Speech Signal Enhancement Techniques

Chouki Zegar, Abdelhakim Dahimene

الملخّـص

لتعزيز نوعية الإشارة الكلامية والحدّ من الضّوضاء تطبيقات واسعة في مجال المعالجة الآليّة للكلام، وغالبا ما تستخدم باعتبارها مرحلة تمهيدية أو سابقة للمعالجة في مختلف التّطبيقات.

العمل الذي سنقوم به من خلال هذا المشروع يتضمن التّالية : تقليل نسبة الضوضاء الملوّنة الخلفيّة المُضمّنة في إشارة الكلام من خلال قناة واحدة لتحسين نوعيّ محسوس، وكلام واضح ببنية جيّدة.

معظم الضّوضاء الموجودة تأتي في شكل «الضّوضاء الملونة» ذات التّأثير غير المنتظم في إشارات الكلام المستعملة على طول مدى الطيف. سندرس ستّ خوارزميّات تتّصل بإشارة الكلام، من خلال قناة واحدة : الطرح الطيفي، الطرح الطيفي متعدد الطبقات، مصفاة وينر، أدنى معدّل خطإ تربيعيّ لسعات طيفية في مدات زمنية قصيرة (باستعمال المعدّل الحساس ودونه لوجود قطع كلامية)، أدنى معدل خطإ تربيعي للوغارتم السّعات الطيفية (باستعمال المعدّل الحساس ودونه لوجود قطع كلامية)، الخوارزمية الفائقة المعدلة للوغارتم السعات الطّيفية.

تتميز الطرق المقترحة بدرجة مرونة معتبرة في معالجة ومراقبة مستويات الضوضاء المحذوفة. من خلال النتائج المتحصل عليها تبين أنّ طريقة المعاجة التمهيدية المقترحة للخوارزمية الفائقة المعدّلة للوغارتم السعات الطّيفية تقدّم نتائج أحسن من الطرق الأخرى، اعتمادا على جملة من الاختبارات الموضوعية والذاتية، مختلفة المصادر، وجدنا أنّ هذه المصادر الكلامية فصلت جيّدا، وذلك تبعا لاختبارات التقييم المنجزة.

الكلمات المفاتيح : تعزيز نوعية الإشارة الكلامية، قناة واحدة، الضوضاء الملونة، الضوضاء الموسيقية، تقنية DFT، اختبارات التقييم الموضوعية والذاتية.

Speech Signal Enhancement Techniques

Chouki Zegar[1], Abdelhakim Dahimene[2]

[1,2] Institute of Electrical and Electronic Engineering, University of Boumerdes, Algeria

inelectr@yahoo.fr, dahimenehakim@yahoo.fr

Abstract

Speech enhancement and noise reduction have wide applications in speech processing. They are often employed as preprocessing stage in various applications. The work to be presented in this paper is denoising a single-channel speech signal in 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 explores a set of DFT-based algorithms as single-channel speech enhancement techniques which are as follows:

- 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 comparison study results based on subjective and objective tests showed that the Optimally Modified Log-Spectral Amplitude Estimator (OM-LSA) method outperforms all the implemented DFT-based single-channel speech enhancement algorithms.

Keywords: speech enhancement, single channel, non-stationary noise, musical noise, DFT-based techniques, evaluation tests.

1. Introduction

Development and widespread deployment of digital communication systems during the last decades 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 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 ques-

tions. 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. Due to their minimal complexity and relative ease in implementation, they have enjoyed a great deal of attention over the past years.

2. 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.

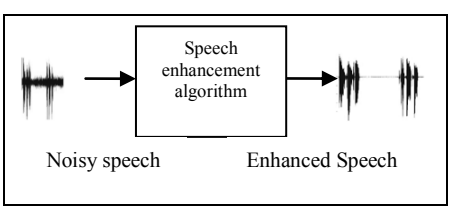

Figure 1: Basic overview of single channel speech enhancement system.

STSA based approaches assume that noise is additive and uncorrelated to the speech signal. 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 by:

$$y(t) = x(t) + d(t) \quad (1)$$

where y(t) is the noisy speech signal, x(t) is the clean speech signal, and d(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, $0 \le n \le N-1$, and N is the frame length. The additive noise model can be represented as shown in Figure 2.

