Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 666 – 676

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Comparison of Speech Enhancement Algorithms

Siddala Vihari, A. Sreenivasa Murthy, Priyanka Soni and D. C. Naik*

U.V.C.E., Bangalore University, Bangalore, India

Abstract

The simplest and very familiar method to take out stationary background noise is spectral subtraction. In this algorithm, a spectral noise bias is calculated from segments of speech inactivity and is subtracted from noisy speech spectral amplitude, retaining the phase as it is. Secondary procedures follow spectral subtraction to reduce the unpleasant auditory effects due to spectral error. The drawback of spectral subtraction is that it is applicable to speech corrupted by stationary noise. The research in this topic aims at studying the spectral subtraction & Wiener filter technique when the speech is degraded by non-stationary noise. We have studied both algorithms assuming stationary noise scenario. In this we want to study these two algorithms in the context of non-stationary noise. Next, decision directed (DD) approach, is used to estimate the time varying noise spectrum which resulted in better performance in terms of intelligibility and reduced musical noise. However, the a priori SNR estimator of the current frame relies on the estimated speech spectrum from the earlier frame. The undesirable consequence is that the gain function doesn't match the current frame, resulting in a bias which causes annoying echoing effect. A method called Two-step noise reduction (TSNR) algorithm was used to solve the problem which tracks instantaneously the non-stationarity of the signal but, not by losing the advantage of the DD approach. The a priori SNR estimation was modified and made better by an additional step for removing the bias, thus eliminating reverberation effect. The output obtained even with TSNR still suffers from harmonic distortions which are inherent to all short time noise suppression techniques, the main reason being the inaccuracy in estimating PSD in single channel systems. To outdo this problem, a concept called, Harmonic Regeneration Noise Reduction (HRNR) is used wherein a non-linearity is made use of for regenerating the distorted/missing harmonics. All the above discussed algorithms have been implemented and their performance evaluated using both subjective and objective criteria. The performance is significantly improved by using HRNR combined with TSNR, as compared to TSNR, DD alone, as HRNR ensures restoration of harmonics. The spectral subtraction performance stands much below the above discussed methods for obvious reasons.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Decision Directed Approach; Harmonic Regeneration; Speech Enhancement; Two-step Noise Reduction; Wiener Filtering.

1. Introduction

The processing of noisy speech signals to improve their perception by humans or better decoding by systems is what speech enhancement deals with. A Formulation of speech enhancement algorithms is to improve the performance of a system when its speech input is ruined by noise. It is usually hard to retain speech undistorted while reducing noise and thus, limitation on speech enhancement system's performance- the compromise between speech distortion and noise

*Corresponding author. Tel.: +91 -96114455231
E-mail address: chethan.naik24@gmail.com

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.032

reduction.For distorted speech with medium to high SNR, objective will be to produce subjectively natural signal by reducing noise level and for those with low SNR, objective could be to decrease the noise level, while preserving the intelligibility.

The most common factor that causes the degradation of speech's quality and intelligibility is background noise which can be stationary or non-stationary and is assumed to be uncorrelated and additive to the speech signal. A broad classification of speech enhancement methods can be given as *spectral processing* and *temporal processing* methods. The degraded speech goes through processing in frequency domain in the spectral processing methods, whereas processing will be in time domain for temporal processing method.

Spectral subtraction[1], known for its minimal complexity and relative ease in implementation is one of the oldest algorithms proposed in the area of background noise reduction. In this technique, the average magnitude of noise spectrum is subtracted from the noisy speech spectrum. The average magnitude of noise spectrum is estimated from the frames of speech absence, usually from initial frames of the signal in case of stationary noise conditions. In case the noise is non-stationary, the noise estimate has to be calculated every time the noise characteristics are changed. So, the spectral subtraction becomes inefficient for speech corrupted with non-stationary noise

Utilizing an MMSE criteria[2], using spectral component distribution models of speech and noise signals, the mean square error between the short time spectral magnitude of the clean speech and enhanced speech may be minimized. Speech enhancement based on noise suppression usually disturbs the spectral balance in speech[3], which results in unpleasant distortions in enhanced speech. LP residual enhancement is a method used for LP residual reconstruction.

Using different SNR parameters[4], we can formulate a short time spectral gain using Wiener filtering with DD approach in which frame delay results in an annoying reverberation effect. The problem is solved by TSNR[5], wherein, a second step is formulated so as to remove the delay. All the classic short time noise reduction algorithms are followed by HRNR algorithm[6] which is used to regenerate harmonics in the reconstructed signal.

