is something, which will be well correlated to the corrupting noise in the received signal. This generated noise is denoted as N'(n) in Figure 2, N'(n) is again passed through adaptive signal processing techniques to cancel the noise content present in the received signal, as shown in Figure 3. Reduction of noise content to an extent of 15-20 db. Was achieved the pseudo-codes for grazing estimation are presented below: [12][13]

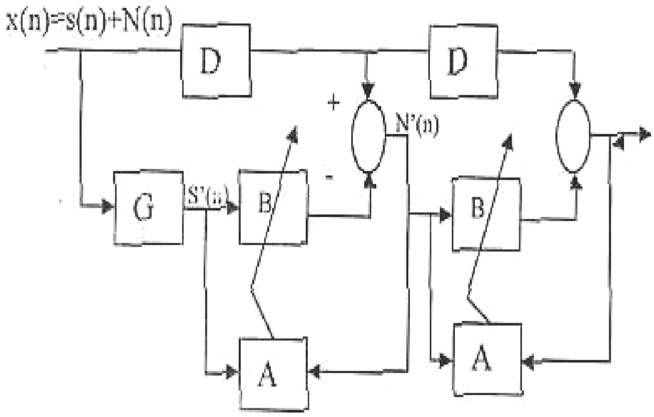

Figure 2: Block diagram of Grazing estimation method

Where the letters denote the following blocks A- Adaptive algorithm, B = Filter, D =delay, G=Grazing Estimation

4. PSEUDO CODE

Grazing {rec.Signal(1)signal{2)),
Load index(1),Indcx{2)=signal(1), signal(2)
Noise1,,noise(2)= rec(1)-index(1},

rec(2)-index{2)
Len = length of the signal
For n=3 to Len
slope= index (n-2)-index{n-1)

index(n}=slope +index(n-2)
noise (n)= rec(n)-index(n)
if absolute value of noise(n)>threshold,
index(n) - index(n) +noise(n)/2
end

It is assumed that the sampled noisy speech signal y, is generated from[14]

$$Y_K = S_K + s_K . n_K \qquad (5)$$

Where *k s* is the clean speech signal, *k n* represents an independent noise source with unit variance ($s_n^2=1$) and s_K is the noise level. Wavelet denoising is an n on – Parametric estimation method that has been proposed in recent years for speech enhancement applications. The goal of wavelet denoisinig is to optimize the mean-squared error Subject to the side condition that with high probability, the estimation, $s^\wedge$ is at least as smooth as s .This constraint provides an optimal trade-off between the bias and variance of the estimate by keeping the two terms the same order of magnitude. The implementation of wavelet denoising is a three step procedure involving wavelet decomposition, nonlinear threshold and wavelet reconstructing although wavelet denoising, provides a theoretical framework to the estimation problem attributes specific lo speech must still be exploited to achieve good performance for the speech enhancement application. Here, the perceptual speech wavelet denoising system using adaptive time-frequency threshold estimation (PWDAT) is utilized. Several new techniques are incorporated into the typical three-step wavelet threshold procedure.[15][16]

5. PREPROCESSING WITH GRAZING ESTIMATION OF SIGNAL

The purpose of preprocessing is to initially lower the noise level *of yk* while minimizing the distortion in s_k (y_k) denotes the output of this preprocessing stage. For this the grazing signal estimation method is applied. The entire structure of implementing wavelet denoising is as given in figure 3 below [17]

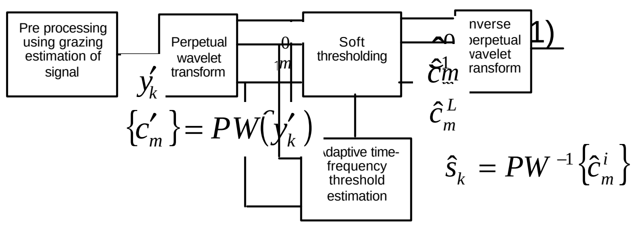

Figure 3 Block diagram for wavelet denoising

Quintile-based noise spectrum estimator to track the slowly varying non-stationary noise statistics. Simulation results show that the grazing estimation of the signal technique achieves modern's levels of noise suppression.[18][19]

6. PERCEPTUAL WAVELET TRANSFORM

A wavelet packet (WP) decomposition designed to mimic the critical bands as widely used m perceptual auditory modeling is utilized The implementation, first proposed by Black and ZeytinogJu, is based *on* an efficient 6-stage tree structure decomposition using 16-tap FIR filters derived from the Daubechres wavelet and provides for an exact invertible decomposition. This perceptual wavelet (PW) transform is used to decompose y_k into subbands,

$$\{c_m^l\} = PW(y_K^1) \qquad (6)$$

Where $\{C_m^i\}$ are the decomposition coefficients with index i *corresponding* to the Subband and *m* corresponding to the 'time' location. For 8 kHz speech, the decomposition results in 16 critical bands. Note that the down sampling operation in the multi-level wavelet packet transform results in a multirate signal representation (i.e., the resulting number of samples corresponding to index *m* differ for each sub band *i).[20]*

