Iranica Journal of Energy & Environment 2 (2): 166-180, 2011
ISSN 2079-2115
IJEE an Official Peer Reviewed Journal of Babol Noshirvani University of Technology

BUT

Optimized Singular Vector Denoising Approach for Speech Enhancement

Amin Zehtabian, Hamid Hassanpour

Shahrood University of Technology, Shahrood, Iran

(Received: May 4, 2011; Accepted: June 8, 2011)

Abstract: In this paper, a novel approach for speech signal enhancement is presented. This approach employs singular value decomposition (SVD) to overlook noise subspace and uses Genetic Algorithm (GA) to optimally set the essential parameters. The method is elicited by analyzing the effects of environmental noises on the singular vectors as well as the singular values of clean speech signals. This article reviews the existing approaches for subspace estimation and proposes novel techniques for effectively enhancing the singular values and vectors of a noisy speech. This results in a considerable attenuation of the noise and retaining quality of the original speech. The efficiency of our proposed method is affected by a number of parameters which are optimally set by utilizing the GA. Extensive sets of experiments have been carried out on speech signals impaired by additive white Gaussian noise and/or different types of realistic coloured noises. The results of applying the six superior speech enhancement methods are compared using the objective (SNR) and subjective (PESQ) measures.

Key words: Speech Enhancement % Singular Vectors % Genetic Algorithm % Savitzky-Golay Filter

INTRODUCTION

Speech enhancement and noise reduction are used in a large number of speech applications such as automatic voice recognition and speaker authentication systems, cellular mobile communication and hearing aid devices [1-4]. There are two important issues often required to be considered in speech enhancement applications; eliminating the undesired noise from the speech to improve Signal-to-Noise Ratio (SNR) and retaining quality of the original speech signal. There is often a trade-off between the residual noise and the speech quality in the speech enhancement systems. The success of speech enhancement approaches often depends on satisfying both the objective and subjective goals.

The existing speech enhancement methods often reduce the noise by considering the prior assumptions; hence they are suitable for specific applications and conditions [5]. For instance, the signal is completely recoverable from noise if the frequency spectra of the signal and the noise are distinct [1]. Therefore, as a traditional solution for signal enhancement, one can use a typical Low-Pass Filter (LPF). But this assumption may not be feasible in most speech enhancement applications. On the other hand, using these types of filters may have phase effect on the signal and hence, slightly changes its shape. This phenomenon seriously affects the quality of the signal; however it may be neglected by the human audition system.

The nature of environmental noise is another important issue which significantly affects performance of the speech enhancement method and constrains its application. For example, in many spectral subtraction based methods it is assumed that the noise proposes a stationary characteristic or its frequency band is limited to a predefined range [6, 7]. Although it may not be feasible to design an approach able to overcome all kinds of the noise sources, an efficient and robust speech enhancement method must be able to deal with a relatively wide range of noise cases; from stationary to non-stationary and from white to coloured.

In this paper, we present a novel subspace-based approach which provides a considerable noise reduction while cares in preserving the quality and audibility of the original speech signal. The proposed approach includes the combination of innovative speech enhancement levels which independently deal with the singular values and vectors of the signal. Despite of the computational

Corresponding Author: Hamid Hassanpour, Shahrood University of Technology, Shahrood, Iran.
E-mail: h_hassanpour@yahoo.com.

complexity of the GA-based optimization procedure utilized in this approach, the significant speech enhancement level is appealing. Meanwhile, the robustness of the approach in relatively extensive noise conditions makes the proposed method more versatile compared to the other well-known speech enhancement techniques.

The rest of the paper is organised as follows: In Section 2, we provide a comprehensive overview of the existing well-known speech enhancement approaches. Section 3 includes the basic theories behind the traditional subspace division techniques. Since determining the optimum threshold point for subspace division has a crucial role in the development of the subspace division methods, this section also provides an introduction to the more efficient threshold point estimation methods. Section 4 introduces the proposed SVD-based speech enhancement method. This section begins with an introduction to the enhancement of singular vectors and values and then concentrates on the proposed GA-based technique for parameter setting. The section also studies the factors determining the relationship between the noise reduction and speech quality. Section 4 concludes with exploring the Savitzky-Golay parameters effects on the performance of the proposed speech enhancement method. Extensive sets of experiments are provided in Section 5. The efficiencies of the threshold point estimation techniques are also compared in this section. The section then concentrates on reducing the noise from the noisy signals infected with the white noise as well as coloured noises. An overall conclusion is finally provided in Section 6.

Background: Existing speech enhancement approaches, depending on the domain of analysis, can be categorized into three main groups: time, frequency and time-frequency/ time-scale domains.

The Wiener filter is actually an effective solution for speech enhancement that can be implemented both in time and frequency domains. This filter has been widely used by researchers and has also been utilized in many technical applications [1, 8]. This method estimates an optimal noise reduction filter by using the signal and noise spectral characteristics. In a typical Wiener filtering method, the noisy signal is passed through a Finite Impulse Response (FIR) filter whose coefficients are estimated by minimizing the Mean Square Error (MSE) between the clean signal and its estimation to restore the desired signal. Since this procedure is often iterated until convergence occurs, the method is usually called as iterative Wiener filtering. Despite the reasonable complexity of the method and its relatively quick response, in some speech enhancement applications using the Wiener filter may result in some signal degradations. When the SNR value for a noisy speech signal is low, using this method may aggravate the quality of the speech. This is due to the fact that in the Wiener filtering techniques, the amount of noise reduction is generally proportional to the final speech degradation [9]. Therefore, the lower SNR conditions lead to the more noise reduction and consequently it causes more speech distortions.

In the time-scale based approaches, the speech signal is initially subdivided into several frequency bands and the noise-reduced sub-signals are then used to reconstruct the enhanced signal. One of the most efficient transforms which can be used for this sub-division is the wavelet transform. Many researchers have developed the wavelet-based approaches and achieved some considerable results [10-12]. One of these methods is based on the Bionic Wavelet Transform (BWT). The BWT is an adaptive wavelet transform based on a non-linear auditory model of the human cochlear, which captures the non-linearity features of the basilar membrane and translates them into adaptive time-scale transformations of the proper fundamental mother wavelet [12]. In this approach, the enhancement is the result of thresholding on the adapted BWT coefficients.

Since keeping the structure of the original signal is one of the main concerns in speech processing, the Time-Frequency (TF) distributions can be suitable tool in noise attenuation as both time and frequency contents of the signal are considered in such distributions. Recently a TF-based approach for signal enhancement was proposed in [13]. This approach produces a data matrix from the TF representation of the noisy signal and then the singular value decomposition technique is applied to the data matrix. Using this technique, the noise subspace and signal subspace are separated and a noise-reduced signal can be derived. This TF-based technique provides a good performance in noise reduction at the cost of higher computational complexity in comparison with the other existing methods. Another drawback of this approach which may dramatically affect its application is that some TF distributions may not be synthesized to the time series.

There are several speech enhancement methods categorized as frequency domain approaches [6, 7, 14-17]. These methods often use spectral subtraction for reducing the noise. In the spectral-based techniques,

the noise spectrum is usually estimated from the non-speech segments of the noisy signal. Then, the estimated noise spectrum is subtracted from the noisy speech spectrum. Finally, the result is transformed into the time domain. These methods are only suitable for specific applications. For example, in Boll's method, the noise is considered to be stationary [6]. However, the noise is usually nonstationary in practice.

The authors in [18] improved the spectral subtraction technique and proposed a novel approach which applies a perceptual weighting filter to remove the musical residual noise from the preliminary noise-reduced speech. This approach which considerably leads to a more desirable speech quality can be called as over-subtraction method. The technique is based upon an advanced spectral subtraction combined with a perceptual weighting filter based on psycho-acoustical properties. The authors also used a modified masking threshold estimation to eliminate the noise influence during the determination of the speech masking threshold.

There are plenty of signal enhancement approaches implemented in time domain. Subspace based approaches which have been widely used in signal processing application are mainly categorized as time domain based methods. These techniques have also wide applications in speech enhancement [19]. They usually represent the noisy speech signal in a time data matrix which often has the Hankel or Toeplitz forms [20]. Using the SVD technique, the noisy speech signal is enhanced by retaining some of the singular values from the decomposition of the noisy data matrix. The eliminated singular values are supposed to be associated with the noisy part of the signal.

