# Adaptive Weiner Filter for Speech Enhancement under Various Noisy Conditions

Akash A. Salmuthe
Dept. of Electronics and Telecommunication
SNJB's KBJ college of Engineering
Chandwad, Nashik (MH)

Rajesh K. Agrawal
Dept. of Electronics and Telecommunication
SNJB's KBJ college of Engineering
Chandwad, Nashik (MH)

## ABSTRACT
The performance of a noisy speech enhancement algorithm depends on the estimation of the priori signal-to-noise ratio (SNR). The most commonly used approach to estimate the priori SNR parameter uses Decision-Directed (DD), Two Step Noise Reduction (TSNR) and Harmonic Regeneration Noise Reduction (HRNR) method. Two-step noise reduction (TSNR) method eliminate this problem decision-directed method. Common short-time noise reduction techniques introduce harmonic distortion in enhanced speech because of the non-reliability of estimators for small signal to-noise ratios. A simple but effective harmonic regeneration method called Harmonic Regeneration Noise Reduction (HRNR) is used to overcome this problem.

## Keywords
Speech enhancement; harmonic regeneration SNR; noise reduction; TSNR; HRNR.

## 1. INTRODUCTION
The impact of additive background noise in a hands free communication system lowers the quality of the transmitted speech signal. The main purpose of speech enhancement is to diminish the noise while keeping the speech undistorted as much as possible. Noise reduction is useful in many applications such as speech communication and automatic speech recognition where effective noise reduction techniques are required. The most widely used approach to estimate the priori SNR parameter is the decision-directed (DD) method. Estimation of DD Priori SNR of current frame follows the shape of the estimated DD priori SNR with a frame delay i.e. previous frame and instantaneous SNR of current frame. DD method was presented in [1], [8] and modified DD (MDD) method is proposed in [1]. The TSNR method is able to suppress the delay and reverberation effect while maintaining the benefits of the DD method. TSNR method was presented in [3], [6], [7], [8], [10]. The improved signal obtained by the TSNR technique presented still suffers from distortions. The distorted signal is processed to create a fully harmonic signal where all the missing harmonics are regenerated. Hence, this method is called as Harmonic Regeneration Noise Reduction (HRNR) method and it was presented in [3], [5], [8], [9].

Two-step noise reduction (TSNR), to refine the estimation of the a priori SNR which get rid of the drawbacks of the DD approach while maintaining its advantage, i.e. highly reduced musical noise level. The major advantage of this approach is the suppression of the frame delay bias leading to the cancellation of the annoying reverberation effect characteristic of the DD approach.

Also, one major restriction that exists in classic short-time suppression techniques, including the TSNR, is that some harmonics are considered as noise only components and accordingly are suppressed by the noise reduction process.

This is inherent to the errors introduced by the noise spectrum estimation which is a very difficult task for single channel noise reduction techniques. Note that in most spoken languages, voiced sounds represent a large amount (around 80%) of the pronounced sounds. Then it is very interesting to overcome this limitation. For that purpose, we propose a method, called harmonic regeneration noise reduction (HRNR) that takes into account the harmonic characteristic of speech. In this approach, the output signal of any classic noise reduction technique (with missing or degraded harmonics) is further processed to create an artificial signal where the missing harmonics have been automatically regenerated [8]

## 2. ADAPTIVE LINEAR FILTER
Fig.1 shows the Adaptive Wiener filter using Decision Directed, Two Step Noise Reduction, and Harmonic Regeneration Noise Reduction techniques. Let y(n)=x(n)+d(n) represents the observed signal. x(n) and d(n) denote speech and uncorrelated noise signals respectively. This observed signal y(n), is transformed into the time-frequency domain by applying the short-time Fourier transform (STFT) with a window function (e.g., Hamming window): y(k , l) where, k=frame index, l=sample index. In STFT, first Noisy speech is decomposed into small frames. Frames are overlapped with previous frame. Each frame is passed through a smoothing window such as Hamming window. Then FFT is taken for signal spectrum, conversion from time to frequency domain.

