

<sup>1</sup>Zegar Chouki<sup>2</sup>Dahimene Abdelhakim

## Comparative Study on Noisy Speech Preprocessing Algorithms



Received:05/02/2024

Published:28/05/2024

**Abstract:** - Speech Enhancement and noise reduction have wide applications in speech processing. They are often employed as pre-processing stage in various applications. The work to be presented in this paper is denoising a single-channel speech signal at the presence of a highly non-stationary background noise in order to improve the perceptible quality and intelligibility of the speech. Real world noise is mostly highly non-stationary and does not affect the speech signal uniformly over the spectrum. This paper investigates various Discrete Fourier Transform-based algorithms as single-channel pre-processing techniques consisting of: Spectral Subtraction using over-subtraction and spectral floor, Multi-Band Spectral Subtraction (MBSS), Wiener Filter, MMSE of Short-Time Spectral Amplitude (MMSE-STSA) estimator with, and without using SPU modifier, MMSE Log-Spectral Amplitude Estimator with, and without using SPU modifier, Optimally-Modified Log-Spectral Amplitude estimator (OM-LSA). The processed speeches from these algorithms are compared at the same set of conditions using visual examinations of signals in the time domain and the spectrograms, and also the objective and subjective tests for quality and perceptual evaluation. All the implemented algorithms provide considerable, different degrees of flexibility and control on noise elimination levels that reduces artifacts in the enhanced speech, resulting in the improved quality, and intelligibility.

**Keywords:** Speech denoising, non-stationary noise, single channel.

### I. INTRODUCTION

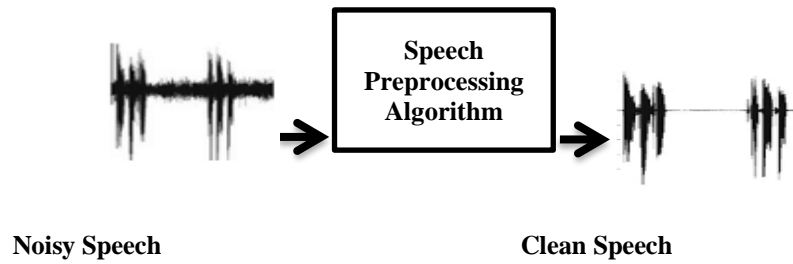
Development and widespread deployment of digital communication systems during the last twenty years have brought increased attention to the role of speech enhancement in speech processing problems. The degradation of the quality and intelligibility of speech signals, due to the presence of background noise severely affects the ability of speech related systems to perform well. Speech enhancement algorithms are used to improve the performance of communication systems when their input or output signals are corrupted by noise. The main objective of speech enhancement or noise reduction is to improve the perceptual aspects of speech, such as the speech quality and intelligibility. However, the problem of cleaning noisy speech still poses a challenge to the area of signal processing. Noise reduction techniques have some problems and questions. One of these problems is to reach a compromise between noise reduction, signal distortion, and the residual musical noise. Complexity and ease of implementation of the speech enhancement algorithms is also of concern in applications especially those related to portable devices such as mobile communications and digital hearing aids. The DFT-based speech enhancement methods have been one of the most well-known techniques for noise reduction. The spectral subtraction estimates the power spectrum of clean speech by explicitly subtracting the noise power spectrum from the noisy speech power spectrum. Due to its minimal complexity and relative ease in implementation, it has enjoyed a great deal of attention over the past years. This approach generally produces a residual noise commonly called musical noise. In this paper, we investigate DFT-based single-channel speech enhancement algorithms as speech signal pre-processing approaches at highly non-stationary noise.

### II. DFT-BASED TECHNIQUES FOR SINGLE CHANNEL SPEECH ENHANCEMENT

This part describes short time DFT-based single channel techniques for additive noise removal. These methods are based on the analysis-modify-synthesis approach. They use fixed analysis window length (usually 20-32ms) and frame by frame based processing. They are based on the fact that human speech perception is not sensitive to spectral phase but the clean spectral amplitude must be properly extracted from the noisy speech to have acceptable quality of speech at output and hence they are called short time spectral amplitude (STSA) based methods. Figure 1 shows the basic overview of a single-channel speech enhancement system.

<sup>1</sup> \* ZEGAR Chouki: Laboratory of Signals and Systems, M'Hamed Bougara University of Boumerdes, c.zegar@univ-boumerdes.dz

<sup>2</sup> Laboratory of Signals and Systems, M'Hamed Bougara University of Boumerdes, a.dahimene@univ-boumerdes.dz



**Figure 1:** Basic overview of single channel speech enhancement system

Most real world noise such as street noise, train station noise, restaurant noise, babble noise...etc. are non-stationary in nature. In the additive noise model the noisy speech is assumed to be the sum of the clean speech and the noise as defined in the following equation:

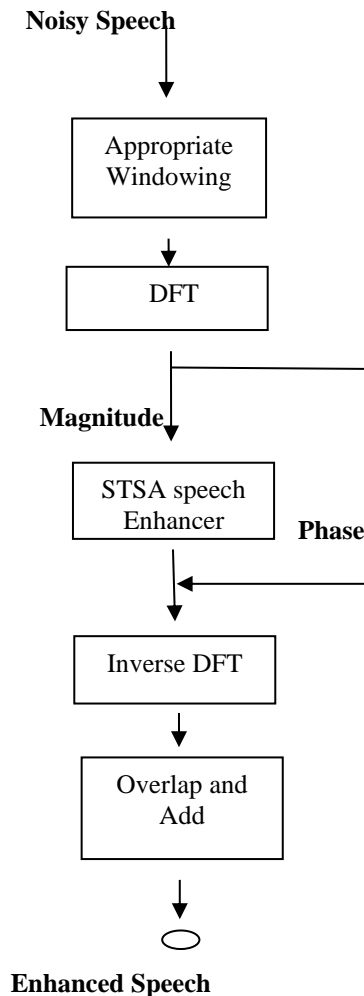
$$Y(t) = x(t) + n(t) \tag{1}$$

Where  $y(t)$  is the noisy speech signal,  $x(t)$  is the clean speech signal, and  $n(t)$  is the background noise signal.

Let  $y[n] = x[n] + d[n]$  be the sampled observed noisy speech signal consisting of the clean signal  $x[n]$  and the noise signal  $d[n]$  where,  $0 \leq n \leq N - 1$ , and  $N$  is the frame length.

*A. General Structure of DFT-Based Speech Enhancement*

The overall structure of the DFT-based speech enhancement techniques is shown in Figure 2.



**Figure 2:** Block diagram of the DFT-based speech enhancement [1].

*B. Noise power spectrum estimation*

Noise spectrum estimation is a challenging task for single-channel speech enhancement, where we have only the noisy speech available at the input. Non-stationary noise spectrum varies rapidly over time, hence it needs to be estimated and updated continuously.

In this paper, we use the algorithm proposed in [2] for estimating highly non-stationary noise environments.

*C. Spectral subtraction*

Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the noise spectrum from the noisy signal spectrum [3].

The first detailed treatment of spectral subtraction was performed by Boll [4], [5]. After that, papers [6], [7] expanded and generalized Boll's method to power subtraction, Wiener filtering.

Spectral subtraction algorithm is derived under Gaussian assumption for each spectral component.

*1. Power spectral subtraction and its generalized form*

The basic power spectral subtraction (PSS) principle involves the subtraction of the estimated noise variance, from the power spectrum of the observed noisy signal, to obtain an estimate of the modulus of speech power spectrum (taking into account that  $|\hat{X}_k|^2$  has to be positive). Mathematically, this is represented as:

$$|\hat{X}_k|^2 = \max(|Y_k|^2 - |\hat{D}_k|^2, 0) \tag{2}$$

However, there are limitations to this subtraction rule. The basic problem has been tackled by deriving several fundamentally and theoretically justified noise suppression rules.

Since the power spectrum of two uncorrelated signals is additive. By generalizing the exponent from 2 to  $a$ , Eq. (2) becomes

$$|\hat{X}_k|^a = \max(|Y_k|^a - |\hat{D}_k|^a, 0) \tag{3}$$

The speech phase is estimated directly from the noisy signal phase. Thus a general form of the estimated speech in frequency domain can be written as:

$$\hat{X}_k = (\max(|Y_k|^a - \alpha|\hat{D}_k|^a, 0))^{\frac{1}{a}} \cdot \exp(j\theta_k) \tag{4}$$

Where  $\alpha > 1$  is used to overestimate the noise to account for the variance in the noise estimate. The inner term  $|Y_k|^a - \alpha|\hat{D}_k|^a$  is limited to positive values, since it is possible for the overestimated noise to be greater than the current signal [8].