We have recently developed a novel non-destructive time domain approach for reducing the noise from the signal which has indicated its effective performance in reducing the additive white Gaussian noise from stationary and non-stationary noisy synthetic signals [21]. This method is an SVD-based approach, in which reduces the effects of additive noise from the singular values as well as the singular vectors (SVs) of the noisy signal.

In this paper, we develop a novel signal enhancement approach to enhance the real speech signals as well as synthetic signals. Meanwhile in this paper the additive noise is not necessarily a white Gaussian noise. Indeed, the proposed speech enhancement method is properly adapted to reduce the white noise as well as the coloured noise from the noisy speech. The results of applying the proposed method to several standard speech signals are compared with that of other well-known speech enhancement methods including the traditional spectral subtraction approach and its improved over-subtraction version, the Plain SVD-based method which only enhances the singular values per se (without filtering the singular vectors), the iterative Wiener filtering and the adaptive Bionic Wavelet Transforming technique (BWT).

Speech Enhancement Using Subspace Division: In speech processing applications, to reduce the computational time of the procedures it is common to divide the speech signal into some overlapping frames. In all frames, the noisy signal model in the time domain is given by

$$X_n = X_s + W_n \tag{1}$$

Where X_n, X_s and W_n denote the noisy signal, clean signal and additive white Gaussian noise, respectively. Then the noisy time-series in each frame is represented as a Hankel matrix. The Hankel matrix is a square matrix, in which all of the elements are the same along any northeast to southwest diagonal. Therefore, supposing X_n (I), i = 0,1,...,N represents the noisy signal in the time domain, the Hankel matrix H , $R^{P\times Q}$ is constructed as follows.

$$H = \begin{bmatrix} X_n(0) & X_n(1) & \cdots & X_n(Q-1) \\ X_n(1) & X_n(2) & \cdots & X_n(Q) \\ \vdots & \vdots & & \vdots \\ X_n(P-1) & X_n(P) & \cdots & X_n(N-1) \end{bmatrix} \tag{2}$$

Where, $P + Q = N + 1$ and $P \geq Q$ [22]. Note from Equation (1) that a similar relation can be established between the Hankel matrices

$$H_n = N_s + H_{wn} \tag{3}$$

Where H_n, H_n and H_{wn} are respectively the Hankel constructions of the noisy signal, original clean signal and the additive white Gaussian noise.

Generally, the singular value decomposition of matrix H with size $P\times Q$ is of the form

$$H = U \Sigma V^T \tag{4}$$

Where $U_{P\times r}$ and $U_{Q\times r}$ are orthogonal matrices and their columns are respectively the left and right singular vectors. The matrix Σ is a $r\times r$ diagonal matrix of singular values and usually can be expressed as below.

$$\Sigma = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \quad (5)$$

Furthermore, the diagonal matrix S has components such that F_{ij} =0 if i≠j and F_{ij} >0 if $i = j$. It can be shown that $F_{11} \geq F_{22} \geq ... > 0$ are the nonzero singular v alues of the matrix H [23, 24].

Mathematically, the subspace separation for the noisy matrix H_n can be expressed as below.

$$Hn = U \sum V^T = (\bar{U}_s \;\; \bar{U}_n) \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_s \end{bmatrix} \begin{pmatrix} \bar{V}_n^T \\ \bar{V}_n^T \end{pmatrix} \quad (6)$$

Where $\hat{\Sigma}_s$ and $\hat{\Sigma}_n$ respectively represent the singular values associated with the clean signal subspace and noise subspace. Similarly, the singular vectors matrices $\bar{U}_s$ and $\bar{V}_S^T$ correspond to the signal subspace and the matrices $\bar{U}_n$ and $\bar{V}_S^T$ belong to the noise subspace. Equation (6) can be rewritten as

$$H_n = \bar{U}_s \hat{\Sigma}_s \bar{V}_S^T + \bar{U}_n \hat{\Sigma}_n \bar{V}_n^T \quad (7)$$

Comparing Equations (3) and (7) yields

$$\hat{H}_s = \bar{U}_s \hat{\Sigma}_s \bar{V}_s^T \quad (8)$$

And

$$\hat{H}_{wn} = \bar{U}_n \hat{\Sigma}_n \bar{V}_n^T \quad (9)$$

Since the matrices $\hat{H}_s$ and $\hat{H}_{wn}$ are respectively the approximation of the initial clean data matrix and the noise matrix, we can reduce the effect of additive noise from the original signal via removing or decreasing the $\hat{H}_{wn}$ subspace and utilizing the $\hat{H}_s$ matrix in reconstruction of the enhanced data matrix.

From Equation (6) it can be deduced that a well-defined threshold point must be determined in the Σ matrix, where the lower singular values from that point may suppose to be from the noise subspace. Finding this point is a critical step in the subspace based enhancement technique since an improper selection may result in an insufficient noise reduction or even an excessive noise removal. Section 3.1 provides a brief review of the existing threshold point estimation (TPE) algorithms and in Section 4, a novel technique will be presented to find the optimal point.

As discussed in [21], in the traditional SVD-based methods, the noise subspace's singular values are set to zero for noise reduction. Then the noise-reduced singular value matrix can be achieved by

$$\Sigma_e = \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & 0 \end{bmatrix} \quad (10)$$

Where Σ_e denotes the singular value matrix of the enhanced speech signal and $\hat{\Sigma}_s$ denotes the approximation of the signal subspace. The enhanced data matrix is finally given by

$$H_e = U \Sigma_e V^T \quad (11)$$

and the enhanced signal is reconstructed as

$$X_e = [H_e(1,1)...H_e(1,Q),H_e(2,Q)...H_e(P,Q)] \quad (12)$$

Threshold Point Estimation Techniques: As stated in the previous subsection, a precise threshold point must be considered on the singular values associated with the matrix of the noisy signal for a proper subspace division. The researchers have developed some methods to calculate this point accurately. These methods are briefly described in the following.

Constant Ratio Method (CRM): In this method, first the singular values are sorted in a decreasing order and then they are normalized with an amplitude range of 1. Afterwards, using an experimentally determined constant ratio (which depends on the application and the signal type), the lower normalized values are supposed to be from the noise subspace and must be filtered. Though it may be a fast trick, but especially for the more complicated signals the results are not good enough to be acceptable.

Least Squares Approximation Method (LSA): In this method, the noise variance is supposed to be calculated from the non-speech frames. Calculating the SVD of the noisy data yields to $H_n = U\Sigma V^T$. Then, an approximation for the original signal matrix H_s can be obtained using Eq. (13):

$$\min_{rank(\hat{H}_{LS})=K} \| H_n - \hat{H}_{LS} \|^2 \quad (13)$$

Where $\hat{H}_{LS}$ is the least square approximation of H_s. In Equation (13), the parameter L which minimizes the mentioned relation can result in the best approximation matrix $\hat{H}_{LS}$. Then the matrix $\hat{H}_{LS}$ can be achieved by.

$$\hat{H}_{LS} = U_s \hat{\Sigma}_{LS} V^T \quad (14)$$

Where $\hat{\Sigma}_{LS}$ is the noise reduced singular values matrix using the rank *K* achieved by the LSA method [25].

Minimum Variance Approximation Method (MVA): In this approach, before reproducing the reduced rank data matrix, the singular values are transformed using a diagonal matrix denoted by F_{MV}. The enhanced matrix $\hat{H}_{MV}$ is supposedly the best approximation of the initial clean matrix H_s and can be achieved as below

$$\hat{H}_{MV} = U(F_{MV} \hat{\Sigma}_{MV}) V^T \quad (15)$$

Where, $\hat{\Sigma}_{MV}$ is the noise reduced singular values matrix and the diagonal matrix F_{MV} can be gained by

$$F_{MV} = diag\ ((1 - \frac{s^2_{noise}}{s^2_1}), .., (1 - \frac{s^2_{noise}}{s^2_k})) \quad (16)$$

In comparison with the LSA approach, using minimum variance approximation method often leads to a better speech recognition performance. For further information please refer to references [26, 27].