6.1TIME-FREQUENCY DEPENDENT THRESHOLD ESTIMATION

Wavelet denoising involves *thresholding in* which coefficients below a specified value *{i.e.* threshold) are set to zero. This is called hard-thresholding. Alternatively, *soft*-thresholding simply shrinks or scales coefficients below the threshold value. A general optimal *universal* threshold for the Gaussian white noise under 3 mean squared error criterion is used. However, in practice this threshold is not ideal for speech signals due the poor correlation between *USE* and subjective quality and the more realistic presence of correlated noise. Here a new adaptive time-frequency dependent thresholds estimation method is used. This involves first estimating the standard deviation of the noise, *FOR* every sub band and time frame. For this a quintile-based noise tracking approach is adapted. Given *X,* the threshold, *a*, is calculated, again for each sub band and time frame. The process starts by segmenting each i-th sub band of decomposed coefficients, cm, into frames of length L^i_{frm}. Denote $s^{\wedge l,p}$ as the corresponding estimated noise level of the P-th frame in the i-th sub band. These are estimated using the segment of previous data \c'*if,m* = *$X*-^,*te* -If where $Z/_{lt|!}$ >L'^ the quintile approach requires sorting the data.[21]

Where int{.) rounds to the nearest integer value. The noise estimate is then given as

$$\hat{s}^{1-P} = b\,.\sum_{j=0}^{\mathrm{int}(q.L^1_{seg})} c_j^{i.p} / L^1_{seg} \qquad (7)$$

Where the constant p is an appropriate scale factor. Nominal values: *q = 0* 2, JJ = 0.33 the corresponding time lengths of U^ and U_m are 512ms and 64ms respectively, and the frame shift is 32ms finally the threshold for each subband at the p-th frame, X^{iP} ,is estimated as given,

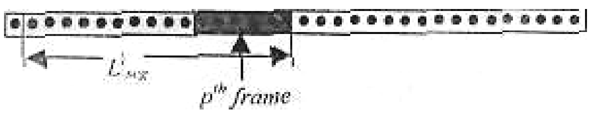

Figure 4 Quintile based Threshold estimation for p^n frame of i^{1h} sub band

6.2. SOFT THRESHOLDING WITH MODIFIED EPHRAIM/MALAH SUPPRESSION RULE

The experiments show that the synthesized speech using a classical hard or soften thresholding operation can result in an unnatural quality. To overcome this, a new wavelet there holding technique by modifying *an* approach by Ephraim and Malah is implemented. In section 5.2.3, an initial threshold A'-'1 for the p-th frame in Ath level is determined. .For simplicity, Let ^=/'"if *m falls* into p-th frame. Define the a *priori* Coefficient to Threshold Ratio (CTR)

$$(R_m^i)^{priori} = \frac{[c_m^i]}{l_m^l}$$
(8)

The corresponding posteriori CTR is then given as

$$(R_m^i)^{priori} = a\frac{|\hat{c}_m^i| - L_{frm}}{l_m^i - L_{frm}} + (1-a)\max(Q[(R_m^i)^{priori} - 1]) \quad (9)$$

Where (Q< <1) is used to control the degree of suppression {nominal setting" *-0.5).* With the a *priori*

$$H_m^i = \frac{(R_m^i)^{post}}{1 + (R_m^i)^{post}} \left(\frac{1}{(R_m^i)^{post}} + \frac{(R_m^i)^{post}}{1 + (R_m^l)^{post}}\right) \quad (10)$$

and *posteriori* CTRs, a Suppression filter can be written Then applied above the suppression filler to the decomposed noisy coefficients *c'm*, that is

$$\hat{c}_m^i = H_m^i . c_m^i \quad (11)$$

6.3 INVERSE PERCEPTUAL WAVELET TRANSFORMS

The last stage simply involves re-synthesizing the enhanced speech using the inverse perceptual wavelet transform,

$$\hat{S}_k = PW^{-1}\{c_m^i\}, j = 0, \ldots\ldots\ldots 17$$
(12)

The results of applying wavelet denoising to speech signals is discussed.[22]

7. SIMULATION RESULTS OF GRAZING ESTIMATION

The results from figure 5 shown on the performance of grazing estimation method and from figure 6 shows the comparisons analysis of grazing estimation method, wavelet de-noising, and combinational approach of grazing estimation and wavelet de-noising in time and spectral domain for Utopia windows start, s1ofwb, s1omwb and ding signals sound signal Ding