*2. Spectral subtraction using over-subtraction and spectral floor*

For more residual musical noise reduction, a modification of the spectral subtraction was proposed by Berouti et al [9]. The technique could be expressed as:

$$|\hat{X}_k|^2 = \max(|Y_k|^2 - \alpha \cdot |\hat{D}_k|^2, \beta \cdot |\hat{D}_k|^2) \tag{5}$$

Where:  $\alpha$  is the over-subtraction factor, and it is given in terms of the frame noisy signal to noise ratio as follows:

$$\alpha = \alpha_0 - \frac{3}{20} \cdot SNR \quad -5dB \leq SNR \leq +20dB \tag{6}$$

$\alpha_0$  is the desired value of  $\alpha$  at 0 dB SNR.

$\alpha$  plays the role of a time-varying factor, which provides a degree of control over the noise removal process between periods of noise update.

The parameter  $\beta$  is the spectral floor which prevents the spectral components of the enhanced spectrum from being below the smallest value  $\beta \cdot |\widehat{D}_k|^2$ . In this case  $\beta$  plays the role of controller (the amount of remaining residual noise and the amount of perceived musical noise).

### 3. Multi-Band Spectral Subtraction (MBSS)

The idea of MBSS method proposed by [10] starts from the fact that the colored noise has different effects at the various frequencies of the speech spectrum. The MBSS technique performs spectral subtraction with different over subtraction factor in different non-overlapped frequency bands. The spectral subtraction rule in  $i^{th}$  frequency band is given by:

$$|\widehat{X}_{k,i}|^2 = \begin{cases} |\overline{Y}_{k,i}|^2 - \delta_i \alpha_i \cdot |\widehat{D}_{k,i}|^2, & \text{if } |\overline{Y}_{k,i}|^2 > \delta_i \alpha_i \cdot |\widehat{D}_{k,i}|^2 \\ \beta \cdot |\overline{Y}_{k,i}|^2 & \text{else} \end{cases}$$

for  $b_i \leq k \leq e_i$  (7)

Where the spectral floor parameter was set to  $\beta = 0.002$ , and  $b_i$  and  $e_i$  are the beginning and ending frequency bins of the  $i^{th}$  frequency band.

$\overline{Y}_{k,i}$  is the  $i^{th}$  frequency band of smoothed and averaged version of the noisy speech spectrum. A weighted spectral average is taken over preceding and succeeding frames of speech as follows:

$$\overline{Y}_{k,j} = \sum_{l=-M}^M W_l Y_{k,j-l}$$

(8)

Where  $j$  is the frame index, and  $0 < W_l < 1$ . The averaging is done over  $M$  preceding and succeeding frames of speech.

The number of frames  $M$  is limited to 2 to prevent smearing of the speech spectral content. The weights  $W_l$  were empirically determined and set to  $W_l = [0.09, 0.25, 0.32, 0.25, 0.09]$  for  $-2 \leq l \leq +2$  [10].

The band-specific over-subtraction factor  $\alpha_i$  is a function of the segmental  $SNR_i$  of the  $i^{th}$  frequency band, which is calculated as:

$$SNR_i(dB) = \left[ \frac{\sum_{k=b_i}^{e_i} |Y_{k,i}|^2}{\sum_{k=b_i}^{e_i} |\widehat{D}_{k,i}|^2} \right]$$

(9)

$\alpha_i$  can be expressed in terms of  $SNR_i$  (defined previously) as follows:

$$\alpha_i = \begin{cases} 4.75 & SNR_i < -5 \\ 4 - \frac{3}{20} SNR_i & -5 \leq SNR_i \leq 20 \\ 1 & SNR_i > 20 \end{cases}$$

(10)

The additional over subtraction factor  $\delta_i$  called tweaking factor provides additional degree of control in each frequency band. The values of this factor are empirically determined and set according to following equation (Usually 4-8 linearly spaced frequency bands are used).

$$\delta_i = \begin{cases} 1 & f_i < 1 \text{ KHz} \\ 2.5 & 1 \text{ KHz} \leq f_i \leq \frac{F_s}{2} - 2 \text{ KHz} \\ 1.5 & f_i > \frac{F_s}{2} - 2 \text{ KHz} \end{cases}$$

(11)

Where  $f_i$  is the upper frequency of the the  $i^{th}$  band, and  $F_s$  is the sampling frequency [10].

D. *Wiener filter*

In terms of our speech enhancement problem the Wiener filter proposed in [11] is given by:

$$|\hat{X}_k| = \frac{\xi_k}{\xi_k + 1} |Y_k| \tag{12}$$

Where  $\xi_k$  is defined as the a priori SNR found by Decision Directed Method.

E. *MMSE of short time spectral amplitude*

Ephraim and Malah [12] formulated an optimal spectral amplitude estimator, which, specifically, estimates the modulus (magnitude) of each complex Fourier coefficient of the speech signal in a given analysis frame from the noisy speech in that frame.

In order to derive the MMSE STSA estimator, the a priori probability distribution of the speech and noise Fourier expansion coefficients should be assumed since these are unknown in reality. Ephraim and Malah [12] assumed that the Fourier expansion coefficients of each process can be modeled as statistically independent Gaussian random variables, real and imaginary parts of each component is independent to each other, and the mean of each coefficient is assumed to be zero and the variance time-varying.

1. *Gaussian based MMSE-STSA estimator*

The desired gain function for the MMSE-STSA estimator, [12]:

$$G_{MMSE}(v_k) = \Gamma(1.5) \frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right) \cdot \left[ (1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right] \tag{13}$$

Where  $\Gamma(\cdot)$  is the Gamma function (with  $\Gamma(1.5) = \sqrt{\pi}/2$ ) and  $I_0(\cdot)$  and  $I_1(\cdot)$  are the zeroth and first order modified Bessel functions, respectively, defined as:

$$I_n(z) = \frac{1}{2\pi} \int_0^{2\pi} \cos(\beta n) \exp(z \cos \beta) d\beta \tag{14}$$

In Eq. (2.30),  $v_k$  is defined as:

$$v_k = \frac{\xi_k}{\xi_k + 1} \gamma_k \tag{15}$$

Where  $\xi_k$  and  $\gamma_k$  are defined by:

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \tag{16}$$

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \tag{17}$$

$\xi_k$  and  $\gamma_k$  are interpreted as the a priori and a posteriori signal-to-noise ratios (SNR), respectively.  $R_k$  denotes the spectral magnitude of the noisy signal.

Essentially, a priori SNR is the Signal-to-Noise Ratio of the  $k^{th}$  spectral component of the “clean” speech signal,  $x[n]$ , while a posteriori SNR is the  $k^{th}$  spectral component of the corrupted signal,  $y[n]$ . Computation of  $\gamma_k$  is straightforward ratio of the variance of the noisy speech signal to the estimated noise variance. However, computation of a priori SNR is more involved, especially since the knowledge of “clean” signal is seldom available in real systems. In this paper “Decision-Directed” estimation [12] has been exploited to compute a priori SNR.

2. *Amplitude estimator under Speech Presence Uncertainty SPU*

Signal absence in noisy observations  $\{y[n], 0 \leq n \leq N\}$  is frequent, as speech signals generally contain large portions of silence, [12]. Nevertheless, it does not mean that speech is never present in noisy sections.

The idea of utilizing the uncertainty of signal presence in the noisy spectral components for improving speech enhancement results was first proposed by McAulay and Malpass [6], [12].

The MMSE estimator which accounts for uncertainty of speech presence in noisy observation was first developed by Middleton and Esposito, [12], [13] and it is based on the model of statistically independent random appearance of signal in noisy spectral components.

In order to derive the new amplitude estimator, we need to calculate the generalized likelihood ratio  $\Lambda(Y_k, q_k)$  while  $q_k$  denotes the *a priori* probability of speech absence in the  $k^{th}$  spectral component.

$$\Lambda(Y_k, q_k, \xi'_k) = \frac{1 - q_k}{q_k} \frac{\exp\left(\frac{\xi'_k}{1 + \xi'_k} \gamma_k\right)}{1 + \xi'_k} \quad (18)$$

Where  $\xi'_k$  is the conditional a priori SNR:

$$\xi'_k \triangleq E\{A_k \setminus H_1^k\} \quad (19)$$

$$\xi'_k = \frac{1}{1 - q_k} \xi_k \quad (20)$$

F. *Speech enhancement using MMSE Log Spectral Amplitude estimator*

Based on [14] Malah and Ephraim proposed a new short time spectral amplitude (STSA) estimator for speech signals which minimizes the mean squared error of the log spectra.

This section will briefly discuss the derivation of the minimum mean squared error log spectral amplitude (MMSE-LSA).