Organization of the paper is, in Section 2, we discuss the implementation of Spectral subtraction method. In Section 3, we discuss the Wiener filtering using DD, TSNR, and TSNR followed by HRNR method. In Section 4, obtained results are presented. And in Section 5, summary of our work and conclusion drawn is given.

2. Spectral Subtraction

A noise spectrum estimate, derived from the signal measured while non-speech activity or beginning/ending of a speech signal was subtracted from the noisy speech spectrum to obtain a spectral subtraction estimator. A spectral

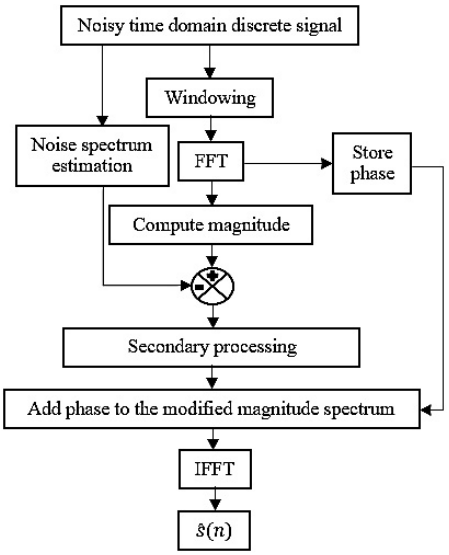


Fig. 1. Flow Chart Showing the Spectral Subtraction Followed by Secondary Processing Needed to Reduce Auditory Effects due to Spectral Error.

error was then calculated and further processing was done to reduce it. In this method, we assumed the noise to be additive background noise and is locally stationary.

Discrete time noisy signal is segmented into short-time frames using Hanning window with an overlap of 50%. The magnitude spectrum of the windowed data is calculated and the noise spectrum, estimated from segments of speech absence is subtracted off. The resulting spectrum is biased down by the noise spectrum. Secondary noise residual reduction is then applied to reduce musical noise. The noisy speech phase, stored earlier is added to the modified magnitude and the time waveform is synthesized from it. The synthesized short time waveforms are added with the overlap of 50% between adjacent frames to get the enhanced speech.

2.1 Additive noise model

Assuming a noisy speech signal $x(t)$ formed due to additive background noise $d(t)$ corrupting a clean speech $s(t)$. It can be discretized and mathematically represented as,

$$x(n) = s(n) + d(n) \tag{1}$$

and the Short Time Fourier Transform of $x(n)$ is,

$$X(p,k) = S(p,k) + D(p,k) \tag{2}$$

where,

$$X(p,k) = \sum_{m=-\infty}^{\infty} x(m)w(p-m)e^{-j\omega m} \tag{3}$$

and $w(n)$ = hanning window where, $X(p,k)$, $S(p,k)$ and $D(p,k)$ represents the k^{th} spectral component of p^{th} time window of $x(n)$, $s(n)$ and $d(n)$ respectively.

2.2 Spectral subtraction estimator

The spectral subtraction estimator $\hat{S}(p,k)$ is obtained by passing the clean speech through spectral subtraction filter $H(p,k)$.

$$\hat{S}(p,k) = H(p,k)X(p,k) \tag{4}$$

where,

$$H(p,k) = 1 - \frac{\mu(k)}{|X(p,k)|} \tag{5}$$

$$\text{and } \mu(k) = E[\{|D(p,k)|\}] \tag{6}$$

Substituting equations (5) and (6), equation (4) becomes,

$$\hat{S}(p,k) = [|X(p,k)| - \mu(k)]e^{j\theta_x(p,k)} \tag{7}$$

2.3 Spectral error

The difference between the estimator and clean speech is called as spectral error $\in (p,k)$ and is given by,

$$\in (p,k) = \hat{S}(p,k) - S(p,k) \tag{8}$$

$$\in (p,k) = D(p,k) - \mu(k)e^{j\theta_x} \tag{9}$$

Further processing done for reducing the spectral error's auditory effects include: 1) Biasing down the noisy speech spectrum, 2) Reduction of Residual noise, and 3) Signal attenuation during speech absence.

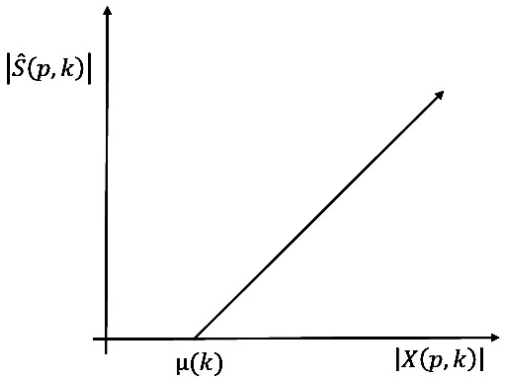


Fig. 2. Relation Between $X(p,k)$ and $\hat{S}(p,k)$.