Magnitude and phase of noisy speech are taken from FFT. Noise parameters are estimated from noisy speech. Modification of Noisy speech magnitude is calculated using Decision Directed (DD), Two Step Noise Reduction (TSNR), and Harmonic Regeneration Noise Reduction (HRNR) techniques. After these techniques, we get new magnitude of noisy speech. The new magnitude and phase is multiplied for Inverse Fast Fourier Transform (IFFT). 'Overlap and add' IFFT of each frame is used to generate enhanced speech.

Most of the speech enhancement techniques require the evaluation of two parameters: the posteriori SNR and the priori SNR, respectively defined by [1].

### 2.1 Decision Directed Method (DD)
In Decision Directed method, Estimation of DD Priori SNR of current frame follows the shape of the estimated DD priori SNR with a frame delay i.e. previous frame and instantaneous SNR of current frame. Estimation of Priori SNR of current frame is given by

$$SNR_{Prior}^{DD}(K, l) = \alpha SNR_{Prior}^{DD}(K - 1, l) + (1 - \alpha)[SNR_{inst}(k, l)] \quad (1)$$

This priori SNR estimator corresponds to the so-called decision-directed approach [8] whose behavior is controlled by the smoothing factor $\alpha = 0.98$. Assume previous frame

$\text{SNR}_{\text{Prior}}^{\text{DD}}$ (0, l) is zero vector of length l for estimation of

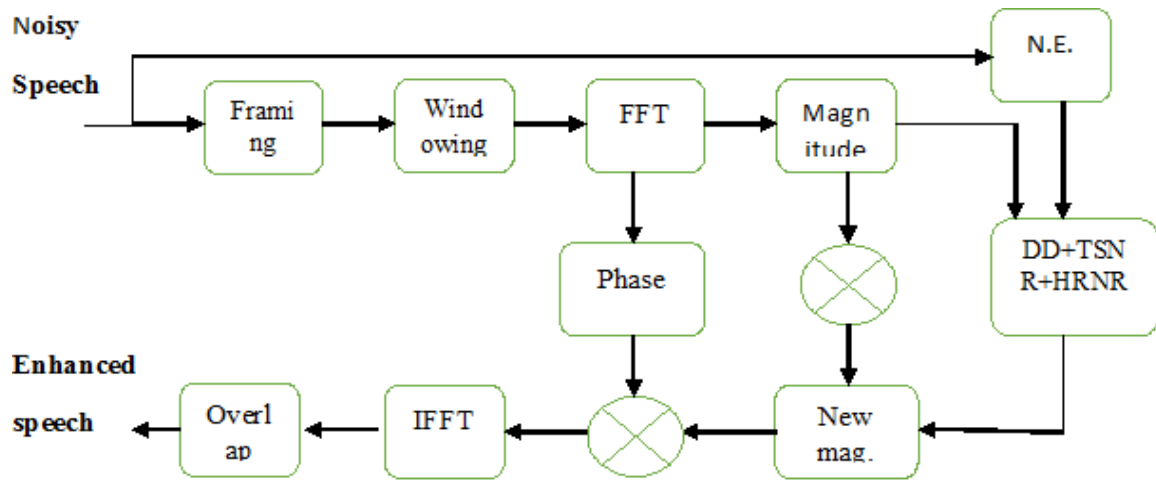first frame $\text{SNR}_{\text{Prior}}^{\text{DD}}$ (1, *l*).



**Fig 1: Adaptive Linear Filter**

## 2.2 Two step Noise Reduction Method (TSNR)

In the well-known Decision Directed method, the priori SNR depends on the speech spectrum estimation of previous frame [1]. The delay introduced by the DD method is a drawback especially when the speech signal is non-stationary like during onset or ending of speech. Furthermore, this delay introduces a bias in the gain estimation and thus limits the noise reduction performance. As a consequence the gain function matches the previous frame rather than the current one which degrades the noise reduction performance and introduces reverberation effect in DD Enhanced speech.