In order to derive MMSE-LSA, Malah and Ephraim used the same formulation of the estimation problem and the same statistical model as in [12] (modeling speech and noise spectral components as statistically independent Gaussian random variables).

With the same definitions for a priori and a posteriori SNR (discussed during the MMSE-STSA derivation), the desired MMSE-LSA gain function is given as follows [14]:

$$G_{MMSE-LSA}(\xi_k, \gamma_k) = \frac{\xi_k}{1 + \xi_k} \left\{ \frac{1}{2} \int_{\nu_k}^{\infty} \frac{e^{-t}}{t} dt \right\}, \quad (21)$$

Where  $\nu_k = \frac{\xi_k}{\xi_k + 1} \gamma_k$  as shown previously during MMSE-STSA estimator derivation.

The MMSE-LSA estimator may be also modified using the multiplicative gain  $G_{SPU}(k)$  defined previously for the MMSE-STSA estimator.

G. *Speech enhancement using the Optimally Modified Log Spectral Amplitude estimator (OM-LSA)*

The purpose of this section is to study the Optimally-Modified Log-Spectral Amplitude estimator (OM-LSA) proposed by I. Cohn [15]. As the name suggests, it estimates  $\hat{A}_k$  by minimizing mean-squared error of the log-spectra for speech signals under signal presence uncertainty where the spectral gain function is obtained as a weighted geometric mean of the hypothetical gains associated with signal presence and absence.

In this algorithm, Cohen [15] proposed two important estimators:

- An estimator for the a priori signal-to-noise ratio.
- An efficient estimator for the a priori speech absence probability (SAP) which is based on the time-frequency distribution of the a priori SNR.

1. *Optimal gain modification*

Let  $H_0^k$  and  $H_1^k$  designate respectively hypothetical speech absence and presence in the  $k^{th}$  frequency bin, and assuming a complex Gaussian distribution of the STFT coefficients for both speech and noise [12]:

- Null Hypothesis  $H_0^k$  : *speech absent*:  $Y_k = D_k$
- Alternate Hypothesis,  $H_1^k$  : *speech present*:  

$$Y_k = X_k + D_k$$

The LSA estimator for the clean speech spectral amplitude (Assuming statistically independent spectral components [14]), which minimizes the mean-squared error of the log spectra, is given by:

$$\widehat{A}_k = \frac{\exp\{E[\ln A_k | Y_k, 0 \leq k \leq N - 1]\}}{G_{k \text{ OM-LSA}} |Y_k|} \quad (22)$$

The Optimally Modified LSA estimator gain is given by:

$$G_{k \text{ OM-LSA}} = \{G_{H1}(\xi'_k, \gamma_k)\}^{p_k} \cdot G_{min}^{1-p_k}, \quad \text{where } 0 \leq k \leq N - 1 \quad (23)$$

Where:  $G_{H1}(\xi'_k, \gamma_k) = \frac{\xi'_k}{1+\xi'_k} \left\{ \frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt \right\}$   $\xi'_k$ : *a priori SNR*,  $\gamma_k$ : *a posteriori SNR*,

and  $v_k = \frac{\xi'_k}{\xi'_{k+1}} \gamma_k$ .

$$p_k = P(H_1^k | Y_k), \quad 0 \leq k \leq N - 1.$$

When speech is absent, the spectral gain is constrained to be larger than a threshold  $G_{min}$ , which is determined by subjective criteria for the noise naturalness [15]. Hence,

$$\exp\{E[\ln A_k | Y_k, H_0^k]\} = G_{min} \cdot |Y_k| \quad (24)$$

When speech is present, we use Ephraim and Malah's MMSE-LSA estimator [14]:

$$\exp\{E[\ln A_k | Y_k, H_1^k]\} = G_{k \text{ H1}} \cdot |Y_k| \quad (25)$$

Where,  $G_{k \text{ H1}}$  is defined (as defined previously in eq (21) by:

$$G_{H1}(\xi'_k, \gamma_k) = \frac{\xi'_k}{1 + \xi'_k} \left\{ \frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt \right\}, \quad (26)$$

where  $\xi'_k$ : *a priori SNR*,  $\gamma_k$ : *a posteriori SNR*, and  $v_k = \frac{\xi'_k}{\xi'_{k+1}} \gamma_k$ .

By substituting (24) and (25) into (22), The Optimally Modified LSA estimator gain is given by:

$$G_{k \text{ OM-LSA}} = \{G_{H1}(\xi'_k, \gamma_k)\}^{p_k} \cdot G_{min}^{1-p_k}, \quad \text{where } 0 \leq k \leq N - 1 \quad (27)$$

2. *Priori SNR estimation*

According to the decision-directed approach, proposed by Ephraim and Malah [12], it provides a useful estimation method for the non-conditional a priori SNR  $\xi_k$  which is given by:

$$\widehat{\xi}_k(l) = \alpha \frac{\widehat{A}_k^2(l-1)}{\lambda_d(k, l-1)} + (1 - \alpha) \max\{\gamma_k(l) - 1, 0\} \quad (28)$$

where  $0 < \alpha < 1$ , and  $l$  is the frame number.

Therefore the estimate for the a priori SNR should be given by:  $\xi'_k = \frac{1}{1-q_k} \xi_k$ . According to this expression, there is an interaction between the estimated  $q_k$  and the a priori SNR which may deteriorate the performance of the speech enhancement system [16], [17], [18].

Hence, Cohen in [15] proposed a new estimator of the Priori SNR which is given as follows:

$$\widehat{\xi}'_k(l) = \alpha G_{H1}(l-1)^2 + (1-\alpha) \max\{\gamma_k(l) - 1, 0\} \tag{29}$$

To explain more this equation, if  $H_1^k$  is true then the spectral gain should degenerate to  $G_{H1}$ , and the a priori SNR  $\xi'_k$  estimate should coincide with  $\xi_k$ . In opposite, if  $H_0^k$  is true, then the spectral gain should decrease to  $G_{min}$ , or equivalently the a priori SNR estimate should be as small as possible[15].

### 3. Priori speech Absence probability (SAP) estimation

In [15], Cohen proposed a new estimator for the speech absence probability  $\widehat{q}_k$ . The estimator utilizes a soft-decision approach in order to find three parameters ( $P_{k_{local}}(l), P_{k_{global}}(l), P_{frame}(l)$ ) based on the time-frequency distribution of the estimated a priori SNR,  $\widehat{\xi}'_k(l)$ . These parameters exploit the strong correlation of speech presence in neighboring frequency bins of consecutive frames [15].

Let  $\xi'_k(l)$  be a recursive average of the a priori SNR:

$$\xi'_k(l) = \beta \xi'_k(l-1) + (1-\beta) \widehat{\xi}'_k(l-1) \tag{30}$$

where  $\beta$  is a time constant.

Local and global averaging windows are applied in the frequency domain to obtain local and global averages of the a priori SNR:

$$\xi'_{k\lambda}(l) = \sum_{i=-w_\lambda}^{w_\lambda} h_\lambda(i) \xi'_{k-i}(l) \tag{31}$$

Where the subscript  $\lambda$  designates either “local” or “global”, and  $h_\lambda$  is a normalized window of size  $2w_\lambda+1$ .

We define two parameters,  $P_{k_{local}}(l)$  and  $P_{k_{global}}(l)$ , which represent the relation between the above averages and the likelihood of speech in the  $k^{th}$  frequency bin of the  $l^{th}$  frame[15]. The local and global parameters are given by the following expression:

$$P_{k\lambda}(l) = \begin{cases} 0, & \text{if } \xi'_{k\lambda}(l) \leq \xi'_{min} \\ 1, & \text{if } \xi'_{k\lambda}(l) \geq \xi'_{max} \\ \frac{\ln(\xi'_{k\lambda}(l)/\xi'_{min})}{\ln(\xi'_{max}/\xi'_{min})}, & \text{otherwise} \end{cases} \tag{32}$$

Where  $\xi'_{min}$  and  $\xi'_{max}$  are empirical constants.

For more noise attenuation in noise-only frames, a third parameter named,  $P_{frame}(l)$  is defined. This parameter is based on the speech energy in neighboring frames. If we average  $\xi'_k(l)$  in the frequency domain we obtain:

$$\xi'_{frame}(l) = \text{mean}_{1 \leq k \leq \frac{N}{2}+1} \{\xi'_k(l)\} \tag{33}$$

Figure 3 shows the block diagram for  $P_{frame}(l)$  computation.