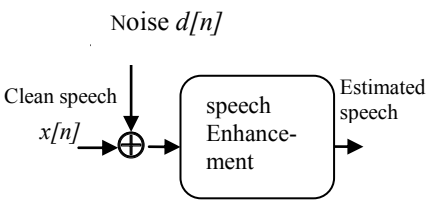

Figure 2: Additive noise model in single-channel speech enhancement [1].

2.1. The general structure of the DFT-based speech enhancement

The overall structure of the DFT-based speech enhancement techniques is shown in figure 3.

2.2. Spectral Subtraction using over-subtraction and spectral floor

For more residual musical noise reduction, a modification of the spectral sub-

traction was proposed by Berouti et al. [3]. The technique could be expressed as:

$$\left|\hat{X}_k\right|^2 = max\left(|Y_k|^2 - \alpha.\left|\hat{D}_k\right|^2, \beta.\left|\hat{D}_k\right|^2\right) \quad (2)$$

where α is the over-subtraction factor, and it is given in terms of the frame noisy signal to noise ratio as follows:

$$\alpha = \alpha_0 - \frac{3}{20}.SNR, \quad -5dB \le SNR \le +20dB \quad (3)$$

α_0 is the desired value of α at 0 dB SNR. α 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 β is the spectral floor which prevents the spectral components of the enhanced spectrum from being below the smallest value $\beta.\left|\hat{D}_k\right|^2$.

2.3. Multi-Band Spectral Subtraction (MBSS)

The MBSS technique performs spectral subtraction with different over subtraction factor in different non-overlapped frequency bands [4]. The spectral subtraction rule in i^{th} frequency band is given by:

$$\left|\hat{X}_{k_i}\right|^2 = \begin{cases} \left|\overline{Y_{k_i}}\right|^2 - \delta_i\alpha_i.\left|\hat{D}_{k_i}\right|^2, & if\ \left|\overline{Y_{k_i}}\right|^2 > \delta_i\alpha_i.\left|\hat{D}_{k_i}\right|^2 \\ \boldsymbol{\beta}.\left|\overline{\boldsymbol{Y_{k_i}}}\right|^2 & else \end{cases}$$

$$\text{for } b_i \le k \le e_i \quad (4)$$

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:

$$\overline{Y_{k_j}} = \sum_{l=-M}^{M} W_l\, Y_{k_{j-l}} \quad (5)$$

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 $-2 \le l \le +2$ [4].

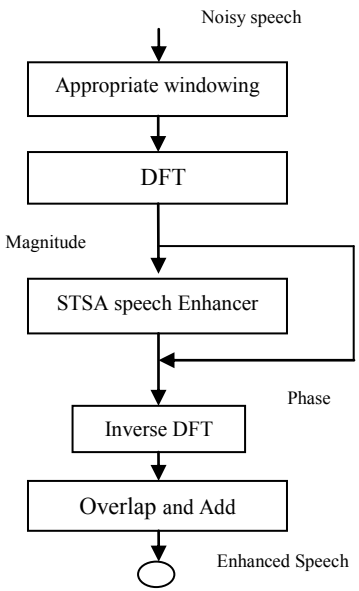

Figure 3: Block Diagram of the DFT based speech enhancement [4].

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

$$SNR_i(dB) = \left[\frac{\sum_{k=b_i}^{e_i}\left|Y_{k_i}\right|^2}{\sum_{k=b_i}^{e_i}\left|\hat{D}_{k_i}\right|^2}\right] \quad (6)$$

α_i can be expressed in terms of SNR_i (defined previously) as follows:

$$\alpha_i = \begin{cases} 4.75 & SNR_i < -5 \\ 4 - \frac{3}{20} SNR_i & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \quad (7)$$

The values of the factor δ_i (tweaking factor) are empirically determined and set according to following equation (Usually 4-8 linearly spaced frequency bands are used).

$$\delta_i = \begin{cases} 1 & f_i < 1\ KHz \\ 2.5 & 1\ KHz \le f_i \le \frac{F_s}{2} - 2\ KHz \\ 1.5 & f_i > \frac{F_s}{2} - 2\ KHz \end{cases} \quad (8)$$

Where f_i is the upper frequency of the the i^{th} band, and F_s is the sampling frequency [4].