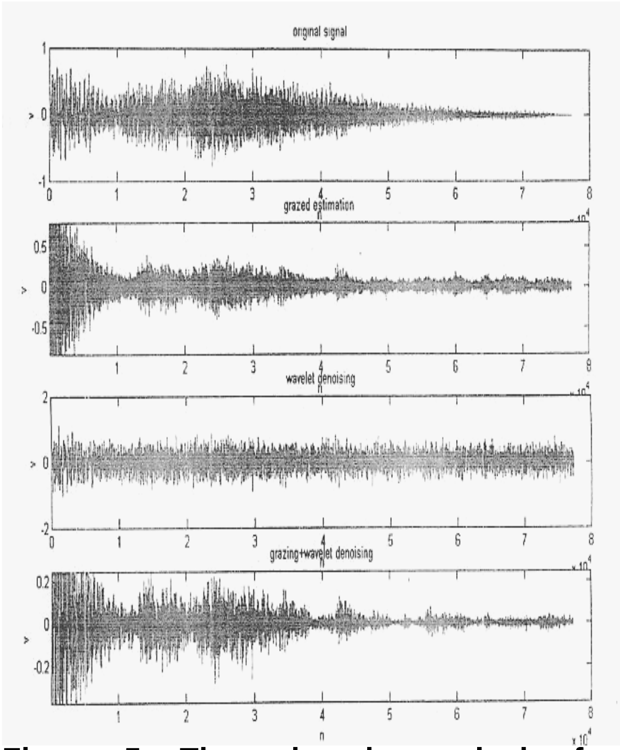

Figure 5: Time domain analysis for grazing estimation method, wavelet denoising and the combination of grazing estimation and wavelet denoising

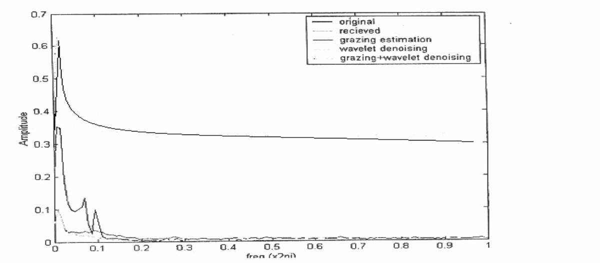

Figure 6: Spectral comparison of the signal, recovered using the mentioned method

The below table is mentioned the performance of each method depends upon the received signal.

Received signal PSNR	PSNR of recovered signal (db)		
	Grazing estimation	Wavelet de-noising	Grazing estimation + wavelet
48.0441	58.7422	56.8878	65.5255
53.0151	59.2054	61.7909	66.9868
58.0323	60.7443	66.2719	68.7585
63.0543	64.1474	70.1104	71.1211
8.0827	68.5550	72.7188	72.9615
73.0773	73.2588	73.9572	73.9424

Table 7.1. Comparisons of various methods for ding

8. CONCLUSION AND FUTURE WORK

This approach is thus very efficient it is cascaded by other noise reducing methods. As can be seen from the results, that when this method was cascaded by wavelet de-noising method it overall was very much improved the efficiency of the combination was better than when either of the techniques were used individually. Thus the combination of this method with some general known other methods, gives the advantage of transmitting signals with low power, than required in case when the other method is used individually, as well as enhancing the SNR to the required level, but of course, this will come at the cost of higher computational time.

Thus in the case of analog communication, this method can be used before the De-emphasis circuit at the receiving end. In case of digital communication, combination of this method and the matched filter will work very efficiently. Thus this method is an efficient way for pre-processing the received signal.

There lie lots of uncovered potential techniques, which can make this method more and more self-reliant. One such thing is a still efficient way of estimating the signal. The above results have through simulation. The method could be tested for real time situations using DSP processors.

References:

[1] *IEEE Trans. On Neural N Woon S. Gan and Sen M. Kuo, "Integrated Headsets Using the Adaptive Feedback Active Noise Control System," in Proc. Int. Congress on Sound and Vibration, July 2001, pp. 277-284.*

[2] *Sen M. Kuo and D. W. Sun, "De-reverberation Using Frequency-Domain Center Clipper," in Workshop Record (CD-ROM), FA1.9, IEEE Workshop on Signal Processing Advances in Wireless Communications, March 2001.*

[3] *Sen M. Kuo and Dennis R. Morgan, "Review of DSP Algorithms for Active Noise Control," in Proc. IEEE Int. Conf. On Control Applications, Sept. 2000, pp. 243-248.*