Maximum Changes in the Slope of Curve (MCSC): In [28], maximum changes in the slope of the singular values curve are evaluated to obtain the threshold point. Although the MCSC method utilizes an approximately straightforward algorithm for effectively finding the threshold point, its application is constrained to a limited range of signals.

The Proposed Speech Enhancement Method: In this section, a novel speech enhancement approach is presented which proposes a technique to determine the optimal threshold point. Meanwhile, the proposed method develops the traditional subspace based techniques and suggests novel ideas for enhancing the singular vectors of a noisy speech signal and optimizing other parameters used for an efficient speech enhancement.

Singular Vectors Enhancement: Figure 1 illustrates the outcomes of filtering the SVs in reducing noise from an arbitrary multi-frequency signal. To reduce the effect of noise from SVs which are treated as time-series, we utilize the Savitzky-Golay filter [29]. In the Savitzky-Golay approach, each value of the series is replaced with a new value which is obtained from a polynomial fit to $2k' + 1$ neighbouring points. The parameter *k'* is equal to, or larger than the order of the polynomial.

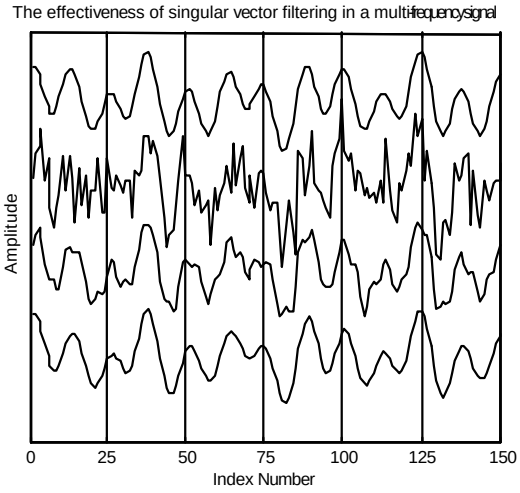

Fig. 1: The result of applying the Savitzky-Golay filter on the singular vectors of a multi-frequency signal. From top to bottom: clean signal, noisy signal with SNR=0 dB, the result of enhancing the singular values of the noise subspace per se, the result of filtering the singular vectors as well as noise subspace subtraction.

The main advantage of this approach in comparison with other adjacent averaging techniques is that it tends to preserve the features of the time series distribution.

In this method, a polynomial is fit to a number of consecutive data points from the time-series. The degree of the Savitzky-Golay polynomial is denoted by S_{deg} and the number of consecutive samples (which can be considered as the window length of the Savitzky-Golay filter) is shown by S_{win}. Filtered SVs can be then obtained as follows

$$U^i_e = F\left(\overline{U}^i_s\right), i = 1, ..., P \quad (17)$$

$$V^i_e = F\left(\overline{V}^i_s\right), i = 1, ..., Q \quad (18)$$

Where *F*(.)denotes the Savitzky-Golay filter function, $\overline{U}_S$ and $\overline{V}_S$ are the singular vectors corresponding to the signal subspace (refer to Equation 7), U^i_e and V^i_s are the enhanced singular vectors after applying the Savitzky-Golay filter and the integer variable *i* is the sample index.

Singular Values Enhancement: In section 3, some of the most common techniques for finding the threshold point used for subspace division were introduced briefly.

The MCSC method which was proposed first by the authors in [28] is able to reduce the effect of white noise from many synthetic signals. Nevertheless, our recent comprehensive researches have shown that for more complicated signals such as speech, determining the proper threshold point seems challenging and needs more attentions.

Hence, in the presented paper we propose a novel technique for finding the most optimum threshold point in comparison with the other existing well-known approaches. This technique utilizes a well-defined cost function and applies the Genetic Algorithm (GA) to minimize this function. This GA-based Threshold Estimation (GA-TE) procedure will be explained in the following subsection.

Utilizing GA as a Parameter Setting Tool: The previous subsections described some crucial parameters affecting performance of the proposed speech enhancement method. They include the number of rows in the Hankel data matrix 0*l*0, the optimum threshold point needed for space subdivision 0P_{thr}0, the degree of polynomial S_{deg} and the window size of the Savitzky-Golay filter S_{win} used for filtering the singular vectors. To optimally set these parameters, we specify a well-defined cost function (Equation 19) and then use the genetic algorithm to minimize this function. The GA is an iterative algorithm which randomly chooses a value within the search space in each repetition [30]. Hence we define our proposed cost function as below

$$Cost(l,P_{thr},S_{deg},S_{win}) = (1-a)\left(\sum_i \| X_e(i) - X_n(i) \|\right) + a\sum_i \| X_e(i+1) - X_e(i) \| \quad (19)$$

In the above equation, x_n, x_e and i represent the noisy speech signal, enhanced signal and the sample index respectively. At the right side of the equation, the first term indicates the distance between the enhanced speech and the noisy speech. The first term of this function indicates that the enhanced signal should be similar to the noisy signal. This is the only thing we know about the original signal. The second term also indicates the smoothness of the enhanced speech signal. The parameter α is the smoothing factor which is chosen between 0 and 1. Where there is no idea about the smoothness level suited for the speech enhancement application, setting this parameter to a balanced value (for example α=0.5) is suggested. It needs to be noted that almost every denoising filter tends to decrease the level of the sudden changes in successive samples of a given noisy signal. Therefore, it is important to precisely manage the smoothness of the final enhanced signal.

Noise Reduction Versus Speech Audibility: There are two important goals often interested in speech enhancement applications; reducing the undesired noise from the speech and improving the perceptional quality or audibility of the noisy speech signal. There is often a trade-off between the residual noise and the speech quality. Reducing the noise without considering the quality of the speech may not be a good solution. In this section we introduce the two parameters which strongly affect the relationship between the noise reduction level and speech quality in our proposed SVD-based method.

αEffect: As discussed before, α is a factor (within 0 and 1) determining the smoothness of the enhanced signal. The value of this factor depends to the signal type and the application, hence is chosen experimentally. For instance, where we deal with linear FM signals, the factor α is supposed to be equal to 0.3; but in speech enhancement applications, the smoothness factor may be determined as a balanced value (α=0.5), whereas the characteristics of the speech signals may vary more randomly.

***K_{red}* Effect:** By applying our novel threshold estimation technique, namely GA-TE, the signal and noise subspaces can be separated effectively. In [21], we have suggested the singular values associated with the noise subspace be set to zero. This approach reduces the effects of additive noise from the signal, but it may not preserve details of the signal. This is an important issue to retain audibility of speech signals. Hence, in this research, since the noisy signals are supposed to be speech, we propose to reduce the noise subspace's singular values by a proper reduction factor. Therefore, the enhanced singular value matrix can be achieved by

$$\sum e = \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_n * K_{red} \end{bmatrix} \quad (20)$$

Where Σ_e denotes the singular value matrix of the enhanced speech signal, $\hat{\Sigma}_s$ and $\hat{\Sigma}_n$ denote the approximations of the signal subspace and noise subspace respectively and K_{red} is the reduction factor.

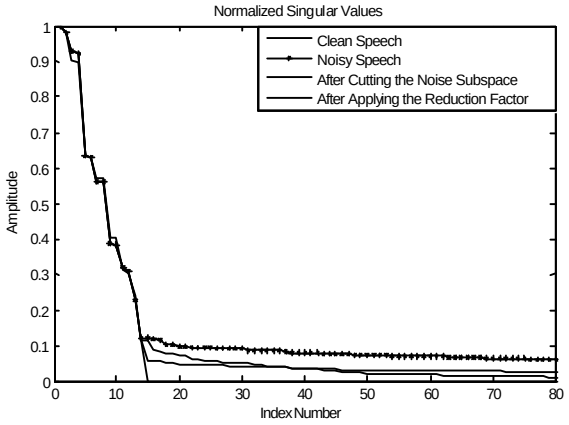

Fig. 2: The effect of applying a reduction factor K_{red}, instead of setting the noise subspace's singular values to zero.