2.4 Biasing down the noisy speech spectrum

For each value of k, if the estimated noise is greater than the noisy speech, the estimated speech magnitude is made zero. The estimator now becomes,

$$|\hat{S}(p,k)| = \begin{cases} |X(p,k)| - \mu(k) & \text{for } |X(p,k)| \geq \mu(k) \\ 0 & \text{for otherwise} \end{cases} \tag{10}$$

This is shown graphically in Fig. 2.

By this step, we make the noisy magnitude spectrum to be biased down at each frequency k by the noise bias determined at that frequency.

2.5 Reduction of residual noise

The difference $D_R = D - \mu\, e^{j\theta_D}$, which is populourly known as residual noise, will be present in the spectrum as narrow bands of magnitude spikes spaced randomly. In time domain, it sounds like group of tone generators, switched on and off at a frquency of 50 Hz with random fundamental frequencies. It can be eliminated by modifying the short time spectrum as given below:

$$|\hat{S}(p,k)| = \begin{cases} |\hat{S}(p,k)|; & |\hat{S}(p,k)| \geq \max|D_R| \\ \min[|\hat{S}(p,k)|p = i-1, i, i+1]; & |\hat{S}(p,k)| < \max|D_R| \end{cases} \tag{11}$$

where, $|D_R|$ = noise residual computed during speech absence.

2.6 Signal attenuation during speech absence

Even after biasing down the noisy spectrum and selecting minimum value, noise still remain at the non-speech activity durations. The parameter T, chosen to provide a means to classify a frame as speech or non-speech is defined as:

$$T = 20\log_{10}\left[\frac{1}{N}\sum_{k-1}^{N}\frac{\hat{S}(p,k)}{\mu(p,k)}\right] \tag{12}$$

where, N represents the FFT size or frame length.

It was observed that average power ratio was down at least 12 dB. If the frame was having speech activity, T would be greater than -12 dB. Before re-synthesis, we have three choices to select for the non-speech activity duration of

speech signal; to let as it is, or to scale down by a factor, or make it equal to zero. By subjective tests, we concluded that it is better to have some signal present during non-speech activity. Empirically, scaling factor was chosen to be c, whose value is 0.032. As a result, the output spectral estimate including output attenuation during non-speech activity is given by,

$$\hat{S}(p,k) = \begin{cases} \hat{S}(p,k); & T \geq -12\,\text{dB} \\ cX(p,k); & T \leq -12\,\text{dB} \end{cases} \tag{13}$$

2.7 Synthesis

After all the processing is done, noisy speech's phase was added to the processed magnitude spectrum and IFFT was done to get the short time-domain signal. As the short time frames are earlier divided with an overlap of 50%, the synthesized short time frames are added appropriately with the overlap of 50% to construct the whole speech signal.

3. Wiener Filtering

A number of methods are available for estimating the coefficients of clean speech, one such method based on MMSE estimation such as Wiener filter. If the noise is independent and additive with respect to speech, the minimization of $E\{(\hat{S}(p,k) - S(p,k))^2\}$ leads to,

$$G(p,k) = \frac{E\{|S(p,k)|^2\}}{E\{|S(p,k)|^2\} + E\{|D(p,k)|^2\}} = \frac{\widehat{SNR}_{\text{prio}}\,(p,k)}{1 + \widehat{SNR}_{\text{prio}}\,(p,k)}$$

As in[2], for practical implementations, the estimation of the *a priori* SNR, $\widehat{SNR}_{\text{prio}}\,(p,k)$ is considered, which is required for the computation of $G(p,k) \cdot \widehat{SNR}_{\text{prio}}\,(p,k)$ is frequently estimated using the DD approach.

In[4], analysis on behavior of the estimator was given and it was proved that the *a priori* SNR of current frame follow the *a posteriori* SNR of previous frame. As a Consequence, the desirable behavior of the spectral gain was not achieved. For refining the *a priori* SNR estimate, TSNR method was utilized. The second step here ensures that annoying reverberation effect of DD approach is removed by suppressing bias due to frame delay while its ability to reduce musical noise level is not lost.

Moreover, all classic techniques for short-time noise suppression suffer from harmonic distortion. These techniques consider some harmonics as noise-only components and suppress them. The error arises from the estimation of noise spectrum which is a tough task in case of uni-channel noise suppression techniques.