[4] *T.G. Tsuei, Anu Srinivasa, and Sen M. Kuo, "Analysis of Adaptive Feedback Noise Control System," in Proc. IEEE Int. Conf. On Control Applications, Sept. 2000, pp. 249-254.*

[5] *. Sen M. Kuo, "Active Noise Control Systems with the TMS320 Family," DSPSFEST 2000, DSP Technology and Education Conference, August 2000. NIU001 in conference proceeding CD-ROM SPRC045.*

[6] *. T.G. Tsuei, Sen M. Kuo, and Anu Srinivasa, "A New Adaptive Feedback Algorithm for Active Noise Control," in Proc. Int. Conf. Signal Processing Applications and Technology, Oct. 2000. In CD-ROM.*

[7] *. Sen M. Kuo and Wei M. Peng, "Asymmetric Crosstalk-Resistant Adaptive Noise Canceler," in Proc. IEEE Workshop on Signal Processing Systems, October, 1999, pp. 605-614.*

[8] *Xuan Kong, Sen M. Kuo, Yuan Qiu, "A New Feedback Active Noise Control System with Adaptive Line Enhancer," in Proc. Int. Conf. On Signal Processing Applications and Technology, Nov. 1999, file 307.pdf in CD-ROM proceedings.*

[9] . Sen M. Kuo and Yung-Jeng Yan, "Development and Implementation of Adaptive Echo Cancellation Techniques," in Proc. Fourth Int. Conf. Signal Processing, Oct. 1998, pp. 494-497.

[10] *Jeanine Z. Pan and Sen M. Kuo, Integrated Adaptive Noise and Echo Cancellation Microphone System," in Proc. Fourth Int. Conf. Signal Processing, Oct. 1998, pp. 461-464.*

[11] *Sen M. Kuo, Xuan Kong, and Shaojie Chen, "Out-of-Band Overshoot in Narrowband Active Noise Control Systems," in Proc. Fourth Int. Conf. Signal Processing, Oct. 1998, pp. 474-477*

[12] *Sen M. Kuo, Xuan Kong, Shaojie Chen, and Wenge Hao, "Analysis and Design of Narrowband Active Noise Control Systems," in Proc. IEEE Int. Conf. Acoust. Speech & Signal Processing, vol. VI, pp. 3557-3560, Seattle, WA, May 1998.*

[13] *Xuan Kong and Sen M. Kuo, "An Analysis of ?-Step Ahead Adaptive Prediction and its Application in Active Noise Control," 35th Allerton Conf., Urbana, IL, Oct. 1997, pp. 798-806.*

[14] *Gregory H. Canfield and Sen M. Kuo, "Dual-Channel Audio Equalization and Cross-Talk Cancellation for Correlated Stereo Signals," in Audio Engineering Society Preprint 4570 (J-6), 103rd AES Convention, Sept. 1997, pp. 2-9.*

[15] *Sen M. Kuo and Ming Dai, "Multirate Subband Techniques for Speech Enhancement," Int. Conf. Signal Processing Applications and Technology, San Diego, CA, Sept. 1997, pp. 1694-1698.*

[16] *Sen M. Kuo and Greg Canfield, "Cross-Talk Cancellation and Dual-Channel Equalization Technique for Audio Applications," Int. Conf. Signal Processing Applications and Technology, San Diego, CA, Sept. 1997, pp. 275-279.*

[17] *Padma P. Mallela and Sen M. Kuo, "Integrated Noise and Echo Cancellation System for Multimedia Applications," IEEE Int. Conf. on Consumer Electronics, Rosemont, IL, June 1996, pp. 122-123.*

[18] *Tahernezhadi, S. M. Kuo, and R. Yellapantula, "A Subband Based Hands-Free Cellular Phone With Integrated Engine Noise and Acoustic Echo Cancellation," IEEE Int. Conf. on Consumer Electronics, Rosemont, IL, June 1996, pp. 224-225.*

[19] *Sen M. Kuo,"Adaptive Acoustic Noise Cancellation Microphone," IEEE Int. Conf. on Consumer Electronics, Rosemont, IL, June 1996, pp. 124-125.*

[20] *Sen M. Kuo, "Multiple-Channel Adaptive Noise Equalizer," 29th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, Oct. 1995, pp. 1250-1254*

[21] *Sen M. Kuo and Gerald D. Miller, "An Innovative Course on Real-Time Digital Signal Processing Applications," 29th Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA, Oct. 1995, pp. 88-92.*

[22] *David C. Zhang, Gerald D. Miller, and Sen M. Kuo, "Reference Signal Generating Techniques for Narrow-Band Active Noise Control," Int. Conf. Signal Processing Applications & Technology, Boston, MA, Oct. 1995, pp. 51-55.*