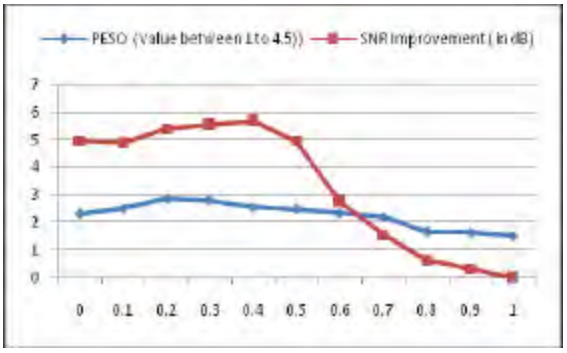

Fig. 3: Plot of PESQ level and SNR improvement (y-axis) versus reduction factor K_{red} (x-axis) for a given speech signal

Since the key parameters α and K_{red} control the noise reduction level and the speech quality enhancement, it is important to evaluate their effects on these two objectives. Following Eq. (19), if is set to zero, the cost function will be equal to the Euclidian distance of the noisy and the enhanced signal. Hence, it does not reflect the smoothness level of the signal at all. Inversely, setting the smoothness factor to its maximum value (α = 1) will neglect the essential similarity between the structures of the enhance signal and the noisy signal. The considerable diversity in characteristics of the noisy speech signals used in the experiments necessitates setting the α factor to a balanced value (α = 0.5).

The effect of the reduction factor K_{red} is even more considerable. Figure 2 demonstrates the effectiveness of this factor in retrieving the singular values of the clean speech signal in comparison with the previous technique, where the noisy singular values lower than threshold point were set to zero.

Considering the singular values curves depicted in Figure 2 may persuade for applying the reduction factor. But for a more comprehensive judgment, it is preferred to evaluate the gained results with a proper quality measure [31]. Hence, we utilize the ITU-T P.862 standard [32] for Perceptual Evaluation of Speech Quality (PESQ). The PESQ quantifies the voice quality and measures the effects of noise, delay, clipping and coding distortions. This can be carried out by comparing an input signal with its corresponding output and measuring the voice quality [33, 34]. For most of the practical applications, the PESQ algorithm produces a value ranging from 1 (the severest degradation) to 4.5 (without any degradation). Figure 3 depicts the PESQ level and SNR improvement for a noise-reduced speech contaminated by an additive white Gaussian noise, where the x-axis is the K_{red} parameter used for noise subspace reduction. As mentioned before, this factor can be chosen based on objectives of the speech enhancement application. It can be inferred from the plot that there is a substantial range across which the overall results are consistent, while either extremely large or extremely small values as the reduction factor level substantially degrade the performance of the method. In this experiment we may choose K_{red} = 0.4 to obtain the most desired results.

It is clear that the enhanced data matrix can be finally achieved by substitution of Equations (17, 18 and 20) in the basic SVD relation (Equation 4) which yields.

$$H_e = U_e^i \Sigma_e V_e^{i^T} \tag{21}$$

The Savitzky-Golay Parameters Effects: In Section 4, we have reviewed the Savitzky-Golay filter and its application for reducing the noise from the singular vectors. As stated before, there are two important parameters strongly affect performance of the Savitzky-Golay smoothing filter in reducing the effect of noise from the SVs; the degree of the polynomial and the frame size of the Savitzky-Golay filter which are denoted respectively by $S.G_{deg}$ and $S.G_{win}$. Figures (4-a) and (4-b) illustrate the effects of choosing various values as the Savitzky-Golay polynomial degree and the frame size, respectively. The figures indicate that an improper parameter selection may result in a disappointing performance and degrading the signal. Conversely, an optimum parameter setting results in a considerably enhanced signal. In Section 4.3, a GA-based technique was introduced for optimally setting the characteristics of the Savitzky-Golay filter. In this experiment, the proposed GA-TE technique provides the optimum results with $S.G_{deg}$ = 3 and $S.G_{win}$ = 15, which are consistent with the results in Figure 4.

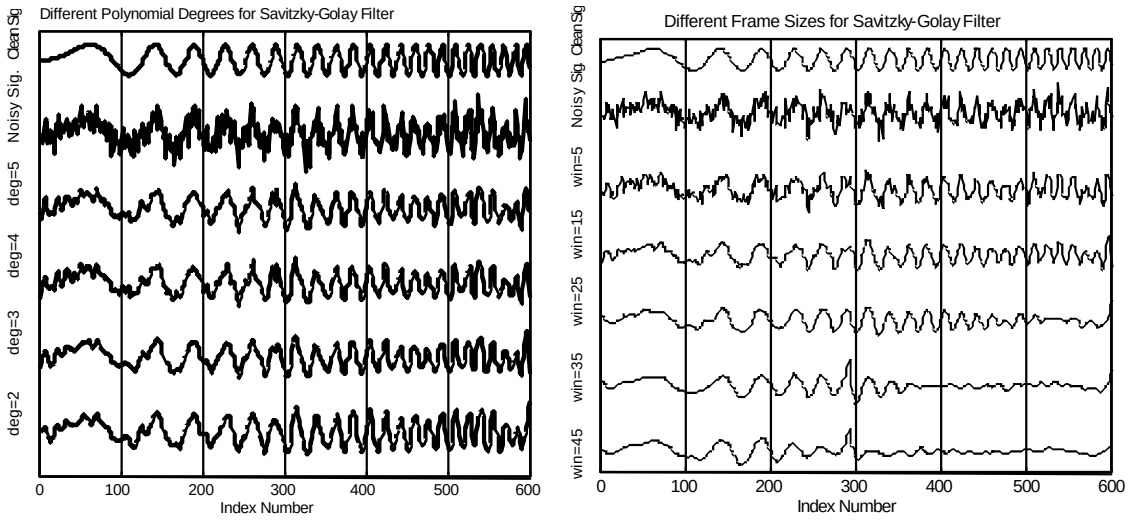

Fig. 4: The Savitzky-Golay parameters effects in reducing the noise from a given noisy linear FM signal, (a) the results of applying different numbers as the degree of the polynomial ($S.G_{deg}$ and (b) the results of applying various Savitzky-Golay frame (window) sizes ($S.G_{win}$)

Reducing Coloured Noise: The coloured noise is defined as a process with unequal power at different frequencies [1]. This makes the spectrum of the noisy signal to have a non-flat shape. Since the frequency distribution of the additive noise and hence the characteristics of the coloured noisy signals are relatively different from that of the white noise, it may be more difficult to discriminate the principal values and vectors associated to the signal from those related to the noise. Two approaches are suggested in this section for such problems. The first approach is to apply a pre-whitening process to the noisy speech. This pre-process transforms the coloured noise to an uncorrelated white noise which its variance is equal to 1. This procedure requires estimating the noise covariance matrix from the non-speech segments of the signal. The pre-whitening algorithm presented in this paper, uses the Cholesky Factor. The second approach is more straightforward and internally performs the whitening stages by employing the Generalized Singular Value Decomposition (GSVD) algorithm. These two techniques are described in the following subsections.

Applying a Pre-Whitening Level: In this section, first we suppose that the coloured noise was added to the clean speech signal and then, we represent them in the form of Hankel matrices:

$$H_{cn} = H_s + N \tag{22}$$

Where, H_{cn} is the Hankel matrix of the clean speech (H_s), infected by an additive coloured noise ((N). Now we apply RG^{1} matrix to H_n from the above equation, which R is the Cholesky Factor of N^TN. Then the following equation can be obtained

$$N^TN = R^TR \tag{23}$$

There are plenty of strategies to calculate the Cholesky Factor R. For the noisy speech case, one solution is to separate the silence or non-speech segments of the noisy signal and estimate the Hankel representation of the additive noise (N) from that frames using:

$$N = QR \tag{24}$$

Where,

$$Q^TQ = I \tag{25}$$

Now, by calculating N^TN, the Cholesky Factor can be obtained. Consequently the pre-whitening process can be yielded as below

$$H_{wn} + H_{cn}R^{-1} \tag{26}$$

Where, H_{cn} was the Hankel representation of the signal infected by the additive coloured noise and H_{wn} is the Hankel form of the noisy signal which its noise is whitened.