To regenerate the removed harmonics, we went for HRNR which takes speech's harmonic characteristic into account. An artificial signal with missing harmonics is produced by processing the output of any classic noise suppression technique. Using this artificial signal, *a priori* SNR was modified and used for computing a spectral gain which can restore the speech signal harmonics is formulated.

3.1 Introduction to noise reduction parameters

Consider the digitized noisy speech given by $x(n) = s(n) + d(n)$ where $s(n)$ and $d(n)$ denote the digitized speech and noise signals respectively. Let $S(p,k)$, $D(p,k)$ and $X(p,k)$ are same as considered in the Section 2. Using noisy features, we compute SNR estimates, and use them to get a spectral gain $G(p,k)$. We apply this $G(p,k)$ to each $X(p,k)$ to obtain an estimate of $S(p,k)$. The techniques implemented for speech enhancement required computation of two parameters: *a priori* SNR, and *a posteriori* SNR defined by:

$$\text{SNR}_{\text{prio}}(p,k) = \frac{E[|S(p,k)|^2]}{E[|D(p,k)|^2]} \tag{14}$$

$$\text{SNR}_{\text{post}}(p,k) = \frac{|X(p,k)|^2}{E[|D(p,k)|^2]} \tag{15}$$

where, the E is the expectation operator. Another parameter, the instantaneous SNR is also defined, as:

$$\mathrm{SNR}_{\mathrm{inst}}(p,k) = \frac{|X(p,k)|^2 - E[|D(p,k)|^2]}{E[|D(p,k)|^2]} \tag{16}$$

$$\mathrm{SNR}_{\mathrm{inst}}(p,k) = \mathrm{SNR}_{\mathrm{post}}(p,k) - 1 \tag{17}$$

Practically, the noisy speech spectrum $X(p,k)$ alone is available, the PSDs of speech $E[|S(p,k)|^2]$ and noise $E[|D(p,k)|^2]$ are unknown. Therefore, we need to estimate both $\mathrm{SNR}_{\mathrm{post}}(p,k)$, and $\mathrm{SNR}_{\mathrm{prio}}(p,k)$. The noise PSD estimate $E[|D(p,k)|^2]$, noted as $\widehat{\gamma}_d(p,k)$, was done using classic recursive relation. The spectral gain $G(p,k)$ is then obtained by

$$G(p,k) = g(\mathrm{SNR}_{\mathrm{post}}(p,k), \mathrm{SNR}_{\mathrm{prio}}(p,k)) \tag{18}$$

The function g is chosen to be wiener filtering in our work and the estimate of speech signal is obtained as,

$$\hat{S}(p,k) = G(p,k)X(p,k) \tag{19}$$

3.2 DD approach

The derivation of $\mathrm{SNR}_{\mathrm{prio}}(p,k)$ here, is based on its definition, and its relation to $\mathrm{SNR}_{\mathrm{post}}(p,k)$, as given below:

$$\mathrm{SNR}_{\mathrm{prio}}(p,k) = \frac{E[|S(p,k)|^2]}{E[|D(p,k)|^2]} \tag{20}$$

$$\mathrm{SNR}_{\mathrm{prio}}(p,k) = E[\mathrm{SNR}_{\mathrm{post}}(p,k) - 1] \tag{21}$$

Adding equations (20) and (21), we can write

$$\mathrm{SNR}_{\mathrm{prio}}(p,k) = E\left\{\frac{1}{2}\frac{[|\hat{S}(p-1,k)|^2]}{E[|D(p,k)|^2]} + \frac{1}{2}[\mathrm{SNR}_{\mathrm{post}}(p,k) - 1]\right\} \tag{22}$$

The proposed estimator $\widehat{SNR}_{\mathrm{prio}}(p,k)$ of $\mathrm{SNR}_{\mathrm{prio}}(p,k)$ is deduced from (23), and is given by

$$\mathrm{SNR}^{\mathrm{DD}}_{\mathrm{prio}}(p,k) = \alpha\frac{[|\hat{S}(p-1,k)|^2}{E[|D(p,k)|^2]} + (1-\alpha)P[\mathrm{SNR}_{\mathrm{post}}(p,k) - 1] \quad \text{for} \quad 0 \le \alpha \le 1 \tag{23}$$

where $|\hat{S}(p-1,k)|^2$ is the amplitude estimator of the k^{th} spectral component of the $(p-1)^{\mathrm{th}}$ frame, and the function $P[\cdot]$ is defined as

$$P[x] = \begin{cases} x & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

This *a priori* SNR estimator as given in equation (23) corresponds to the DD approach and is referred to as $\widehat{SNR}^{\mathrm{DD}}_{\mathrm{prio}}(p,k)$. The behaviour of $\widehat{SNR}^{\mathrm{DD}}_{\mathrm{prio}}(p,k)$ is controlled by α, a parameter whose typical value is 0.98. $G(p,k)$ In equation (19) was chosen to be the Wiener filter, which results in

$$G_{\mathrm{DD}}(p,k) = \frac{\widehat{SNR}^{\mathrm{DD}}_{\mathrm{prio}}(p,k)}{1 + \widehat{SNR}^{\mathrm{DD}}_{\mathrm{prio}}(p,k)} \tag{25}$$