2.4. Wiener Filter

In terms of our speech enhancement problem the Wiener filter proposed in [5] is given by:

$$\left|\hat{X}_k\right| = \frac{\xi_k}{\xi_k + 1} |Y_k| \quad (9)$$

Where ξ_k *is* defined as the a priori SNR found by *Decision Directed Method*

2.5. MMSE of Short-Time Spectral Amplitude

Ephraim and Malah [6] 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.

2.5.1. The Gaussian based MMSE-STSA Estimator

The desired gain functions for the MMSE-STSA estimator [6]:

$$G_{MMSE}(\nu_k) = \Gamma(1.5) \frac{\sqrt{\nu_k}}{\gamma_k} \exp\left(-\frac{\nu_k}{2}\right) . \left[(1 + \nu_k) I_0\left(\frac{\nu_k}{2}\right) + \nu_k I_1\left(\frac{\nu_k}{2}\right)\right] \quad (10)$$

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

$$\nu_k = \frac{\xi_k}{\xi_k + 1} \gamma_k \quad (11)$$

where ξ_k and γ_k are defined by:

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (12)$$

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \quad (13)$$

ξ_k and γ_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.

2.5.2. Decision-Directed Estimation Approach

In the proposed estimator, the a priori SNR ξ_k is unknown and we have to estimate it in order to implement the estimator. The reason ξ_k is unknown is because the clean signal is unavailable. The proposed estimator $\widehat{\xi_k}$ of ξ_k [7] is given by:

$$\widehat{\xi_k}(n) = \alpha \frac{\widehat{A_k}^2(n-1)}{\lambda_d(k, n-1)} + (1 - \alpha) P\{\gamma_k(n) - 1\}, \quad 0 \le \alpha < 1, \quad (14)$$

where $\widehat{A_k}(n-1)$ is the amplitude estimator of the k^{th} signal spectral component in the $(n-1)^{th}$ analysis frame and α is a weighting constant that is deduced from

experimental data. The operator $P\{.\}$ is defined by:

$$P\{x\} = f(x) = \begin{cases} x, & x \geq 0 \\ 0, & otherwise \end{cases} \quad (15)$$

The above estimator for $\xi_k(n)$ is a "decision-directed" type estimator, since $\widehat{\xi_k}(n)$ is updated on the basis of a previous amplitude estimate. The initial conditions need to be determined and $\widehat{\xi_k}(0) = \alpha + (1-\alpha)P\{\gamma_k(0) - 1\}$ is found appropriate based on simulations since it minimizes initial transition effects in the enhanced speech [6].

2.5.3. Amplitude Estimator under Speech Presence Uncertainty (SPU)

We consider a two-state model for speech events, that is, either speech is present at a particular frequency bin (hypothesis H_1^k) or that is not (hypothesis H_0^k). This is expressed mathematically using the following binary hypothesis model [8]:

Null hypothesis speech absent:

$$\mathrm{H_0^k : Y_k = D_k}.$$

Alternate hypothesis, speech present:

$$H_1^k : Y_k = X_k + D_k.$$

The multiplicative modifier on the optimal estimator under the signal presence hypothesis is given by:

$$G_{SPU}(k) = \frac{\Lambda(Y_k, q_k)}{1+\Lambda(Y_k, q_k)}, \quad (16)$$

$\Lambda(Y_k, q_k)$ is the generalized likelihood ratio while q_k denotes the *a priori* probability of speech absence in the k^{th} spectral component. By using the Gaussian statistical model assumed for the spectral components, the generalized likelihood ratio:

$$\Lambda\left(Y_k, q_k, \xi'_k\right) = \frac{1-q_k}{q_k} \frac{\exp\left(\frac{\xi'_k}{1+\xi'_k}\gamma_k\right)}{1+\xi'_k} \quad (17)$$

Where ξ'_k is the conditional a priori SNR

$$\xi'_k = \frac{1}{1-q_k}\xi_k \quad (18)$$

2.6. Speech Enhancement using a MMSE Log-Spectral Amplitude estimator

Based on [9] 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. The desired MMSE-LSA gain function (for more details refer to [9]):

$$G_{MMSE-LSA}(\xi_k, \gamma_k) = \frac{\xi_k}{1+\xi_k}\left\{\frac{1}{2}\int_{\nu_k}^{\infty}\frac{e^{-t}}{t}dt\right\} \quad (19)$$

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

2.7. 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 [10]. 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 [10] 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.