The TSNR method is able to suppress the delay and reverberation effect while maintaining the benefits of the DD method. TSNR is second step for noise reduction. Hence, it is called as Two Step Noise Reduction method. Estimation of TSNR Priori SNR of current frame follows the DD Gain of current frame and it is given by,

$$\text{SNR}_{\text{prior}}^{\text{TSNR}}(k, l) = \frac{\left|\text{Gain}^{\text{DD}}(k, l) * y(k, l)\right|^2}{E[|d(k, l)|^2]} \quad (2)$$

Where, $\text{Gain}^{\text{DD}}(k,l)$ = DD Gain of current frame

TSNR Gain of current frame is a function of TSNR estimated priori signal-to-noise ratio (SNR).

$$\text{Gain}^{\text{TSNR}}(k, l) = \frac{\text{SNR}_{\text{prior}}^{\text{TSNR}}(k, l)}{1 + \text{SNR}_{\text{prior}}^{\text{TSNR}}(k, l)} \quad (3)$$

TSNR Estimated magnitude of noisy speech current frame is:

$$\text{XTSNR}(k, l) = \text{Gain}_{\text{TSNR}}(k, l) * y(k, l) \quad (4)$$

Where, y(k, l)= Noisy speech spectrum of current frame

$\text{XTSNR}(k, l)$ = TSNR Estimated speech spectrum of current frame.

## 2.3 Harmonic Regeneration Noise Reduction Method (HRNR)

The output signal $\text{XTSNR}(k, l)$ is obtained by the TSNR method presented in the previous section still suffers from distortions. Each of the harmonic components of speech are lower in level or amplitude than the fundamental, therefore some harmonics are considered as noise-only components and are suppressed. The harmonic structure of speech is used to prevent this distortion. The distorted signal is processed to create a fully harmonic signal where all the missing harmonics are regenerated. Hence, this is called as Harmonic Regeneration Noise Reduction (HRNR) method.

This method is used to restore harmonics at the desired frequencies i.e. the same positions as in clean speech. A simple and efficient way to restore speech harmonics consists of applying a nonlinear function NL (e.g., absolute value, minimum, or maximum relative to a threshold, etc.) to TSNR Enhanced speech. Then, the artificially restored signal is obtained by,

$$\text{XTSNR}^{\text{Harmo}}(k, l) = \text{NL}(\text{XTSNR}(k, l)) \quad (5)$$

$$= \text{Max}(\text{XTSNR}(k, l), 0)$$

Estimation of HRNR Priori SNR of current frame follows the TSNR Gain of current frame, TSNR Estimated speech squared spectrum of current frame and it's harmonic squared spectrum of current frame.

$$\text{SNR}_{\text{prior}}^{\text{HRNR}}(k, l) = \frac{A + B}{E[|d(k, l)|^2]} \quad (6)$$

Where, $A = [\text{Gain}_{\text{TSNR}}(k, l) * |\text{XTSNR}(k, l)^2|]$

$B = [1 - \text{Gain}_{\text{TSNR}}(k, l) * |\text{XTSNR}^{\text{Harmo}}(k, l)|^2]$

$\text{Gain}^{\text{TSNR}}(k, l)$ = TSNR Gain of current frame

$XTSNR\ (k,l)$ = TSNR Estimated speech spectrum of current frame.

HRNR Gain of current frame is a function of HRNR estimated priori signal-to-noise ratio (SNR).

$$Gain^{HRNR}(k,l) = \frac{SNR_{prior}^{HRNR}(k,l)}{1+SNR_{prior}^{HRNR}(k,l)} \qquad (7)$$

HRNR Estimated magnitude of noisy speech current frame is:

$$XHRNR\ (k,l) = Gain_{HRNR}(k,l) * y(k,l) \qquad (8)$$

where, $y(k,l)$ = Noisy speech spectrum of current frame

$XHRNR\ (k,l)$ = HRNR Estimated speech spectrum of current frame.

## 3. DATABASE DESCRIPTION

We have used NOIZEUS database[11] for noisy speech. NOIZEUS have 6 speakers' speech (with 5 different sentences) are recorded with clean speech and 8 different types of noise at different SNR such as 0 DB, 5 DB, 10 DB, 15 DB. Different noises are like car, train station, train, restaurant, babble, exhibition hall, airport and street.

## 4. EXPERIMENTAL RESULT

Based on the database described in the previous section, Table 1 corresponding to average segmental SNRs obtained using TSNR and HRNR techniques and Fig.1. shows the pictorial view.