In DD method, two effects which can stressed out and were interpreted in[4] are:

- *For large instantaneous SNR*: If $\widehat{SNR}_{\mathrm{inst}}(p,k) \gg 0\,\mathrm{dB}$, $\widehat{SNR}_{\mathrm{prio}}(p,k)$ is equal to the $\widehat{SNR}_{\mathrm{inst}}(p,k)$ delayed by a frame.
- *For smaller instantaneous SNR*: When $\widehat{SNR}_{\mathrm{inst}}(p,k)$ is small or closer to 0 dB, $\widehat{SNR}_{\mathrm{prio}}(p,k)$ matches to a highly smoothened and delayed version of $\widehat{SNR}_{\mathrm{inst}}(p,k)$. Thusly the variance of $\widehat{SNR}_{\mathrm{prio}}(p,k)$ is reduced compared to $\widehat{SNR}_{\mathrm{inst}}(p,k)$.

The main setback for the DD algorithm is the delay that is inherent and its effect, particularly during the speech transitions, i.e. onset and offset. This delay leads to a bias in estimating gain, which results in a reverberation effect.

3.3 TSNR method

In the pursuit of improving noise suppression performance, we went on to implement a method in two steps for estimating the *a priori* SNR, known as TSNR method. In the DD algorithm, musical noise was greatly reduced when the parameter α is selected to be 0.98, we would not want to disturb the elimination of musical noise and hence one of the two steps was exactly same as the one we did in DD algorithm. The second step must be able to eliminate the delay which led to the problem discussed as the drawback of the DD algorithm. So, the spectral gain that is calculated for the frame $(p+1)^{\text{th}}$ in first step is applied to p^{th} frame of noisy speech to get the enhanced p^{th} frame. The two steps are mathematically given as: $S\widehat{N}R_{\text{prio}}^{\text{TSNR}}(p,k) = S\widehat{N}R_{\text{prio}}^{\text{DD}}(p+1,k)$

$$S\widehat{N}R_{\text{prio}}^{\text{TSNR}}(p,k) = \beta' \frac{|G_{\text{DD}}(p,k)X(p,k)|^2}{\hat{\gamma}_n(p,k)} + (1-\beta')P[S\widehat{N}R_{\text{post}}(p+1,k)-1] \tag{26}$$

where, the role of β' is same as that of α but we can choose some other value. We can observe that to calculate $S\widehat{N}R_{\text{post}}(p+1,k)$, information about $X(p+1,k)$, the next frame is needed and because of this an additional delay is introduced. Thence, we preferred to choose $\beta' = 1$. With this modification, equation (26) now becomes:

$$S\widehat{N}R_{\text{prio}}^{\text{TSNR}}(p,k) = \frac{|G_{\text{DD}}(p,k)X(p,k)|^2}{\hat{\gamma}_n(p,k)} \tag{27}$$

This way, we avoided the additional processing delay as the information about future frame is not needed. Moreover the first step ensures that the level of musical noise is minimized to the least obtained by the DD approach. In the end, Wiener filtering was used to calculate the gain as:

$$G_{\text{TSNR}}(p,k) = \frac{S\widehat{N}R_{\text{prio}}^{\text{TSNR}}(p,k)}{S\widehat{N}R_{\text{prio}}^{\text{TSNR}}(p,k)} \tag{28}$$

The gain is then multiplied with the noisy speech spectrum to get clean speech spectrum estimate,

$$\hat{S}(p,k) = G_{\text{TSNR}}(p,k)X(p,k) \tag{29}$$

Highlighting two important characteristics of TSNR can be as follow:

- *For large instantaneous SNR*: If $S\widehat{N}R_{\text{inst}}(p,k) \gg 0\,\text{dB}$, $S\widehat{N}R_{\text{prio}}(p,k)$ is equal to the $S\widehat{N}R_{\text{inst}}(p,k)$ without any delay as contrary to the DD algorithm.
- *For smaller instantaneous SNR*: When $S\widehat{N}R_{\text{inst}}(p,k)$ is smaller or closer to 0 dB, the $S\widehat{N}R_{\text{prio}}^{TSNR}(p,k)$ is furthermore decreased compared to $S\widehat{N}R_{\text{prio}}^{\text{DD}}(p,k)$

In summary, the noise suppression performance was improved by the TSNR algorithm, thanks to the second step which made sure that the gain at a particular frame matches to itself irrespective of SNR, as contrary with the DD approach. The preservation of speech transitions, i.e. onset and offset, and success in removing the annoying reverberation effect as in DD approach are two reasons that made TSNR an obvious select for removing background noise of additive type.