Substitution of Equation (22) in Equation (26) yields

$$H_{wn} = H_sR^{-1} + NR^{-1} \tag{27}$$

After applying the pre-whitening level described above, the proposed GA-SVD speech enhancement method can be used for reducing the effect of noise from the H_{wn} matrix. This must be noted that after reproducing the noise reduced matrix constructed by the enhanced singular values and singular vectors, a de-whitening level must be employed on the matrix. Finally, the enhanced speech can be easily extracted from this de-whitened matrix.

The Proposed GA-GSVD Algorithm: Although the pre-whitening technique may be a proper solution when we deal with the non-white noises, it may cause some degradation to the final speech signal due to its numerical instabilities. In other words, by adding a pre-whitening stage prior to our proposed SVD-based algorithm and a de-whitening level afterwards, the speech enhancement level is not encouraging enough. Avoiding this problem, we apply the GSVD (Generalized Singular Value Decomposition) algorithm which has well-defined implicit whitening levels interiorly and consequently decreases the quality lost caused by applying the pre-whitening and de-whitening stages manually.

Indeed, the GSVD concept is an extension of the truncated Quotient SVD (QSVD) theory, which is clearly described in [35] and its effectiveness in reducing the coloured noise is well proved [36]. Utilizing the GSVD, the novel speech enhancement procedure described in the previous sections can be modified and easily extended to reduce the effect of coloured noise from the speech. The results of applying the proposed method to the speech signals infected by coloured noises are described in Section 5.2.

EXPERIMENTAL RESULTS

Efficiency Evaluation of the TPE Techniques: In this section, we evaluate performance of the existing threshold point estimation algorithms, as described in Section 3.1, in calculating the proper threshold value (P_{thr}). In this evaluation, ten noisy speech signals are provided using AURORA database [37] and then impaired by additive white Gaussian noise with 0, +2, +5 and +10dB SNR in different experiments. Table 1 represents the averaged SNR improvement after applying the five TPE algorithms to the ten noisy speech signals. Note that in this experiment, after estimation of the threshold point, the lower singular values were set to zero for space subdivision. Then, the noise-reduced singular value matrix is used for reconstructing the enhanced data matrix. The constant ratio selected for the CRM method was empirically set to 0.2;

To have a better insight into the circumstances of carrying out the TPE methods, we have plotted the normalized singular values and depicted the threshold points determined by each of the techniques on a given noisy speech (Figure 5). The results of this experiment can reasonably convince us to apply the proposed GA-TE method to find the optimized threshold point.

Table 1: Averaged SNR improvements for the existing threshold estimation techniques

Initial SNR (in dB)	CRM	LSA	MVA	MSCS	GA-TE
0	4.67	7.98	7.81	8.65	10.81
2	3.86	7.04	6.90	8.07	10.39
5	3.23	6.60	6.42	6.73	8.66
10	1.76	4.05	4.26	4.51	5.73

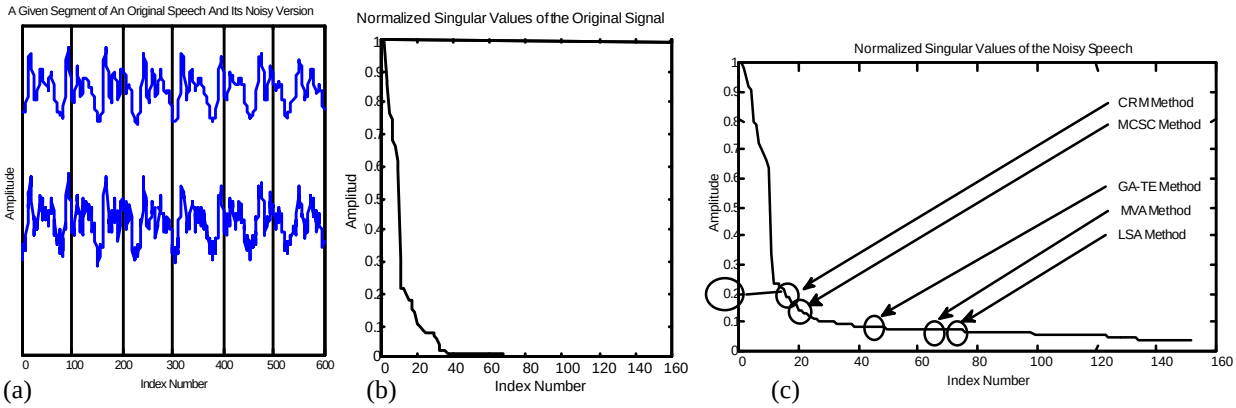

Fig. 5: Visual comparison of the five TPE methods: (a) a given segment of an original speech and its 5 dB noisy version, (b) Normalized singular values of the original signal, (c) threshold point determined by the CRM, LSA, MVA, MCSC and GA-TE algorithms..

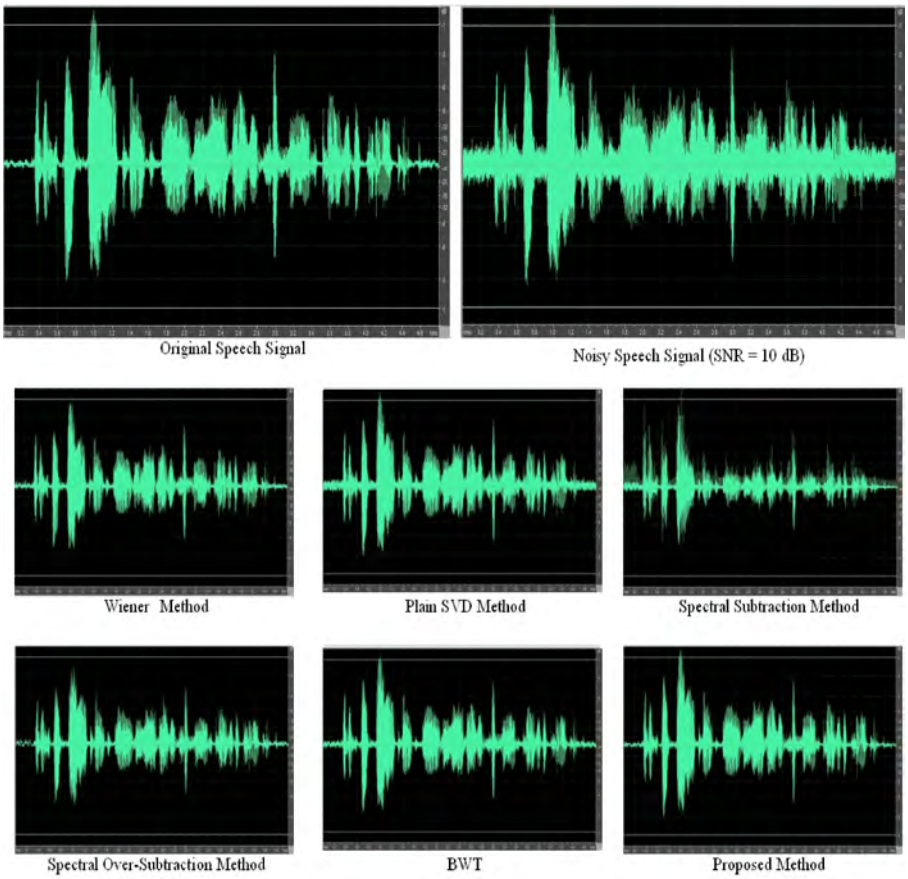

Fig. 6: Time-domain representation of the six speech enhancement approaches

Performance Comparison

The White Noise Case: In this section, the speech enhancement approaches are implemented and their performance in reducing the effect of additive white Gaussian noise is investigated. The compared methods include the iterative Wiener filtering, the traditional SVD-based noise subspace subtraction method which only deals with the singular values and there is no enhancement for the singular vectors (namely, Plain SVD (PSVD) method), the spectral subtraction approach and its improved version called as spectral over-subtraction, the Bionic Wavelet Transform (BWT) and the proposed method (called as GSVD method). Note that all of the methods are first precisely optimized with respect to the speech enhancement applications. Afterward, the quantitative and qualitative measurements are employed to provide a comprehensive insight on performance of the existing speech enhancement approaches.