Where  $I(l)$  is given by:

$$I(l) \triangleq \begin{cases} 0, & \text{if } \xi'_{frame}(l) \leq \xi'_{peak}(l) \cdot \xi'_{min} \\ 1, & \text{if } \xi'_{frame}(l) \geq \xi'_{peak}(l) \cdot \xi'_{max} \\ \frac{\ln(\xi'_{frame}(l)/\xi'_{peak}(l)/\xi'_{min})}{\ln(\xi'_{max}/\xi'_{min})}, & \text{otherwise} \end{cases} \tag{34}$$

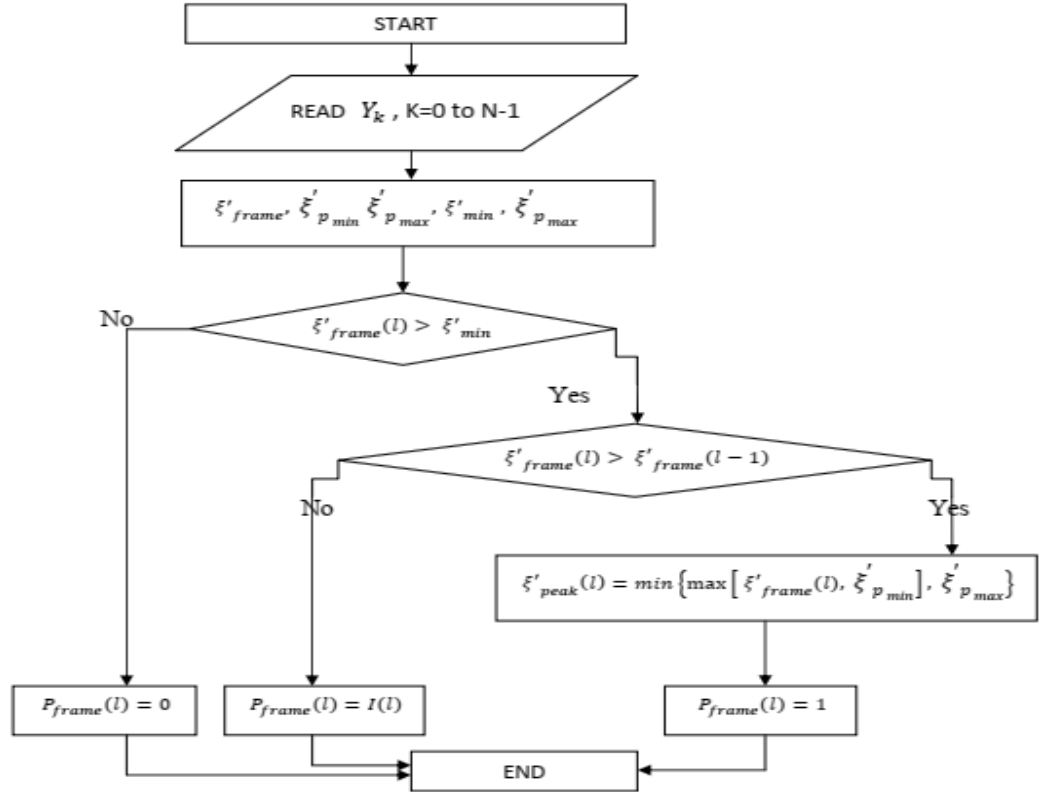
And it represents a soft transition from “speech” to “noise”,  $\xi'_{peak}$  is a confined peak value of  $\xi'_{frame}$ , and  $\xi'_{p_{min}}, \xi'_{p_{max}}$  are empirical constants that determine the delay of the transition [15].

Hence, the proposed estimate for the a priori probability for speech absence is obtained by:



$$\hat{q}_k(l) = 1 - P_{k_{local}}(l) \cdot P_{k_{global}}(l) \cdot P_{frame}(l) \quad (35)$$

$\hat{q}_k(l)$  is larger if either previous frames, or recent neighboring frequency bins, do not contain speech. In order to reduce the possibility of speech distortion we restrict  $\hat{q}_k(l)$  to be smaller than a threshold  $q_{max} (<1)$ .



**Figure 3:** Block diagram for  $P_{frame}(l)$  computation

### III. IMPLEMENTATION AND PERFORMANCE EVALUATION

This section describes the implementation details and performance evaluation of the proposed pre-processing algorithms to understand their functionality and behavior. Evaluation of speech enhancement algorithms is not simple. While objective quality assessment methods can indicate an improvement or degradation in speech quality based on mathematical measures, the human listener does not believe in a simple mathematical error criterion. Therefore, subjective measurements of intelligibility and quality are also required.

The IEEE standard database NOIZEUS (noisy corpus) [19] is used to test algorithms. The database contains clean speech sample files as well as real world noisy speech files at different SNRs and noise conditions like street, car, restaurant, train, station, babble...etc.

#### A. Implementation details

The factors contributing in the efficient implementation of some of the functional blocks are discussed below.

- Frame size: 20 ms was chosen as the optimum frame size for our implementations.
- Window Type and Overlap: the most commonly used Hamming window [7], [3]. After a few informal listening tests and comparing spectrograms, the Hamming window was chosen. The amount of overlap between consecutive frames is also associated with the frame-size, and is required to prevent discontinuities at frame

boundaries. For this study we chose the overlap to be 50%, which is also usually the percentage overlap found commonly in the literature.

- The enhanced signal is obtained by taking the IFFT of the enhanced spectrum using the phase of the original noisy spectrum.
- The standard overlap-and-add method is used to obtain the enhanced signal.

For the Spectral Subtraction using over-subtraction and spectral floor, the spectral floor parameter is set to  $\beta = 0.002$ , and  $\alpha = \begin{cases} 4.75 & SNR < -5 \\ 4 - \frac{3}{20}SNR & -5 \leq SNR \leq 20 \\ 1 & SNR > 20 \end{cases}$

For the Multi-Band Spectral Subtraction (MBSS) implementation, the spectral floor parameter is also set to  $\beta = 0.002$ , and all other parameters are taken as given in chapter two.

For the Wiener filter, MMSE-STSA, and MMSE-LSA algorithms implementations, the *a priori* SNR  $\xi_k$  is calculated using the Decision-Directed estimation approach with  $\alpha = 0.98$  in Eq. (29).

For Speech Presence Uncertainty (SPU) multiplicative modifier implementation, the *a priori* probability of speech absence  $q_k$ , is set to  $q_k = 0.3$ .

For the OM-LSA estimator implementation, the value  $\alpha = 0.92$ , and the values of parameters used for the estimation of the *a priori* SAP are given as follows:

$$\begin{aligned} \beta &= 0.7 & \xi'_{min} &= -10dB & \xi'_{max} &= -5dB \\ \xi'_{p_{min}} &= 0dB & \xi'_{p_{max}} &= 10dB \\ w_{local} &= 1 & w_{global} &= 15 \\ q_{max} &= 0.95 & h_\lambda &: \text{Hanning window} \end{aligned}$$

For the implementation of noise estimation algorithm discussed in section 2.2, the following parameters are used:

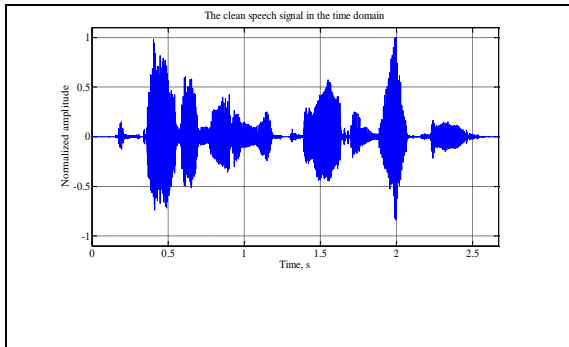
$$\begin{aligned} \eta &= 0.7, & \text{the threshold } \sigma &= 1.3, \\ \lambda &= 0.8, & \gamma &= 0.998, & \beta &= 0.8 & \alpha_1 &= 0.8, & \alpha_2 &= 1. \end{aligned}$$

### B. Visual examinations for the implemented algorithms

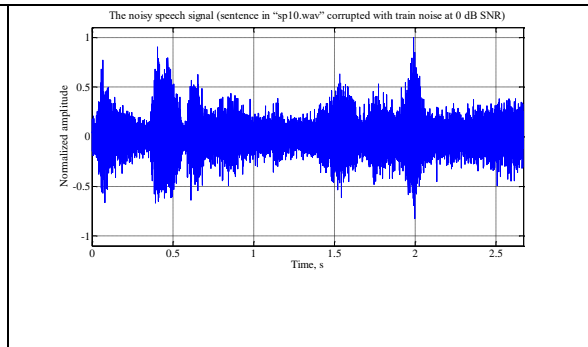
Applying the implemented algorithms to the noisy speech signal sentence in “sp10.wav” corrupted with train noise at 0 dB SNR, and car noise at 5 dB SNR yields to the results presented along with the original noisy signal in following figures:

From Figure 4 to Figure 13 show the signals in the time domain of the original sentence in “sp10.wav” along with the same corrupted with speech-shaped train noise at 0 dB SNR, and the enhanced speech obtained by the implemented algorithms.