3.4 Speech harmonic regeneration

The difficulty in getting accurate noise estimates in single channel noise suppression techniques leads to estimation errors. This leads for the spectrum estimate $\hat{S}(p,k)$, or time domain waveform $\hat{s}(t)$, obtained by techniques like DD, and TSNR to be suffering from distortions. Most of the distortions were found out to be harmonic in nature. Indeed some of the harmonics were considered by the algorithms as noise-only components and were undesirably suppressed.

For preventing the distortion, we processed the distorted signal and an artificial signal whose frequency response is similar to a harmonic comb which has the harmonics that were missing in the distorted signal was produced. The artificial signal was used to calculate a spectral gain that is capable of restoring harmonics.

This step can be implemented by simply applying a non-linear function to the time domain signal $\hat{s}(t)$, as given by

$$s_{\text{harmo}}(t) = NL(\hat{s}(t)) \tag{30}$$

It can be observed that the positions where harmonics of $s_{\text{harmo}}(t)$ will be present are exactly same as that of the clean speech ones, but with biased amplitudes. Hence it was used only for refining the *a priori* SNR:

$$\widehat{SNR}_{\text{prio}}^{\text{HRNR}}(p,k) = \frac{\rho(p,k)|\hat{S}(p,k)|^2 + (1-\rho(p,k))|S_{\text{harmo}}(p,k)|^2}{\hat{\gamma}_d(p,k)} \tag{31}$$

where, $\rho(p,k) = G_{\text{TSNR}}(p,k)$.

$\widehat{SNR}_{\text{prio}}^{\text{HRNR}}(p,k)$ was then used for computing a gain that is capable of preserving harmonics. As the harmonics that were removed by earlier speech enhancement technique are restored, the reconstructed speech after HRNR has all the harmonics as the clean speech and hence will sound natural. The spectral gain is obtained as,

$$G_{\text{HRNR}}(p,k) = \frac{\widehat{SNR}_{\text{prio}}^{\text{HRNR}}(p,k)}{1+\widehat{SNR}_{\text{prio}}^{\text{HRNR}}(p,k)} \tag{32}$$

and $\hat{S}(p,k)$ was computed as:

$$\hat{S}(p,k) = G_{\text{HRNR}}(p,k)X(p,k) \tag{33}$$

4. Results

All the algorithms discussed above have been studied and implemented and their behavior have been analyzed for different kinds of background noises such as babble noise, car noise, added to clean speech with different SNRs. The spectrogram of the clean speech considered in the experiment is shown in Fig. 3.

4.1 Spectrographic analysis

To illustrate the behavior and performance of the implemented techniques, spectrograms after each step are plotted as shown in the Fig. 4(a–e).

The Fig. 4(a) shows the spectrogram of the noisy speech signal where we can observe the noise in yellow color distributed all over the spectrogram. In the Fig. 4(b), enhanced speech using spectral subtraction has removed much of the noise. Fig. 4(c) shows the enhanced speech using Wiener filtering with DD approach wherein we can find that along with noise removal, some of the harmonics are removed. From the Fig. 4(d), it can be observed that noise removal is better with TSNR approach than the one with DD approach, but harmonics are still not preserved. The Fig. 4(e) shows that spectrogram of enhanced speech using Wiener filtering with TSNR combined with HRNR has the ability to restore missing harmonics.