As discussed before, to overcome the complexity of the time-series to Hankel matrix conversion process and simplify the mathematical operations, in the proposed method each speech signal must be initially divided into several fixed- length frames. Hence, after sampling the input speech with a sampling rate of 8 kHz, we divide the time-series signal into several frames with a N samples hanning window and then represent each of these frames in a Hankel matrix. In the following experiments, the number of samples in each frame is equal to 600. On the other hand, the smoothness factor α and the reduction factor K_{red} are experimentally set to 0.5 and 0.2, respectively.

Figure 6 illustrates an arbitrary original speech signal which is infected then by a 10 dB white Gaussian noise. The six pre-mentioned speech enhancement methods have been applied to the noisy speech and their relevant time-domain representations are drawn.

For a more precise and a thorough visual comparison of the six eminent methods, we represent all of the speech signals in the Time-Frequency Domain (TFD). According to Figure 7, it is clear that the proposed GA-SVD approach has the best performance in retrieving the TFD characteristics of the original speech in this noise condition, compared to the other methods.

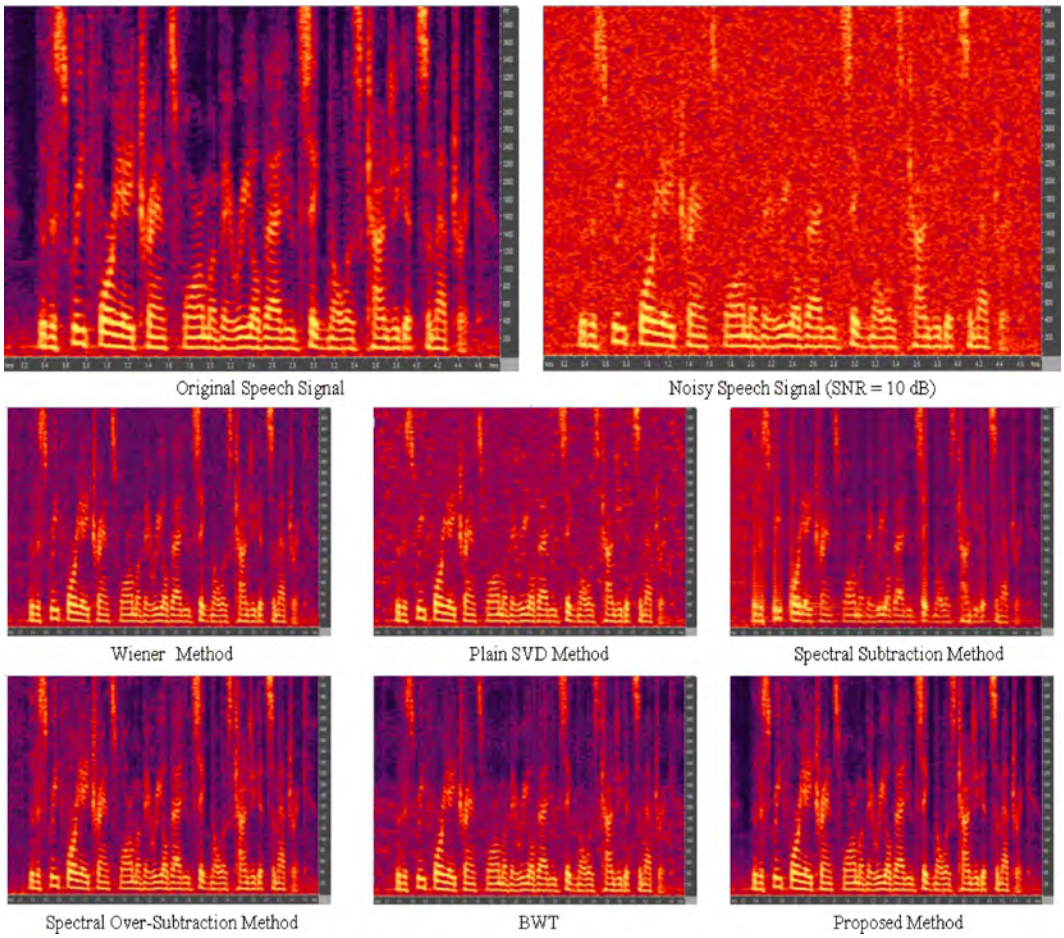

Fig. 7: Time-Frequency representation of the six speech enhancement approaches

Table 2: The SNR and PESQ improvement for the six methods applied on a given noisy speech signal corrupted by a 10 dB white additive noise

Method	Wiener	Plain SVD	Spectral Subtraction	Spectral Over-Subtraction	BWT	Proposed Method
SNR Improvement(dB)	4.18	4.02	- 0.56	1.93	4.90	6.48
PESQ Improvement	0.61	0.48	- 0.13	0.33	0.75	0.90

In addition to the visual demonstrations, the quantitative comparison between the methods applied in this experiment is drawn in Table 2. In the next subsection, the efficiency of the speech enhancement approaches are precisely examined in a relatively wide range of the initial SNR levels.

For a more comprehensive comparison between the pre-mentioned speech enhancement techniques, in this section the Monte-Carlo simulation of the techniques is available. In the presented experiment, ten different clean speech signals are randomly selected from the database and then infected by various levels of white additive noise (from 0 dB to 15 dB). The six speech enhancement algorithms are then applied on each noisy speech and consequently the averaged SNR and PESQ results are drawn as shown in Figures 8 and 9. Note that in Figure 9, each initial PESQ level is determined at the corresponding initial SNR value of the noisy speech.

The Realistic Coloured Noise Case: In this section, the performance of the proposed method is evaluated at the presence of coloured noise process and then compared to that of the other well-known speech processing techniques. Since the proposed approach applies the GSVD, it is called as the GA-GSVD method. All of the six pre-mentioned speech enhancement methods are applied to a variety of speech signals disturbed by three sorts of the coloured noises; the Pink, the Factory and the Babble noise. In the presented experiment, each method is implemented ten times on the signals and the gained results are then averaged as summarized in Table 3.

Table 3: SNR Improvement results for coloured noise case at varying SNR levels (0, +5 and +10 dB)

	SNR Improvement (in dB)								
	Pink Noise			Factory Noise			Babble Noise		
Methods	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB	0 dB	5 dB	10 dB
Iterative Wiener	2.40	1.30	0.88	1.97	1.04	0.80	2.25	1.05	0.64
Plain GSVD	2.57	2.12	1.67	2.11	2.00	1.60	1.53	1.39	1.18
Spectral Subtraction	0.95	- 1.54	- 4.60	0.75	- 1.06	- 4.04	0.40	- 1.65	-4.23
Spectral Over-Subtraction	3.76	0.54	- 2.33	3.62	0.41	- 2.10	2.43	- 0.26	-2.64
BWT	7.16	4.58	2.21	5.78	4.05	2.04	2.79	2.18	1.76
Proposed GA-GSVD Method	6.40	4.97	3.90	5.65	4.44	3.88	3.92	3.32	3.06

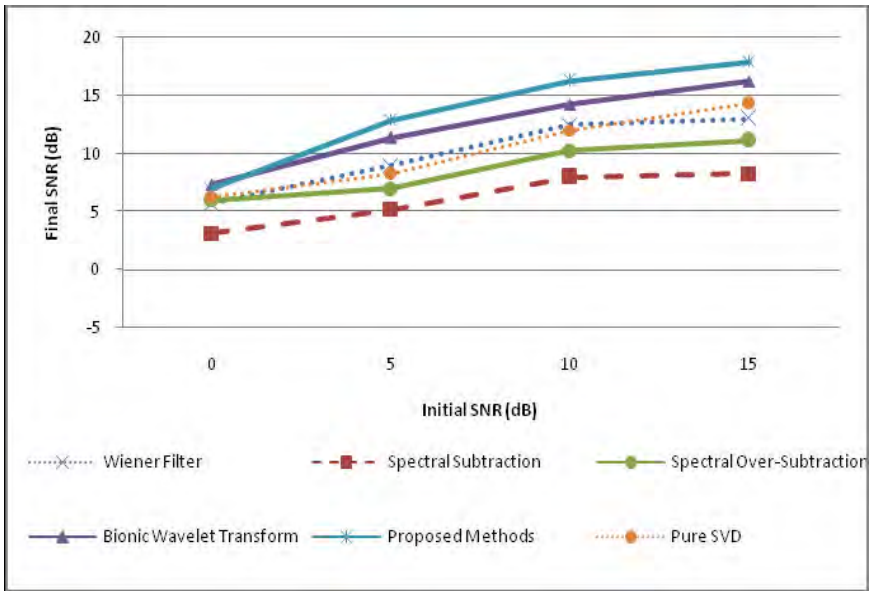

Fig. 8: SNR results for white Gaussian noise case at varying SNR levels (0, +5, +10 and +15 dB)