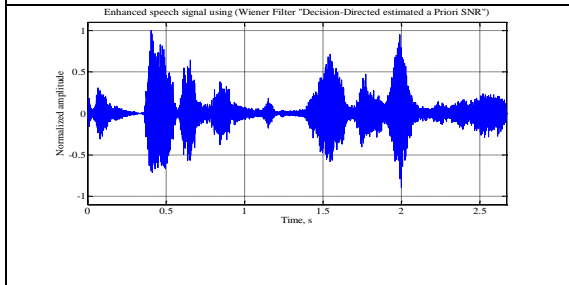
From Figure 14 to Figure 23 show the spectrograms of the original sentence in “sp10.wav” along with the same corrupted with speech-shaped car noise at 5 dB SNR, and the enhanced speech obtained from implemented algorithms.



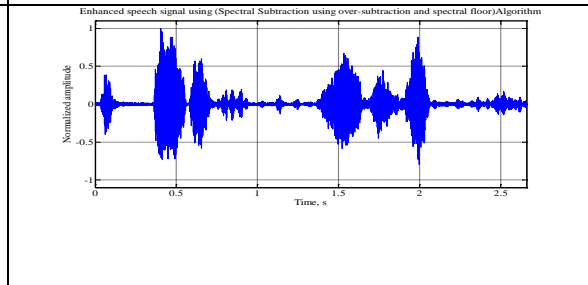
**Figure 4:** Clean speech signal in the time domain.



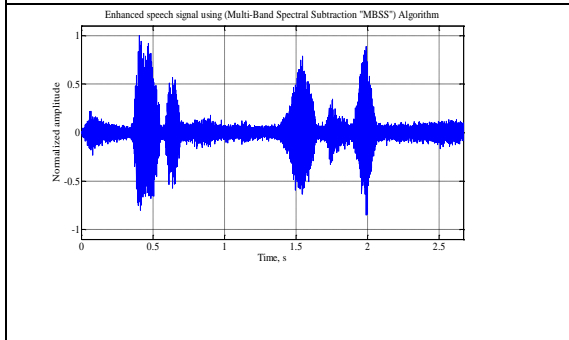
**Figure 5:** Noisy speech signal (sentence in “sp10.wav” corrupted with train noise at 0 dB SNR).



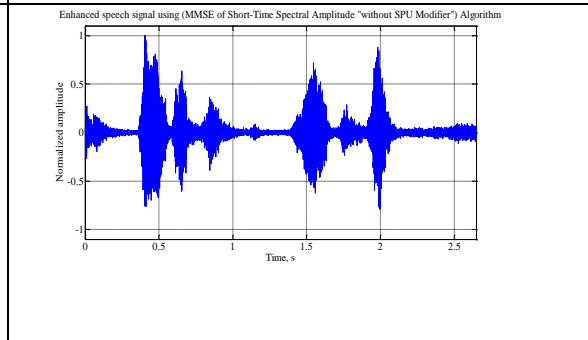
**Figure 6:** Enhanced speech signal using Wiener Filter (Decision-Directed estimated a priori SNR).



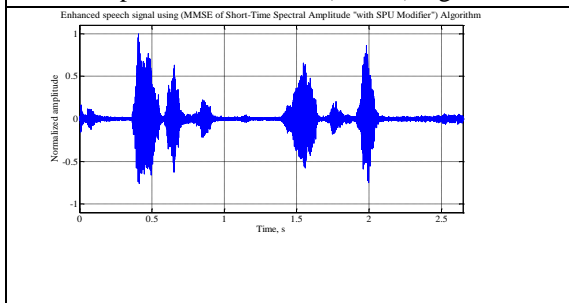
**Figure 7:** Enhanced speech signal using Spectral subtraction using over-subtraction and spectral floor algorithm



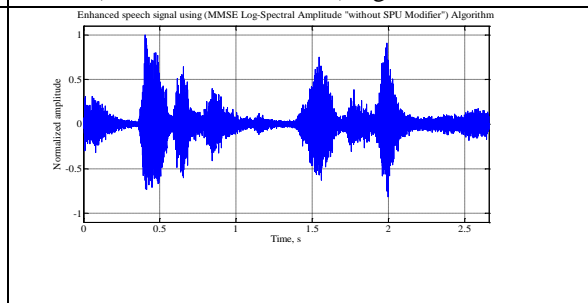
**Figure 8:** Enhanced speech signal using Multi-band spectral subtraction (MBSS) algorithm.



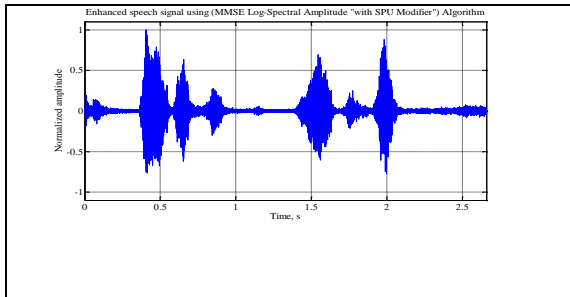
**Figure 9:** Enhanced speech signal using MMSE-STSA (without SPU modifier) algorithm.



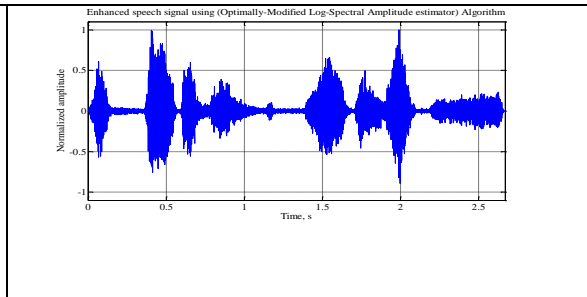
**Figure 10:** Enhanced speech signal using MMSE-STSA (with SPU modifier) algorithm.



**Figure 11:** Enhanced speech signal MMSE-LSA (without SPU modifier) Algorithm.



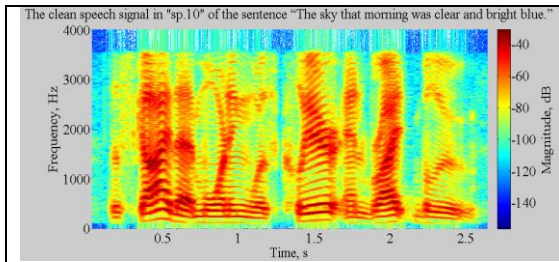
**Figure 12:** Enhanced speech signal using MMSE-LSA (with SPU modifier) algorithm.



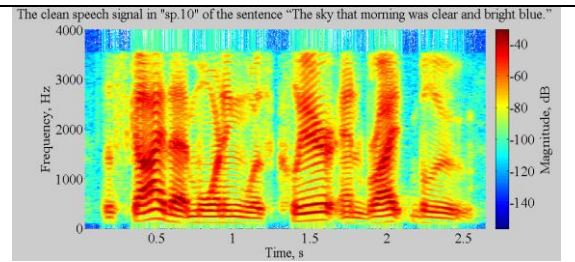
**Figure 13:** Enhanced speech signal using the optimally modified log-spectral Amplitude (OM-LSA) algorithm.

From visual examinations of figures presented above, we can notice that:

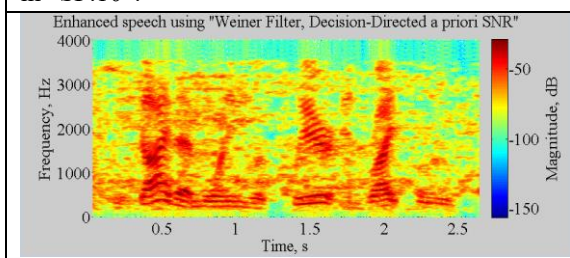
- A significant amount of noise has been reduced from the noisy speech signal after applying each of the DFT-based speech enhancement algorithms.
- The enhanced speech panels using Wiener filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method show more distortions in the shape of the signals when compared to the original clean speech signal.
- The enhanced speech panels using MMSE-STSA, MMSE-LSA (without SPU modifier) algorithms show small distortions in the shape of the signals when compared to the original clean speech signal.
- The enhanced speech panels using MMSE-STSA, MMSE-LSA (with SPU modifier), and the OM-LSA algorithms show that the obtained processed signal shapes are more nearer to the original clean speech signal.