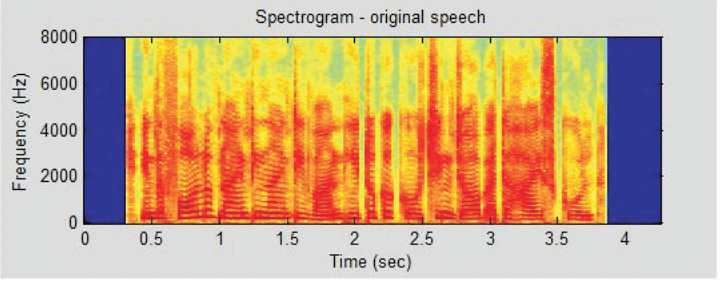
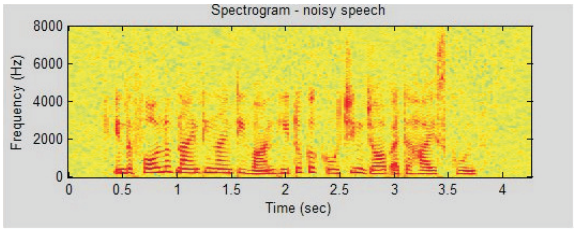
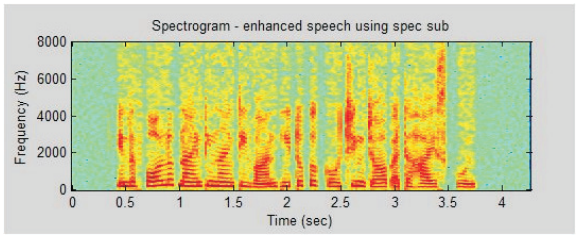


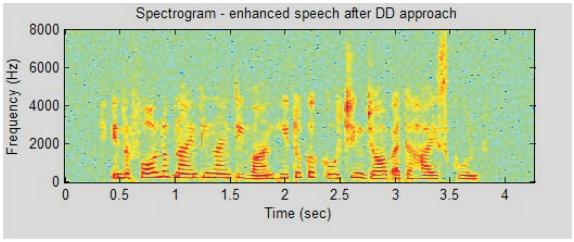
Fig. 3. Spectrogram of the Clean Speech Signal.



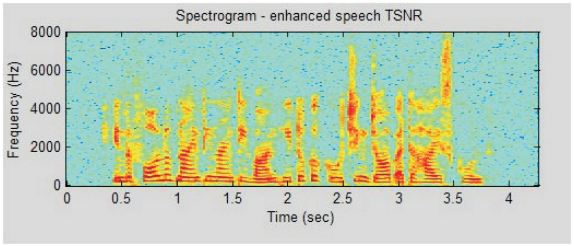
(a)



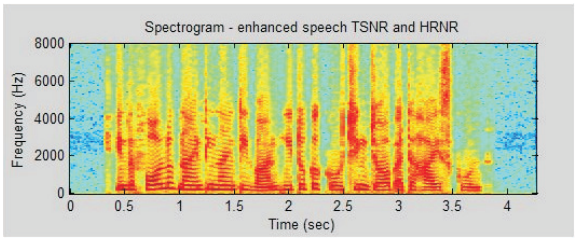
(b)



(c)



(d)



(e)

Fig. 4. Spectrogram of (a) Noisy Speech; Enhanced Speech using (b) Spectral Subtraction; (c) DD Approach; (d) TSNR Method and (e) TSNR Method Followed by HRNR Method.

Table 1. Segmental SNRs Calculated with Different Noise Types, Input SNR Values for Various Techniques Implemented.

Noise Type	Input SNR (dB)	Spectral Subtraction	DD Approach	TSNR Method	TSNR and HRNR
	−5	8.55	6.96	8.73	9.7
	−2	10.6	9.73	11.21	12.03
White noise	0	11.82	11.7	12.83	13.40
	2	13.09	13.68	14.35	45.27
	5	14.76	16.57	16.26	17.61
	−5	1.68	1.67	4.14	4.19
	−2	3.16	2.9	4.99	5.22
Helicopter noise	0	4.48	3.99	5.83	6.13
	2	5.46	5.25	6.79	7.09
	5	7.27	7.41	8.55	8.76
	−5	0.41	0.29	1.55	1.74
	−2	0.92	0.75	2.09	2.21
Babble noise	0	1.47	1.25	2.59	2.66
	2	2.25	1.99	3.42	.39
	5	3.58	3.42	4.78	4.79
	−5	1.60	1.59	1.90	1.91
	−2	2.8	2.71	3.01	3.10
Car noise	0	3.99	4.02	4.38	4.44
	2	5.42	5.28	5.58	5.76
	5	7.05	7.27	7.36	7.70

Table 2. Average Scores after the Subjective Test.