2.7.1. The 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 [6]: Null hypothesis H_0^k speech absent: $Y_k = D_k$.

Alternate hypothesis H_0^k speech present: $Y_k = X_k + D_k$. The LSA estimator for the clean speech spectral amplitude (Assuming statistically independent spectral components [9]), which minimizes the mean-squared error of the log spectra, is given by:

$$\widehat{A_k} = exp\{E[lnA_k \backslash Y_k, 0 \le k \le \mathrm{N}-1]\} \triangleq G_{k\ OM-LSA}.|Y_k| \quad (20)$$

The Optimally Modified LSA estimator gain is given by:

$$G_{k\ OM-LSA} = \{G_{H1}(\xi'_k, \gamma_k)\}^{p_k}.G_{min}{}^{1-p_k},\ where\ 0 \le k \le \mathrm{N}-1 \quad (21)$$

2.7.2. A Priori SNR Estimation

According to the decision-directed approach, proposed by Ephraim and Malah [6], it provides a useful estimation method for the non-conditional a priori SNR ξ_k which was given previously by eq (14)

$$\widehat{\xi_k}(n) = \alpha\frac{\widehat{A_k}^2(n-1)}{\lambda_d(k,n-1)} + (1-\alpha)max\{\gamma_k(n)-1, 0\}$$

where $0 < \alpha < 1$, and n 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 [11], [12], [13].

Hence, Cohen in [10] proposed a new estimator of the *Priori SNR* which is given as follows:

$$\widehat{\xi'_k}(n) = \alpha G_{H1}(n-1)^2 + (1-\alpha)\ max\{\gamma_k(n)-1, 0\} \quad (22)$$

2.7.3. A priori Speech Absence Probability (SAP) Estimation

In [10], 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 $(P_{k\ local}(n), P_{k\ global}(n), P_{frame}(n))$ based on the time-frequency distribution of the estimated a priori SNR

These parameters exploit the strong correlation of speech presence in neighboring frequency bins of consecutive frames [10]. Hence, the proposed estimate for the a priori probability for speech absence is obtained by:

$$\hat{q}_k(n) = 1 - P_{k\ local}(n).P_{k\ global}(n).\ P_{frame}(n) \quad (23)$$

$\hat{q}_k$ 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 $\hat{q}_k(n)$ to be smaller than a threshold q_{max} (<1).

3. Implementation and Performance Evaluation

This Section describes the implementation and performance evaluation of six DFT-based single channel speech enhancement algorithms (explained before). The IEEE standard database NOIZEUS (noisy corpus) [14] 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: car, train, babble…etc. The performance comparisons of various implemented algorithms are carried out which are based on visual examinations of signals in the time domain and the spectrograms (clean, noisy, and enhanced speech signals), and also the objective and subjective tests.

3.1. Implementation Details

- Frame size: 20 ms.
- Window Type and Overlap: the most commonly used Hamming window [15].
- 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.
- 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
- $\alpha = \begin{cases} 4.75 & SNR < -5 \\ 4 - \frac{3}{20} SNR & -5 \le SNR \le 20 \\ 1 & SNR > 20 \end{cases}$
- 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 $\alpha = 0.98$.
- 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:

- $\beta = 0.7$, $\xi'_{min} = -10dB$, $\xi'_{max} = -5dB$, $\xi'_{p\,min} = 0dB$, $\xi'_{p\,max} = 10dB$
  $w_{local} = 1$, $w_{global} = 15$
  $q_{max} = 0.95$ h_λ: $Hanning\ window$

- For the implementation of noise estimation algorithm discussed before, the following parameters are used:
  $\eta = 0.7$,
  the threshold $\sigma, is\ to\ \sigma = 1.3$,
  $\lambda = 0.8$,
  $\gamma = 0.998$, $and\ \beta = 0.8$,
  $\alpha_1 = 0.8$, $and\ \alpha_2 = 1$.

3.2. Visual Examinations for the implemented algorithms

Applying the implemented algorithms to the noisy speech signal sentence in

"sp10.wav" corrupted with 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 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. 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 spectrogram using M-LSA algorithm is the nearest to the to the original clean speech spectrogram.