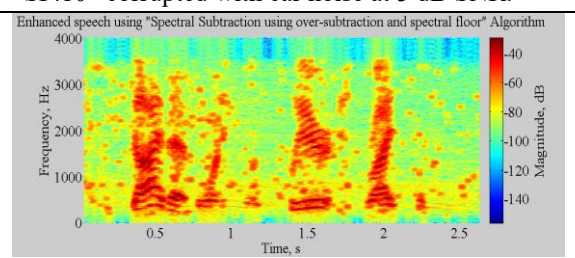
**Figure 14:** Spectrogram of the clean speech signal in "SP.10".



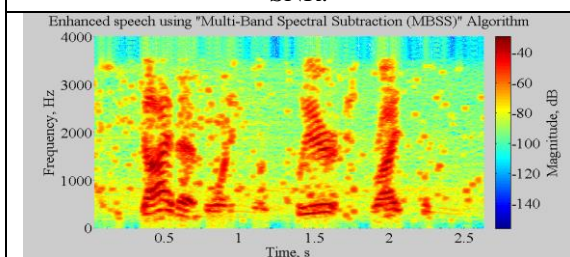
**Figure 15:** Spectrogram of the noisy signal in "SP.10" corrupted with car noise at 5 dB SNR.



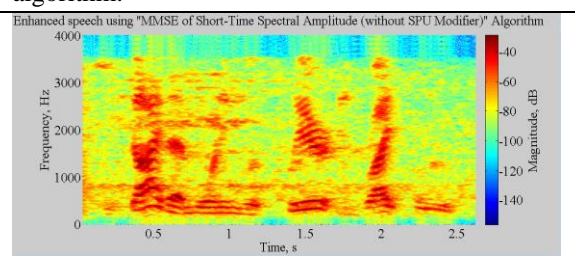
**Figure 16:** Spectrogram of the enhanced speech using Wiener Filter, Decision-Directed a priori SNR.



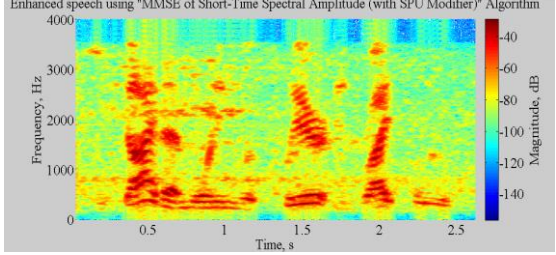
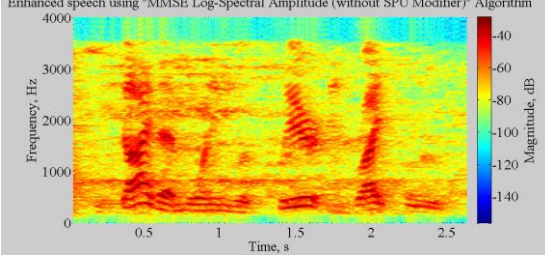
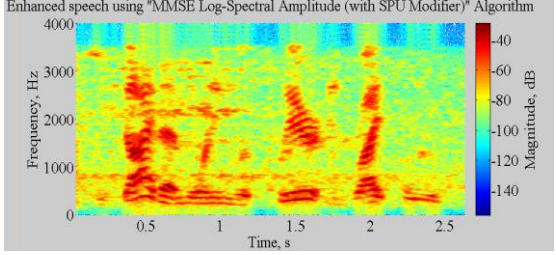
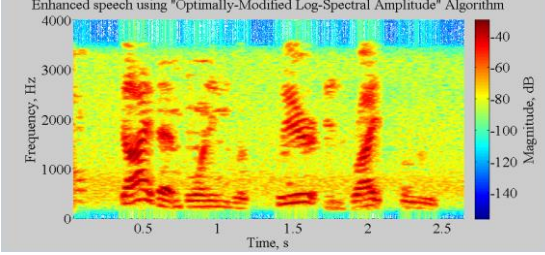
**Figure 17:** Spectrogram of the enhanced speech using Over-Subtraction and spectral floor algorithm.



**Figure 18:** Spectrogram of the enhanced speech



**Figure 19:** Spectrogram of the enhanced speech

using Multi-band spectral subtraction (MBSS) algorithm.	using the MMSE-STSA (without SPU modifier) algorithm.
	
<p><b>Figure 20:</b> Spectrogram of the enhanced speech using the MMSE-STSA (with SPU modifier) algorithm.</p>	<p><b>Figure 21:</b> Spectrogram of the enhanced speech using the MMSE-LSA (without SPU modifier) algorithm.</p>
	
<p><b>Figure 22:</b> Spectrogram of the enhanced speech using the MMSE-LSA (with SPU modifier) algorithm.</p>	<p><b>Figure 23:</b> Spectrogram of the enhanced speech using optimally modified log-spectral (OM-LSA) algorithm.</p>

From the visual examinations of the spectrograms in figures presented above, we can remark that:

- In all the enhanced speech spectrograms, the formants are much clearer and visible than in the noisy speech spectrogram, which indicates that there is a considerable amount of noise has been reduced from the noisy speech.
- The enhanced speech spectrograms using Wiener filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method contain some random isolated dots which cause audible artifact known as musical noise.
- The enhanced speech spectrograms using MMSE-STSA, and MMSE-LSA algorithms show better results, and less amount of isolated dots compared with the spectrograms of Wiener filter, MBSS method, and Spectral Subtraction (using over-subtraction and spectral floor) method.
- Applying the multiplicative SPU modifier to the MMSE-STSA, and MMSE-LSA algorithms provides more noise reduction in spectrograms.
- The enhanced speech spectrogram using OM-LSA algorithm is the nearest to the to the original clean speech spectrogram.

*C. Objective measures for implemented algorithms performance evaluation*

Objective measures are based on a mathematical comparison of the original and enhanced speech signals.

**Signal-to-Noise Ratio (SNR)**

As the name suggests, SNR is the ratio of the signal energy to the noise energy:

$$SNR_{dB} = 10 \cdot \log_{10} \left( \frac{\sum_n s^2[n]}{\sum_n (s[n] - \hat{s}[n])^2} \right) \tag{36}$$

Where  $s(n)$  is the clean signal and  $\hat{s}(n)$  is the processed signal. If the summation is performed over the whole signal length, the operation is called global SNR.

**Segmental Signal-to-Noise Ratio (  $SNR_{Seg}$  ):**

The SNRseg in dB is the average SNR computed over short frames of the speech signal. The SNRseg over M frames of length N is computed as:

$$SNR_{seg} = \frac{1}{M} \sum_{i=0}^{M-1} 10 \log_{10} \left[ \frac{\sum_{n=iN}^{iN+N-1} s^2[n]}{\sum_{n=iN}^{iN+N-1} (s[n] - \hat{s}[n])^2} \right] \text{ dB}, \quad (37)$$

In order to perform our objective tests, each algorithm is evaluated using all the sentences from NOIZEUS data base corrupted by 4 different SNR values (0, 5,10 and 15dB) in 6 colored noise environments which are as follows:

- Train
- Car
- Street
- Restaurant
- Train station
- Babble

In addition to that, a synthesized white noise added to clean speech sentences of NOIZEUS database at SNR range 0-15dB is also used to test the algorithms.

The results (all the obtained SNR and SNRseg values are averages of 30 measures, the number of sentences in the database and are given in dB) are shown in tables from 1 to 7.

<b>Table 1:</b> Objective quality evaluation of the implemented algorithms for train noise.									<b>Table 2:</b> Objective quality evaluation of the implemented algorithms for car noise								
Train noise	SNR=0 dB SNRseg ≅ -4.50		5 dB SNRseg ≅ -1.67		10 dB SNRseg ≅ 1.50		15 dB SNRseg ≅ 4.50		Car noise	SNR=0dB SNRseg ≅ -4.95		5 dB SNRseg ≅ -2.00		10 dB SNRseg ≅ 1.05		15 dB SNRseg ≅ 4.05	
Objective test	SNR	SNRseg	SNR	SNRseg	SNR	SNRseg	SNR	SNRseg	Objective test	SNR	SNRseg	SNR	SNRseg	SNR	SNRseg	SNR	SNRseg
Weiner DD	6.05	-0.20	8.77	1.51	13.01	4.19	16.16	8.00	Weiner DD	6.08	-0.3	10.17	2.32	13.83	5.19	17.37	8.50
SS using over-subtraction and spectral floor	6.14	-0.14	8.07	1.51	12	4.26	17.44	8.73	SS using over-subtraction and spectral floor	4.75	-0.4	9.46	1.83	13.47	4.90	18.75	9.00
Mband	6.40	-0.10	8.56	2.32	12.07	5.12	17.44	8.54	Mband	4.90	-0.2	9.40	2.33	13.00	5.20	17.30	9.05
MMSE-STSA	6.48	0.50	8.32	2.40	12.18	5.76	15.10	8.64	MMSE-STSA	5.16	0.50	9.53	2.80	13.46	5.30	16.51	9.10
MMSE-LSA	6.23	0.55	8.44	2.44	12.22	5.77	15.03	8.68	MMSE-LSA	5.34	0.55	9.38	2.80	13.23	5.33	16.54	9.10
MMSE-STSA using SPU modifier	6.10	1.20	8.00	3.11	11.81	5.80	15.32	8.70	MMSE-STSA using SPU modifier	4.67	0.80	9.21	3.24	13.28	5.41	16.15	9.12
MMSE-LSA using SPU modifier	6.31	1.25	8.24	3.10	12.09	5.81	15.53	8.73	MMSE-LSA using SPU modifier	6.05	0.90	9.49	3.25	13.42	5.43	15.01	9.15
OM-LSA	6.40	1.50	8.76	3.20	13.66	6.00	17.20	8.90	OM-LSA	6.10	1.30	10.48	3.50	14.43	5.55	18.38	9.20