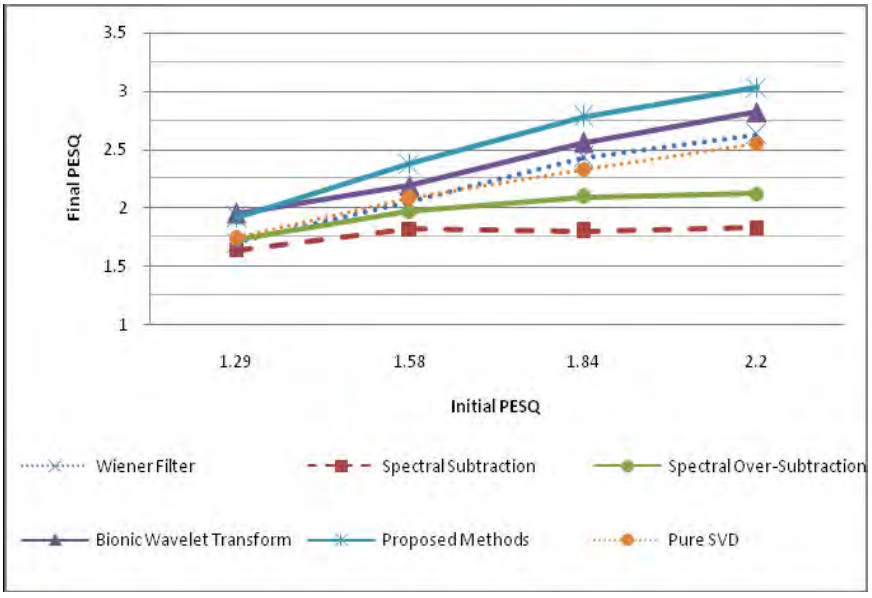

Fig. 9: PESQ results for white Gaussian noise at varying SNR levels (0, +5, +10 and +15 dB)

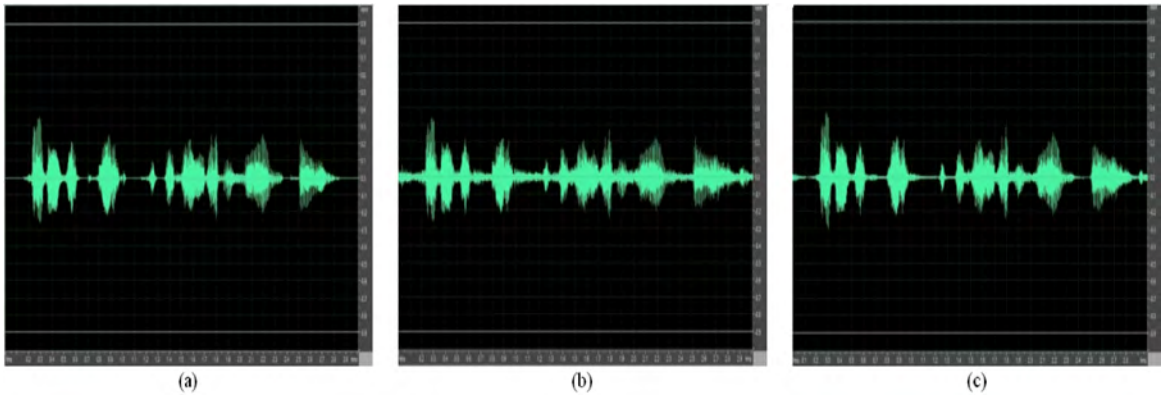

Fig. 10: (a) an arbitrary clean speech signal, (b) the speech signal corrupted by a 10 dB Babble noise, (c) the noise reduced speech with SNR= 13.7 dB

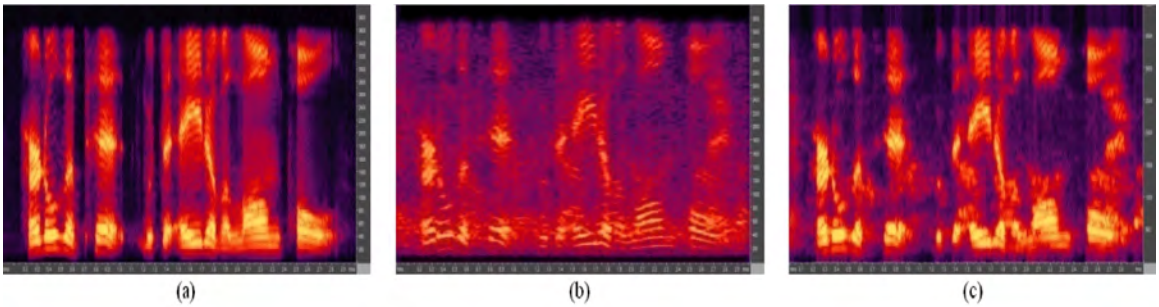

Fig: 11: The Time-Frequency representation of (a) an arbitrary clean speech signal, (b) the speech signal corrupted by a 10 dB Babble noise, (c) the noise reduced speech

The Babble noise process is considered as one of the most well-known coloured noises. Figure (10-a) shows an arbitrary speech signal. The clean speech is then corrupted with a 10 dB Babble noise process. The noisy speech is illustrated in Figure (10-b). The proposed GA-GSVD method is then applied to the noisy speech. Consequently, the enhanced speech is indicated in Figure (10-c). Calculating the SNR level of the signal attests the considerable enhancement in the signal-to-noise ratio.

In addition to the time domain representation of the signals, the time-frequency spectrums of the speech signals are provided in Figure 11.

DISCUSSION

Results represented in Figures 6 to 9 and Table 2 clearly indicate prominence of the proposed GA-SVD method in retrieving the quality of the noisy speech signal as well as reducing the effect of additive white noise from the signal. Indeed, the considerable enhancement in SNR level is guaranteed especially for SNR values higher than about 3 dB. The other encouraging evidence is the noticeable increment in the PESQ value. In other words, utilising the novel proposed technique, a twofold speech enhancement is assured: significant noise reduction and audibility improvement of the enhanced signal. At realistic coloured noise conditions, the proposed GA-GSVD method also outperforms the other approaches (Table 3). Applying the GSVD operator instead of SVD makes the proposed method more reliable in dealing with the signals infected by coloured noises.

From the figures, the Bionic Wavelet Transform (BWT) approach also excels the four other methods at nearly all noise levels. Since the method applies the auditory model of the human cochlear, hence it represents a significant adoption with the human audition system. Therefore the BWT method can properly retrieve the quality of the speech signal and enhance its PESQ level. From Table 3, in lower initial SNR values at the presence of coloured noises, the performance of BWT method is close to or even better than that of the proposed method. But while the SNR increases, the GA-GSVD method excels the BWT.

The iterative Wiener approach is also competitive with the two pre-mentioned prominent methods, especially in enhancing the PESQ criterion. Indeed, the Wiener parameters are precisely tuned to achieve a proper balance between the noise reduction and speech

distortion [7]. This equilibrium results in a considerable PESQ improvement as well as a desirable SNR enhancement at the same time. Once the expected trade-off is not reached, although the SNR improvement at low SNR conditions may seem appreciable, but the amount of speech degradation surely decreases the appeal of using this method. According to Figures 8 and 9, the optimized iterative Wiener filter may present its most satisfying performance at the medium levels of the noise, however the desired balance between the SNR improvement and the speech quality cannot be guaranteed at extremely high or low SNR values. From the application diversity point of view, the Wiener filter may be the best alternative in reducing the effect of noise in real-time applications such as hearing aid devices. This arises from its desirable speech quality enhancement as well as the reasonable complexity of the algorithm.