3.3. Objective measures for implemented algorithms performance evaluation

Objective measures are based on a mathematical comparison of the original and enhanced speech signals. 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 3 colored noise environments which are as follows: **(**Train, Car, 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 SNR_{seg} values are averages of 30 measures the number of sentences in the database and are given in dB**)** are shown in the following tables (from Table 1 to Table 4).

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

- 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 more 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.

4. Conclusion and Further Research

The work in this paper addressed the problem of single-channel speech enhancement at the presence of highly non-stationary background noise. A set of six 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 three colored noise environments (train, car, 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.

In the future, we plan to study the real effects of the implemented pre-processing techniques on the various speech communication applications.

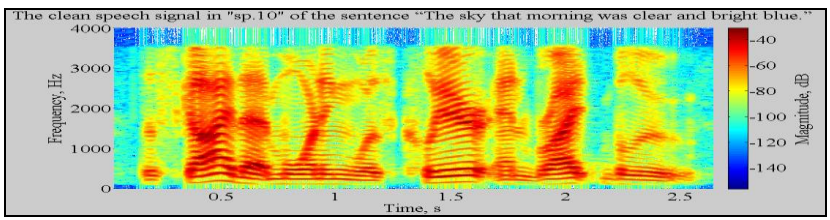

Figure 4**:** The spectrogram of the clean speech signal in "SP.10".

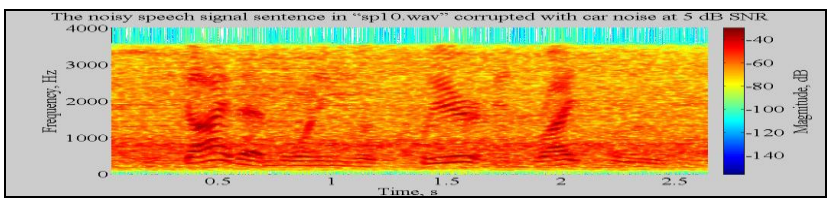

Figure 5: The spectrogram of the noisy signal in "SP.10" corrupted with car noise at 5 dB SNR.

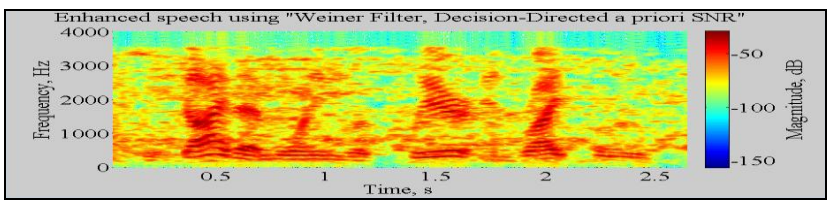

Figure 6: The spectrogram of the enhanced speech using "Weiner Filter, Decision-Directed a priori SNR".

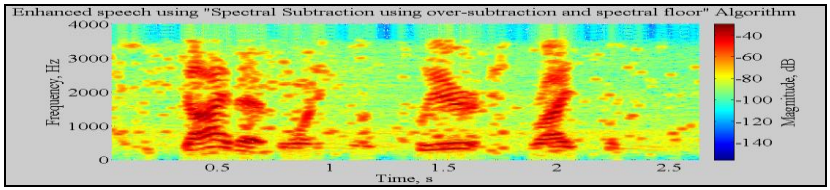

Figure 7: The spectrogram of the enhanced speech using Over-Subtraction and spectral floor" algorithm.

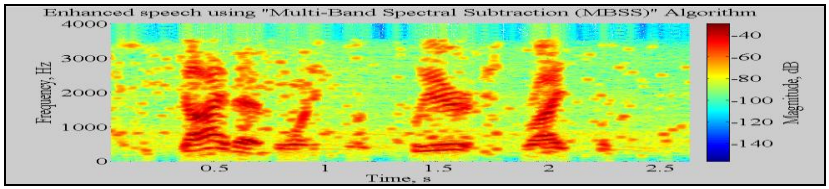

Figure 8**:** The spectrogram of the enhanced speech using "Multi-Band Spectral Subtraction (MBSS)" Algorithm.