<b>Table 3:</b> Objective quality evaluation of the implemented algorithms for street noise	<b>Table 4:</b> Objective quality evaluation of the implemented algorithms for restaurant noise
---	---

Street noise	SNR=0 dB SNR <sub>Seg</sub> ≅ -4.25		5 dB SNR <sub>Seg</sub> ≅ -1.20		10 dB SNR <sub>Seg</sub> ≅ 1.75		15 dB SNR <sub>Seg</sub> ≅ 4.70	
	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>
Objective test								
Weiner DD	4.00	-0.15	7.74	1.85	11.81	3.94	16.10	8.13
SS using using over-subtraction and spectral floor	4.12	-0.2	8.35	1.86	12.21	4.70	16.01	8.94
Mband	4.50	0.05	7.83	2.00	11.80	4.90	16.01	9.00
MMSE-STSA	5.02	0.40	7.96	2.30	12.17	5.00	15.02	9.01
MMSE-LSA	5.05	0.44	7.67	2.49	12.00	5.12	15.01	9.04
MMSE-STSA using SPU modifier	5.10	0.70	7.83	2.90	12.18	5.30	15.06	9.05
MMSE-LSA using SPU modifier	5.20	0.90	8.00	3.05	12.18	5.37	15.17	9.10
OM-LSA	5.57	1.26	8.72	3.40	13.00	5.50	16.24	9.12

Restaurant noise	SNR=0 dB SNR <sub>Seg</sub> ≅ -4.18		5 dB SNR <sub>Seg</sub> ≅ -1.15		10 dB SNR <sub>Seg</sub> ≅ 1.80		15 dB SNR <sub>Seg</sub> ≅ 4.80	
	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>
Objective test								
Weiner DD	4.00	-0.02	7.71	1.62	11.51	4.21	16.10	8.49
SS using using over-subtraction and spectral floor	4.02	-0.10	7.68	1.60	11.52	4.20	17.31	8.50
Mband	4.45	0.04	7.70	1.80	12.00	4.39	17.42	8.66
MMSE-STSA	5.05	0.10	8.28	1.90	11.90	4.80	15.48	8.68
MMSE-LSA	5.60	0.40	8.22	1.95	12.05	5.02	15.30	8.71
MMSE-STSA using SPU modifier	6.00	0.52	8.20	2.40	12.05	5.02	15.44	8.72
MMSE-LSA using SPU modifier	6.05	0.80	8.34	2.60	12.10	5.10	15.60	8.90
OM-LSA	6.12	1.22	8.56	3.12	12.40	5.30	17.43	9.30

**Table 5:** Objective quality evaluation of the implemented algorithms for train station noise.

Train station noise	SNR=0 dB SNR <sub>Seg</sub> =-4.7 dB		5 dB SNR <sub>Seg</sub> =-1.95 dB		10 dB SNR <sub>Seg</sub> = dB		15dB SNR <sub>Seg</sub> =4.60dB	
	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>
Objective test								
Weiner DD	4.98	0.30	9.06	2.01	12.50	4.31	15.23	8.60
SS using using over-subtraction and spectral floor	5.88	0.44	8.89	2.04	13.00	4.20	16.38	8.94
Mband	5.80	0.50	8.73	2.11	12.89	4.50	16.50	8.80
MMSE-STSA	5.86	0.61	8.92	2.42	13.02	4.65	15.58	8.95
MMSE-LSA	5.80	0.66	8.77	2.70	13.05	4.80	15.70	9.05
MMSE-STSA using SPU modifier	6.05	0.80	8.56	2.95	13.00	4.91	15.77	9.10
MMSE-LSA using SPU modifier	6.14	1.10	8.84	3.02	13.06	5.10	15.51	9.22
OM-LSA	6.20	1.66	9.84	3.50	13.5	5.70	16.54	9.61

**Table 6:** Objective quality evaluation of the implemented algorithms for babble noise.

Babble noise	SNR=0 dB SNR <sub>Seg</sub> ≅ -4.48		5 dB SNR <sub>Seg</sub> ≅ -1.50		10 dB SNR <sub>Seg</sub> ≅ 1.48		15 dB SNR <sub>Seg</sub> ≅ 4.40	
	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>
Objective test								
Weiner DD	4.00	-0.10	8.26	1.55	12.00	4.22	15.26	8.40
SS using using over-subtraction and spectral floor	4.69	-0.20	9.15	1.60	13.33	4.60	17.15	8.61
Mband	4.90	-0.05	8.90	2.00	13.20	4.80	17.50	8.70
MMSE-STSA	4.92	0.10	8.26	2.50	12.00	5.01	15.26	8.95
MMSE-LSA	4.90	0.25	7.90	2.68	12.05	5.10	15.31	9.05
MMSE-STSA using SPU modifier	5.50	0.30	8.56	2.80	12.03	5.33	15.20	9.11
MMSE-LSA using SPU modifier	5.40	0.60	8.44	3.01	12.07	5.41	15.02	9.20
OM-LSA	5.60	1.15	9.30	3.19	13.00	5.60	16.96	9.55

**Table 7:** Objective quality evaluation with a white noise.

white noise	SNR=0 dB		5 dB		10 dB		15 dB	
	SNR <sub>Seg</sub> $\cong$ 4.50	SNR <sub>Seg</sub> $\cong$ 1.3	SNR <sub>Seg</sub> $\cong$ 2.03	SNR <sub>Seg</sub> $\cong$ 4.2				
Objective test	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>	SNR	SNR <sub>Seg</sub>
Weiner DD	6.65	<b>1.02</b>	10.32	<b>3.00</b>	14.00	<b>5.34</b>	16.94	<b>8.00</b>
SS using using over- subtraction and spectral floor	5.99	<b>1.05</b>	9.71	<b>3.01</b>	13.23	<b>5.31</b>	18.32	<b>9.23</b>
Mband	6.60	<b>1.10</b>	10.20	<b>3.20</b>	13.55	<b>5.34</b>	18.00	<b>9.25</b>
MMSE-STSA	7.12	<b>1.50</b>	10.46	<b>4.00</b>	13.65	<b>6.40</b>	16.50	<b>8.61</b>
MMSE-LSA	6.86	<b>1.71</b>	10.20	<b>4.15</b>	13.54	<b>6.55</b>	16.35	<b>8.72</b>
MMSE-STSA using SPU modifier	6.68	<b>1.80</b>	10.10	<b>4.22</b>	13.24	<b>6.61</b>	16.14	<b>8.88</b>
MMSE-LSA using SPU modifier	7.01	<b>1.88</b>	10.34	<b>4.30</b>	13.47	<b>6.89</b>	16.22	<b>9.01</b>
OM-LSA	7.73	<b>2.00</b>	10.89	<b>4.42</b>	14.16	<b>7.00</b>	18.00	<b>9.80</b>

According to the objective test results presented above, we can observe the following:

- There are remarkable improvements in both global and segmental SNRs of noisy speech signals after being processed by the implemented DFT-based speech enhancement algorithms and the noise reduction is more when it is white.
- The speech enhancement using Wiener filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method provides less segmental SNR values when compared to the other implemented algorithms in most cases.
- The speech enhancement using MMSE-STSA, and MMSE-LSA algorithms provides better segmental SNR values, and using the SPU modifier gives a remarkable improvement in segmental SNRs.
- The speech enhancement using Optimally Modified Log-Spectral Amplitude estimator (OM-LSA) provides the best results (global SNR, and Segmental SNR) in most cases.