Noise Type	Parameters	Input Global SNR (dB)				
		-5	-2	0	2	5
White noise	Musical noise	4.1	4.5	4.9	5	5
	Intelligibility	4.6	4.7	4.8	5	5
	Quality	3.2	3.7	4.2	4.7	5
Helicopter noise	Musical noise	3	3.3	3.6	3.9	4.1
	Intelligibility	3.2	3.2	3.7	3.9	4.1
	Quality	1.6	2.3	2.8	3.4	4
Babble noise	Musical noise	2.7	3.1	3.4	3.8	3.9
	Intelligibility	2.4	2.9	3.1	3.4	3.9
	Quality	2.4	2.8	3.2	3.8	4.1
Car noise	Musical noise	2.5	3	3.6	3.9	4.1
	Intelligibility	3	3.1	3.4	3.9	4.4
	Quality	1.9	2.7	3.1	3.6	3.8

4.2 Objective results

For measuring the performance of the implemented techniques, we chose to calculate the average segmental SNR, given by,

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{M}\sum_{m=0}^{M-1} 10\ \log_{10}\ \frac{\sum_{l=Lm}^{Lm+L-1} s^2(l)}{\sum_{l=Lm}^{Lm+L-1} [\hat{s}(l) - s(l)]^2} \tag{34}$$

where, M denotes number of frames with active speech, and L represents frame length. For different noise types and SNR values, the segmental SNR was computed by aligning the clean and reconstructed signals in time and neglecting phase errors. For stationary as well as non-stationary noises the HRNR combined with TSNR technique achieves the best segmental SNRs as shown in the Table 1.

4.3 Formal subjective test

The objective results were confirmed by conducting a formal subjective test. Parameters in the implemented algorithms were selected with a fair balance among quality and intelligibility. We conducted the test with 5 listeners,

who were asked to listen to enhanced speech using different techniques randomly, and then were asked give scores from 1 to 5 for the parameters listed in the table. A score of 1 represents poor and 5 represents excellent. Table 2 provides the average scores for enhanced speech using TSNR combined with HRNR, while the scores for other techniques were considerably poorer and hence are not given in the paper.

5. Conclusions

We have presented an analysis of different noise reduction techniques and evaluated their performance for different noise types and SNRs. In spectral subtraction, as an estimate of noise spectrum is computed from segments of speech absence, and is subtracted from noisy speech spectrum, the method is not efficient for speech corrupted with non-stationary noise such as car noise, babble noise, helicopter noise. Another method is Wiener filtering in which multiplicative gain is calculated as a function of *a priori* SNR. In DD method, the frame delay and resulting reverberation effect leaves us in pursuit of a better method.

The TSNR technique is then applied to resolve the drawback of DD method. In this algorithm consisting of two steps, first step ensures musical noise reduction whereas the second step ensures the removal of frame delay, preserving the speech transitions.

TSNR performed well in terms of reducing noise but introduces harmonic distortion due to errors in estimating noise PSD. To resolve this problem, that is to restore any missing harmonics efficiently, a non-linearity was used in time domain to generate an artificial signal. The artificial signal was used for refining the *a priori* SNR, using which a spectral gain that avoids the distortion was computed.

Results, in terms of spectrographic analysis, objective, and subjective tests are given for evaluation of performance of various techniques. All the results demonstrate that TSNR followed by HRNR technique has the best performance among the others analyzed in terms of both the objective and subjective tests.

References

[1] S. F. Boll, Suppression of Acoustic Noise in Speech using Spectral Subtraction, *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. ASSp-27, no. 2, pp. 113–120, April (1979).
[2] Y. Ephraïm, and D. Malah, Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, December (1984).
[3] O. Capp'e, Elimination of the Musical Noise Phenomenon with the Ephra¨ım and Malah Noise Suppressor, *IEEE Transactions on Speech Audio Processing*, vol. 2, no. 2, pp. 345–349, April (1994).
[4] P. Scalart and J. Vieira Filho, Speech Enhancement Based on *a Priori* Signal to Noise Estimation, *IEEE International Conference on Acoustics, Speech, Signal Processing*, Atlanta, GA, USA, vol. 2, pp. 629–632, May (1996).
[5] C. Plapous, C. Marro, P. Scalart and L. Mauuary, A Two-Step Noise Reduction Technique, *IEEE International Conference on Acoustics, Speech, Signal Processing*, Montral, Qu'ebec, Canada, vol. 1, pp. 289–292, May (2004).
[6] C. Plapous, C. Marro and P. Scalart, Speech Enhancement using Harmonic Regeneration, *IEEE International Conference on Acoustics, Speech, Signal Processing*, Philadelphia, PA, USA, vol. 1, pp. 157–160, March (2005).
[7] R. Martin, Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics, *IEEE Transactions on Speech Audio Processing*, vol. 9, no. 5, pp. 504–512, July (2001).
[8] R. F. Kubichek, Standards and Technology Issues in Objective Voice Quality Assessment, *Digital Signal Processing*, vol. 1, pp. 38–44, (1991).
[9] ITU-T Recommandation, Telephone Transmission Quality – Objective Measuring Apparatus, pp. 56, March (1996).
[10] ITU-T Recommendation, Methods for Subjective Determination of Transmission Quality, pp. 800, August (1996).