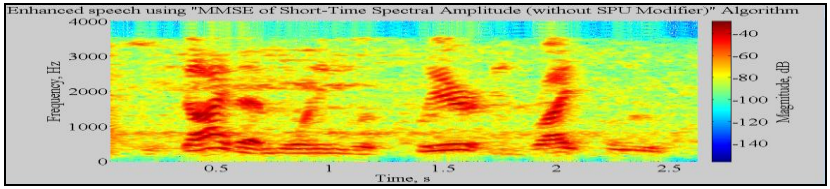

Figure 9**:** The spectrogram of the enhanced speech using "3MMSE-STSA (without using SPU modifier)" Algorithm.

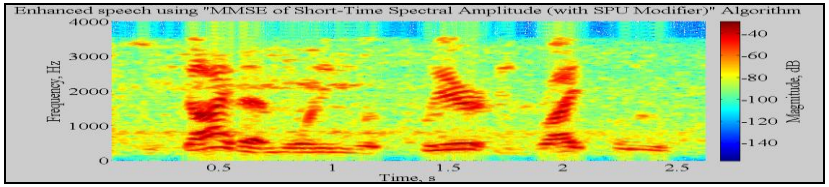

Figure 10: The spectrogram of the enhanced speech using "MMSE-STSA (using SPU modifier)" Algorithm.

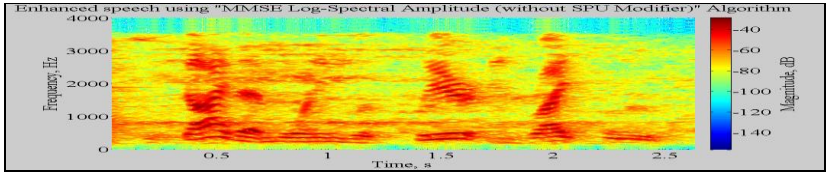

Figure 11: The spectrogram of the enhanced speech using "MMSE-LSA (without using SPU modifier)" Algorithm.

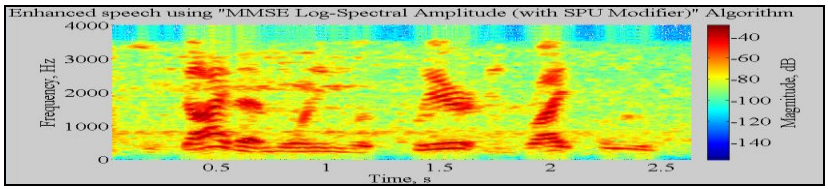

Figure 12**:** The spectrogram of the enhanced speech using "MMSE-LSA (using SPU modifier)" Algorithm.

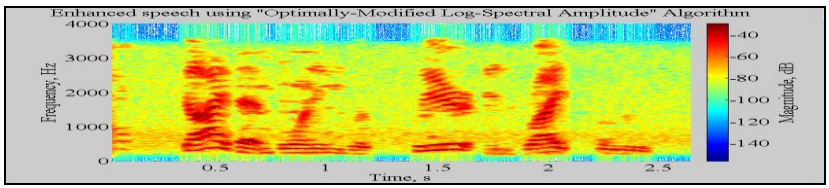

Figure 13: The spectrogram of the enhanced speech using "OM-LSA" Algorithm.

Train noise	SNR=0 dB **SNR_{Seg}** ≅-4.50		5 dB **SNR_{Seg}** ≅-1.67		10 dB **SNR_{Seg}** ≅1.50		15 dB **SNR_{Seg}** ≅4.50	
Objective test	SNR	**SNR_{Seg}**	SNR	**SNR_{Seg}**	SNR	**SNR_{Seg}**	SNR	**SNR_{Seg}**
Weiner DD	6.05	***-0.20***	8.77	***1.51***	13.01	***4.19***	16.16	***8.00***
SS using using over-subtraction and spectral floor	6.14	***-0.14***	8.07	***1.51***	12	***4.26***	17.44	***8.13***
Mband	6.40	***-0.10***	8.56	***2.32***	12.07	***5.12***	17.44	***8.54***
MMSE-STSA	6.48	***0.50***	8.32	***2.40***	12.18	***5.76***	15.10	***8.64***
MMSE-LSA	6.23	***0.55***	8.44	***2.44***	12.22	***5.77***	15.03	***8.68***
MMSE-STSA using SPU modifier	6.10	***1.20***	8.00	***3.11***	11.81	***5.80***	15.32	***8.70***
MMSE-LSA using SPU modofier	6.31	***1.25***	8.24	***3.10***	12.09	***5.81***	15.53	***8.73***
OM-LSA	6.40	***1.50***	8.76	***3.20***	13.66	***6.00***	17.20	***8.90***