#### D. Subjective measures for implemented algorithms performance evaluation

Subjective tests rely heavily on the opinion of a group of listeners to judge the quality or intelligibility of processed speech. These tests are often time consuming as they require proper training of listeners. In addition to this, a constant listening environment (e.g., playback volume), identically tuned output device (e.g., headphones and/or speakers) are necessary. Nevertheless, subjective test results present the most accurate system of performance, insofar as intelligibility and speech quality are concerned, as they are determined perceptually by the human auditory system. These tests can be structured under two types of evaluation procedures: speech quality evaluation and also intelligibility testing. Quality refers to the clarity, freedom of distortion and ease for listening whereas Intelligibility refers to the number of words that can be identified correctly by a listener or to the likelihood of being correctly understood.

##### 1. Subjective test for speech quality evaluation

In this test, we asked five normal-hearing students who speak, and understand English very well to listen twice to the different samples of speech for each input SNR used in the previous objective tests. In the first time, we presented to them the signals in their noisy form, whereas in the second time we presented to them the processed ones. After that, we asked our listeners to grade each speech heard on a scale from 1 to 5, based on how pleasant their listening experience was, the highest grade corresponding to the most pleasant one. Then we averaged the respective grades and results are given in table 8.



**Table 8:** Subjective test for speech quality evaluation

Input SNR (dB)	0	5	10	15
Noisy speech grade	0.5	1.0	1.7	2.2
Weiner DD	1.9	2.1	2.6	2.8
SS using using over-subtraction and spectral floor	2.0	2.5	2.6	2.9
Mband	2.4	2.6	2.8	3.2
MMSE-STSA	2.5	2.7	3.0	3.1
MMSE-LSA	2.6	2.9	3.3	3.6
MMSE-STSA using SPU modifier	2.8	3.1	3.4	3.9
MMSE-LSA using SPU modofier	3.0	3.3	3.8	4.0
OM-LSA	3.4	3.8	4.3	4.5

According to this quality subjective test, we can say that:

- The quality of the noisy speech samples has been considerably improved after the enhancement by the implemented algorithms.
- The speech enhancement using Wiener filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method obtained the smallest grades for speech quality evaluation which confirm the annoying musical noise shown during the spectrograms visual examinations (random isolated dots).
- The speech enhancement using the OM-LSA algorithms provides the best speech quality.

2. *Subjective test for speech intelligibility evaluation*

In this test, we asked our listeners to give the percentage of intelligibility (according to the number of words that can be identified correctly by a listener) in the same speech signals. The results are shown below in table 9.

**Table 9:** Subjective test for intelligibility evaluation.

Input SNR (dB)	0	5	10	15
Noisy speech percentage	10%	30%	47%	58%
Weiner DD	53%	56%	66%	69%
SS using using over-subtraction and spectral floor	53%	60%	67%	72%
Mband	56%	67%	70%	75%
MMSE-STSA	65%	71%	74%	82%
MMSE-LSA	67%	73%	75%	85%
MMSE-STSA using SPU modifier	68%	73%	76%	88%
MMSE-LSA using SPU modofier	71%	74%	77%	90%
OM-LSA	73%	75%	80%	92%

According to table 9, we can say that:

- The implemented algorithms have considerably improved the intelligibility of the noisy speech signals.
- Wiener filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method provide the smallest percentages of intelligibility in comparison to the other implemented algorithms and that's due to the amount of distortions caused them.
- The optimally modified Log-Spectral Amplitude estimator (OM-LSA) algorithm shows the highest percentages of intelligibility.

### E. Comments

The implemented algorithms performance evaluation based on visual examinations, objective and subjective tests show that the optimally modified Log-Spectral Amplitude estimator (OM-LSA) algorithm outperforms all the implemented algorithms (low signal distortion and the best amount of noise reduction). However, we would like to note the following:

- The global SNR is a poor estimator of subjective quality. A high SNR value, is thus, not necessarily indicative of good perceptual quality of the speech.
- The segmental SNR objective test is more related to the subjective tests.
- Weiner filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method show acceptable amounts of non-stationary noise reduction but produce some distortions in the shape of the enhanced speech signals.
- MMSE-STSA, and MMSE-LSA algorithms provide more non-stationary noise reduction and less distortions.
- Applying the SPU multiplicative modifier with MMSE-STSA, and MMSE-LSA algorithms increases the quality and the intelligibility of the enhanced speech signals.

## IV. CONCLUSION

The work presented in this paper addressed the problem of single-channel speech enhancement at the presence of highly non-stationary background noise, as pre-processing stage for various speech applications.

A set of DFT-based single-channel speech enhancement algorithms have been implemented using highly non-stationary noise estimator, and each implemented algorithm has been evaluated using the NOIZEUS data base corrupted by 4 different SNR values (0, 5,10 and 15dB) in six colored noise environments (train, car, street, restaurant, train station, and babble) and a synthesized white noise.

The performance evaluation results establish the superiority of the Optimally-Modified Log-Spectral Amplitude estimator (OM-LSA) algorithm over all the implemented DFT-based single-channel speech enhancement algorithms with respect to perceptible quality and intelligibility improvements of the enhanced speech signals. Therefore, OM-LSA can be considered as good pre-processing technique for single-channel speech applications. MMSE-STSA, MMSE-LSA (using SPU multiplicative modifier) algorithms provide acceptable levels of speech intelligibility and quality in most cases and the second one behaves a little bit better than MMSE-STSA especially in reducing the musical noise. Weiner filter, Spectral Subtraction (using over-subtraction and spectral floor) method, and MBSS method show more distortions in the shape of the enhanced signals at low SNRs (0-5dB) range in most cases.

In addition to all the obtained results, we may say that, the most suitable technique for speech enhancement is the one which provides robustness to environmental noise contributing factors and robustness to acoustical inputs.

The works on implementing the DFT-based techniques for single-channel speech enhancement as pre-processing stages for various speech applications should definitely continue considering the good results we managed to achieve. Here is a short list of items that we think could be subjected to further studies:

- Investigating the speech enhancement using Laplacian-based MMSE estimator of the magnitude spectrum rather than MMSE estimator, which is based on a Gaussian model.
- The error between the processed signal and the clean speech signal can be strongly minimized if the estimate of the noise spectrum is more accurate. Hence, it is desirable to estimate the noise signal at every available instant to get a more accurate estimate of the noise spectrum.
- Single channel blind source separation is a challenging task. Hence working on this provides a good contribution to speech signals pre-processing techniques.

## ACKNOWLEDGMENT

This work is supported by Laboratory of Signals and Systems, Institute of Electrical and Electronic Engineering, M'Hamed Bougara University of Boumerdes, Algeria.

## REFERENCES

- [1] S.China Venkateswarlu, Dr. K.Satya Prasad, Dr. A.SubbaRami Reddy., "Improve Speech Enhancement Using Weiner Filtering," Global Journals Inc. (USA), Volume 11 Issue 7 Version 1.0 May 2011.
- [2] Rangachari, S. and Loizou, P. (2006). A noise estimation algorithm for highly non-stationary environments. *Speech Communication*, 28, 220-231
- [3] S. V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*. John Wiley and Sons, Ltd, 3<sup>rd</sup> Edition Book, pages 24, 246,247, 298, 2006.
- [4] Boll, S.F., "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. on Acoust., Speech, Signal Proc.*, Vol. ASSP-27, No.2, pp.113-120, April 1979.
- [5] S. F. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, (Washington, DC), pp. 200–203, Apr. 1979.
- [6] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-28, pp. 137–145, Apr. 1980.
- [7] Lim, J.S. and Oppenheim, A.V., "Enhancement and bandwidth compression of noisy speech", *Proc. IEEE*, Vol. 67, No.12, pp. 1586-1604, December 1979.
- [8] Erhan Deger, "Noise thresholding with empirical mode decomposition for low distortion speech enhancement," University of Tokyo, Master's thesis, pages 2-8, Feb 2008.
- [9] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Procs.*, pp. 208- 211, Apr. 1979.
- [10] Kamath, S. and Loizou, P. (2002). A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*.
- [11] Scalart, P. and Filho, J. (1996). Speech enhancement based on a priori signal to noise estimation. *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 629-632.
- [12] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.
- [13] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 434–444, May 1968 H. L.
- [14] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, No 2, pp. 443–445, April. 1985.
- [15] Cohen "Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator," *Lamar Signal Processing Ltd*.2003.
- [16] Y. Soon, S. N. Koh and C. K. Yeo, "Improved Noise Suppression Filter Using Self Adaptive Estimator of Probability of Speech Absence," *Signal Processing*, vol. 75, pp. 151–159, 1999
- [17] R. Martin, I. Wittke and P. Jax, "Optimized Estimation of Spectral Parameters for the Coding of Noisy Speech" in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP-2000*, pp. 1479–1482.
- [18] Cohen and B. Berdugo, "Speech Enhancement for Non-Stationary Noise Environments," to appear in *Signal Processing*.
- [19] Hu, Y. and Loizou, P. (2007). "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, 49, 588-601.