The performance of the two Spectral-based techniques seems disappointing compared to the other methods, at least for these noise conditions. After a more critic review of Figure 7, some horizontal lines may be recognized in the spectrums related to the Spectral Subtraction and Spectral Over-Subtraction methods. These lines imply some disadvantageous in the quality and audibility of the enhanced speech which strongly affect the enhancement criteria. On the other hand, the performance of the Spectral-based methods is also heavily dependent on the initial SNR value of the noisy speech. It means that the large initial SNR values result in a so-called saturation effect which leads to poor enhancement results.

The so-called Plain SVD and Plain GSVD approaches are also able to reduce the noise without considerable degradation of the speech quality, but the criteria improvements are marginally fewer than that of the iterative Wiener method. In these traditional forms of the subspace based speech enhancement techniques, the singular vectors of the noisy data matrix are not filtered. From the tables, the performance of the Plain SVD and Plain GSVD methods show a meaningful distance from that of the proposed method and this may clearly indicate the effectiveness of filtering the singular vectors by a well-defined smoothing filter, as discussed in the presented paper.

CONCLUSIONS

In this paper a new algorithm for speech enhancement is presented. In the proposed approach, the effect of noise is reduced from both singular values and singular vectors. We utilize the Genetic Algorithm to optimally set the parameters needed for our proposed speech enhancement process. Some techniques are also proposed in the presented paper for controlling the trade-off between the level of noise reduction and the enhancement level of the speech quality criteria. The overall evaluation clearly indicates the better performance of our proposed method in comparison with other well-known speech enhancement techniques.

REFERENCES

1. Vaseghi, S.V., 2006. Advanced Digital Signal Processing and Noise Reduction, Third Edition. John Wiley & Sons Ltd.
2. Kim, G. and P. Loizou, 2010. Improving Speech Intelligibility in Noise Using Environment-Optimized Algorithms, IEEE Transaction on Audio, Speech, Language Processing, 18(8): 2080-2090.
3. Lee, K.C., J.S. Ou and M.C. Fang, 2008. Application of SVD Noise reduction Technique to PCA-Based Radar Target, Progress In Electromagnetic Research, PIER, 81: 447-459.
4. Krishnamoorthy, P. and S.R.M. Prasanna, 2009. Reverberant speech enhancement by temporal and spectral processing IEEE Transaction on Audio, Speech, Language Processing, 17(2): 253-266.
5. Hanssanpour, H.M. and Mesbah, B. Boashash, 2004. Time-Frequency Feature Extraction of Newborn EEG Seizure Using SVD-Based Techniques. EURASIP J. Appl. Signal Processing, 16: 2544-2554.
6. Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Transaction on Acoustic Speech Signal Processing, 27(2): 113-120.
7. Yamauchi, J. and T. Shimamura, 2002. Noise estimation using high frequency regions for spectral subtraction. IEICE Transaction. E85-A, (3): 723-727.
8. Deller, J.R., J.H.L. Hansen and J.G. Proakis, 2000. Discrete-Time Processing of Speech Signals, second edition. IEEE Press, New York.
9. Chen, J.J. Benesty, Y. Huang and S. Doclo, 2006. New Insights Into the Noise Reduction Wiener Filter. IEEE Transaction On Audio, Speech and Language Processing, 14(4): 1218-1234.
10. Gopalakrishna, V., V. Kehtarnavaz and P. Loizou, 2010. A Recursive Wavelet-Based Strategy for Real-Time Cochlear Implant Speech Processing on PDA Platforms. IEEE Trans. Biomedical Engineering, 57(8): 2053-2063.

11. Hu, Y. and P.C. Loizou, 2009. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Trans. Speech Audio Process, 12(1): 59-67.
12. Johnson, M.T., X. Yuan and Y. Ren, 2007. Speech signal enhancement through adaptive wavelet thresholding. Speech Communication, 49: 123-133.
13. Hassanpour, H., 2008. A Time-Frequency Approach for Noise Reduction. Digital Signal Processing, 18: 728-738.
14. Paliwal, K.K., 1988. Estimation of noise variance from the noisy AR signal and its application in speech enhancement. IEEE Transaction on Acoustic Speech Signal Processing, 36(2): 292-294.
15. Yamashita, K. and T. Shimamura, 2005. Nonstationary Noise Estimation Using Low-Frequency Region for Spectral Subtraction. IEEE Signal processing letters, 12(6): 105-114.
16. Martin, R., 1994. Spectral subtraction based on minimum statistics. in Proc. EUSIPCO, pp: 1182-1185.
17. Murakami, T., T. Hoya and Y. Ishida, 2005. Speech Enhancement by Spectral Subtraction Based on Subspace Decomposition. IEICE Transaction. E88-A, NO. 3.
18. Mihnea Udrea, R., N.D. Vizireanu and S. Ciochina, 2007. An improved spectral subtraction method for speech enhancement using a perceptual weighting filter. ELSEVIER, Digital Signal Processing. doi:10.1016/j.dsp.2007.08.002
19. Dendrinos M., S. Bakamidis and G. Carayannis, 1991. Speech enhancement from noise: A regenerative approach. Speech Communication, 10(2): 45-57.
20. Gray, R.M., 2010. Toeplitz and Circulant Matrices: A review. Department of Electrical Engineering, Stanford University, Stanford 94305, USA.
21. Zehtabian, A. and H. Hassanpour, 2009. A Non-destructive Approach for Noise Reduction in Time Domain. World Appl. Sci. J., 6(1): 53-63.
22. Andrews, M.S., 1998. Structured Subspace and Rank Techniques for Signal Processing Applications. Dissertation presented to the Faculty of The University of Texas at Dalllas.
23. Golub, G.H. and C.F. Van Loan, 1989. Matrix Computations. Baltimore, MD: John Hopkins University Press, 2nd ed., 1989.
24. Virginia C. Klema and Alan J. Laub, 1980. The Singular Value Decomposition: Its Computation and Some Applications. IEEE Transactions on Automatic Control, VOL AC025, NO, 2.
25. Hermus, K. and P. Wambacq, 2004. Assessment of Signal Subspace Based Speech Enhancement for Noise Robust Speech Recognition. IEEE International Conference on Acoustics, Speech and Signal Processing, pp: 17-21.
26. Huffel, S. Van, 1993. Enhanced resolution based on minimum variance estimation and exponential data modeling. Signal Processing, 33(3): 333-355.
27. Lilly, B.T. and K.K. Paliwal, 1997. Robust Speech Recognition Using Singular Value Decomposition Based Speech Enhancement. IEEE TENCON - Speech and Image Technologies for Computing and Telecommunications, pp: 257-260.
28. Hassanpour, H., S.J. Sadati and A. Zehtabian, 2008. An SVD-Based Approach for Signal Enhancement in Time Domain. IEEE International Workshop on Signal Processing and Its Applications, WOSPA 2008, Sharjah, U.A.E, pp: 10-20.
29. Luo, J., K. Ying and J. Bai, 2005. Savitzky-Golay smoothing and differentiation filter for even number data. Signal Processing, 85(7): 1429-1434.
30. Sivanandam S.N. Deepa, 2008. Introduction to Genetic Algorithms. Springer.
31. Kitawaki, N. and T. Yamada, 2007. Subjective and Objective Quality Assessment for Noise Reduced Speech. ETSI Workshop on Speech and Noise in Wideband Communication.
32. ITU-T Rec, P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, International Telecommunications Union, Geneva, Switzerland, 2001.
33. AQM in TEMS Automatic- PESQ. Technical Paper, www.ericsson.com /solutions /tems /library/tech_papers/automatic/AQM_in_TEMS_Automatic_PESQ, 2006.
34. Hu, Y. and P. Loizou, 2006. Evaluation of objective measures for speech enhancement. Proceedings of INTERSPEECH2006, Philadelphia, PA..
35. Jensen, S.H., P.C. Hansen, S.D. Hansen and J.A. Sørensen, 1995. Reduction of broad-band noise in speech by truncated QSVD. IEEE Transactions on Speech Audio Processing, 3(6): 439-448.
36. Ju, G.H. and L.S. Lee, 2002. Speech enhancement based on generalized singular value decomposition approach. in Proc. ICSLP, pp: 1801-1804.
37. Hirsch, H.G. and D. Pearce, 2006. The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions. ISCA ITRW ASR2000, Paris, France.