Table 1. Train noise reduction

Car noise	SNR=0dB $SNR_{Seg} \cong -4.95$		5 dB $SNR_{Seg} \cong -2.00$		10 dB $SNR_{Seg} \cong 1.05$		15 dB $SNR_{Seg} \cong 4.05$	
Objective test	SNR	SNR_{Seg}	SNR	SNR_{Seg}	SNR	SNR_{Seg}	SNR	SNR_{Seg}
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	4.75	***-0.4***	9.46	***1.83***	13.47	***4.90***	18.75	***9.00***
Mband	4.90	***-0.2***	9.40	***2.33***	13.00	***5.20***	17.30	***9.05***
MMSE-STSA	5.16	***0.50***	9.53	***2.80***	13.46	***5.30***	16.51	***9.10***
MMSE-LSA	5.34	***0.55***	9.38	***2.80***	13.23	***5.33***	16.54	***9.10***
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 modofier	6.05	***0.90***	9.49	***3.25***	13.42	***5.43***	15.01	***9.15***
OM-LSA	6.10	***1.30***	10.48	***3.50***	14.43	***5.55***	18.38	***9.20***

Table 2. Car noise reduction

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

Table 3. Babble noise reduction

Babble noise	SNR=0 dB $SNR_{Seg} \cong -4.48$		5 dB $SNR_{Seg} \cong -1.50$		10 dB $SNR_{Seg} \cong 1.48$		15 dB $SNR_{Seg} \cong 4.40$	
Objective test	SNR	SNR_{Seg}	SNR	SNR_{Seg}	SNR	SNR_{Seg}	SNR	SNR_{Seg}
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 modofier	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 4. White noise reduction

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.

5. References

[1] Plourde Eric, "Bayesian short-time spectral amplitude estimators for single-channel speech enhancement", PHD thesis, McGill University Montreal, Canada, pp 30, October 2009.

[2] Soon Ing Yann, "Transform based speech enhancement techniques", PHD thesis, Nanyang Technological University, pp.9-22, 2003. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[3] Berouti M., Schwartz R. and Makhoul J., "Enhancement of speech corrupted by acoustic noise," Proc. IEEE Int. Conf. on Acoust., Speech, Signal Procs., pp. 208-211, Apr. 1979.

[4] Thiemann Joachim, "Acoustic Noise Suppression for SpeechSignals using Auditory Masking Effects" Department of Electrical & Computer Engineering McGill University Montreal, Canada July 2001, page 43.

[5] 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.

[6] Ephraim Y. and Malah D., "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.

[7] 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.

[8] Middleton D. and Esposito R., "Simultaneous optimum detection and estimation of signals in noise," IEEE Trans. Inform. Theory, vol. IT-14, pp. 434–444, May 1968 H. L.

[9] Cohen I. "Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator," Lamar Signal Processing Ltd.2003.

[10] Soon I. Y., Koh S. N. and Yeo C. K., "Improved Noise Suppression Filter Using Self Adaptive Estimator of Probability of Speech Absence," Signal Processing, vol. 75, pp. 151–159, 1999.

[11] Martin R., Wittke I.and Jax P., "Optmized Estimation of Spectral Parameters for the Coding of Noisy Speech,"in Proc. Int. Conf. Acoustics, Speech and Signal Processing, ICASSP 2000, pp. 1479–1482.

[12] Cohen I. and Berdugo B., "Speech Enhancement for Non-Stationary Noise Environments," to appear in Signal Processing.

[13] Cohen I. and Berdugo B., "Speech Enhancement for Non-Stationary Noise Environments," to appear in Signal Processing.

[14] Hu Y. and Loizou P., "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, 49, 588-601.

[15] Sovka, P., "Extended Spectral Substraction:Description and Preliminary Results",[Research Report]. R95-2. Prague, CTU, Faculty of Electrical Engineering 1995. pp 15.