The segmental SNR measure takes into account both residual noise level and speech degradation and can be computed, during speech activity, as follows:

$$segSNR = \frac{1}{M}\sum_{m=0}^{M-1}10\log_{10}\frac{\sum_{l=Lm}^{Lm+L-1}s^2(l)}{\sum_{l=Lm}^{Lm+L-1}(\hat{s}(l) - s(l))^2} \qquad (9)$$

Where M is the number of frames that contain active speech, and l is a discrete-time index. For each noise type and SNR value, the average segmental SNR is the result of the averaging of the segmental SNRs obtained for 36 sentences.

The HRNR technique achieves the best results under all noise conditions. The segmental SNR improvement brought by the HRNR technique is explained by its ability to preserve the harmonics degraded by the TSNR.

TABLE I.

| Table Head | Input SNR (dB) | Average Segmental SNR (dB) | |
|---|---|---|---|
| | | *TSNR* | *HRNRR* |
| Babble | 0 | 1.5821 | 1.5703 |
| | 05 | 2.1178 | 2.1122 |
| | 10 | 2.6359 | 2.6367 |
| | 15 | 2.9922 | 2.9941 |
| Car | 0 | 1.2122 | 1.2054 |
| | 05 | 1.8915 | 1.8957 |

| Table Head | Input SNR (dB) | Average Segmental SNR (dB) | |
|---|---|---|---|
| | | *TSNR* | *HRNRR* |
| | 10 | 2.3927 | 2.4012 |
| | 15 | 2.8698 | 2.8797 |
| Street | 0 | 1.5239 | 1.5172 |
| | 05 | 2.1654 | 2.1673 |
| | 10 | 2.7328 | 2.7386 |
| | 15 | 2.9935 | 2.9988 |

## 5. CONCLUSION

HRNR method yields better performance than TSNR method for comparative analysis of segmented SNR ratio between clean and enhanced noisy speech. Result shows HRNR method can also be improved for next level.

## 6. REFERENCES

[1] Wantao Zhang, Shifeng Ou, Suojin Shen, Ying Gao ,"Gain Factor Linear Prediction Based Decision-Directed Method for the a priori SNR Estimation", IEEE, 2015 8th International Congress on Image and Signal Processing (CISP 2015), page-no. 1199-1203.

[2] Dipal J Thathagar, Manish P Patel, and Kinjal M Vagadia Adaptive "Weiner Filter for Speech Enhancement Under Various Noisy Conditions", IJIERE, 2016.

[3] Pankaj Goel, Prateek Saxena, Mahesh Chandra, V.K.Gupta, "Comparative Analysis of Speech Enhancement Methods", IEEE, 2013.

[4] Santosh Chapaneri, "Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping", International Journal of Computer Applications (0975 – 8887), Volume 40– No.3, February 2012, pageno. 1-8.

[5] Hou Xuchu, Zhu Xiaojing, "Speech Enhancement using Harmonic Regeneration", IEEE, 2011, page-no. 150-152.

[6] Xuemin Zhang, Hang Jiang, Xiuyan Wang ,"Improved Priori SNR Estimation for Sound Enhancement Based on DCT ", IEEE, 2011.

[7] R. Thoonsaengngam and N. Tangsangiumvisai, Member, IEEE , "Performance Improvement of acoustic Echo and Noise Suppression using a Priori SNR and Echo Spectrum Variance Estimation Techniques", IEEE, Proceedings of APCC-2008.

[8] Cyril Plapous, Member, IEEE, Claude Marro, and Pascal Scalart, Member, IEEE ," Improved signal to noise ratio Estimation for Speech enhancement", IEEE transactions on audio, speech, and language processing, VOL. 14, NO. 6, NOVEMBER 2006, page-no. 2098-2108.

[9] Cyril Plapous, Claude Marro and Pascal Scalart, "speech enhancement using harmonic regeneration", ICASSP, IEEE, 2005, page-no. I-157 to I-160.

[10] Cyril Plapous, Claude Marro, Laurent Mauuary, Pascal Scalart, "A two-step noise reduction technique", ICASSP, IEEE, 2004, page-no. I-289 to I-29.

[11] The University of Texas at Dallas http://ecs.utdallas.edu/