

A Novel Speech Enhancement Approach Based on Singular Value Decomposition and Genetic Algorithm

Amin Zehtabian, Hamid Hassanpour, Shahrokh Zehtabian
School of Information Technology and Computer Engineering
Shahrood University of Technology
Shahrood, Iran
amin_zehtabian@yahoo.com

Vicente Zarzoso
I3S Laboratory
University of Nice Sophia Antipolis
Sophia Antipolis, France

Abstract— The Singular Value Decomposition (SVD) is a powerful tool used for subspace division. In this paper a novel approach for speech signal enhancement is presented which is based on SVD and Genetic Algorithm (GA). The method is derived from the effects of environmental noises on the singular vectors as well as the singular values of a clean speech. 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. The proposed approach clearly results in a considerable attenuation of the noise as well as retrieving the quality of the original speech. The efficiency of our proposed method is affected by a number of crucial parameters which are optimally set by utilizing the GA. Extensive sets of experiments have been carried out for both of additive white Gaussian noise as well as different types of realistic colored noise cases. The results of applying six superior speech enhancement methods are then evaluated by the objective (SNR) and subjective (PESQ) measures.

Keywords - speech enhancement; SVD; singular vectors; Savitzky-Golay filter; genetic algorithm

I. INTRODUCTION

In a large number of speech applications such as automatic voice recognition and speaker authentication systems, cellular mobile communication and hearing aid devices, speech enhancement and noise reduction are the necessary pre-processing stages [1-4].

There are two important points often required to be considered in speech enhancement applications; eliminating the undesired noise from the speech to improve the Signal-to-Noise Ratio (SNR), and retrieving the quality of the original speech signal which leads to improvement of the speech intelligibility. Therefore, there may be a trade-off between the residual noise and the speech quality in a speech enhancement application and the success of the speech enhancement approaches often depends on satisfying both of the objective and subjective goals. In practice, it is really hard or even impossible to satisfy all of the goals at the same time [5].

The nature of the environmental noise is another important factor which significantly affects the performance of the speech enhancement method and constrains its application. Although it seems impossible to design an approach which is able to overcome all kinds of the noise process, but 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 colored.

II. BACKGROUND

Among the existing speech enhancement methods, the Wiener filter is an actually effective solution that is widely used by researchers and is utilized in many technical applications [1]. This method estimates the optimal noise reduction filter by using the signal and noise spectral characteristics. In the 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. In some speech enhancement applications using the Wiener filter may result in some signal degradations. Especially when the SNR value for a noisy speech signal is low, using this method may just aggravate the quality of the speech. This is due to this fact that in the Wiener filtering techniques, the amount of noise reduction is generally proportional to the speech degradation [6]. Therefore, the lower SNR conditions necessitate the more noise reduction and consequently it causes more speech distortion. Fortunately there are some ways to control the balance between the noise reduction and speech distortion that make the Wiener filter still desirable [6].

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 [7]. 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 [8]. In this approach, the enhancement is the result of thresholding on the adapted BWT coefficients.

On the other hand, there are several speech enhancement methods that use spectral subtraction for reducing the noise and can be categorized as frequency domain approaches [9-12]. In the spectral-based methods, 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.

The authors in [13] improved the spectral subtraction technique and proposed a novel technique 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 is 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.

The subspace based approaches have also wide applications in speech enhancement. These techniques usually represent the noisy speech signal in a time data matrix which often has the Hankel or Toeplitz forms [14]. 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 the signals [15]. The mentioned SVD-based technique was designed for a twofold noise reduction and was able to decrease the effects of additive noise from the singular values as well as the singular vectors of a noisy signal. The results of applying the mentioned approach to some stationary and non-stationary synthetic noisy signals have demonstrated its prominence in signal enhancement compared with other time domain methods.

In the presented 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 colored 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 traditional SVD-based method which only enhances the singular values (without filtering the singular vectors), the iterative Wiener filtering and finally the adaptive Bionic Wavelet Transforming technique (BWT).

III. THE TRADITIONAL SUBSPACE-BASED SPEECH ENHANCEMENT

The signal subspace based approaches have very extensive applications in speech processing. The basic idea behind this sort of approaches is to approximate the matrix derived from the noisy data, with another matrix of lower rank from which the reconstructed signal is derived [16]. The rank of a matrix can be directly determined by the number of nonzero singular values from its SVD.

A. Basic Theory

When a signal is corrupted by the noise, its singular values are affected and changed in a random manner. Therefore, the main target of any subspace-based signal enhancement technique may be retrieving the original singular values as much as possible. At the beginning, it is assumed that the clean speech is corrupted by an additive white Gaussian noise process and then we will develop our proposed technique in the next sections to overcome the colored noises. The white noise is an uncorrelated process with a wide frequency activity and equal power at all frequencies [1].

In speech processing applications, to reduce the complexity of the procedures it is common to divide the speech signal into some overlapping frames. From all frames, the noisy signal model in the time domain is given by

$$X_n = X_s + W_n \quad (1)$$

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 is a square matrix, in which all of the elements are the same along any northeast to southwest diagonal.

Supposing $X_n(i)$, $i = 0, 1, \dots, N$ represents the noisy signal in the time domain, the Hankel matrix $H \in \mathbb{R}^{P \times Q}$ is constructed as follows

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

where, $P + Q = N + 1$ and $P \geq Q$ [17]. Note from (1) that a similar relation can be established between the Hankel matrices

$$H_n = H_s + H_{wn} \quad (3)$$

where H_n , H_s 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

$$H = U \Sigma V^T \quad (4)$$

where $U_{P \times P}$ and $V_{Q \times Q}$ 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 $\sigma_{ij} = 0$ if $i \neq j$ and $\sigma_{ij} > 0$ if $i = j$. Consequently, it can be shown that $\sigma_{11} \geq \sigma_{22} \geq \dots > 0$ are the nonzero singular values of the matrix H .

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

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

Where $\hat{\Sigma}_s$ and $\hat{\Sigma}_n$ respectively represent the singular values which are supposed to be relevant to 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}_n^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 (3) and (7) yields

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

and

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

Since the matrices \dot{H}_s and \dot{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 \dot{H}_{wn} subspace and utilizing the \dot{H}_s matrix in reconstruction of the enhanced data matrix.

From (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 be supposed to belong to the noise subspace. Finding this point is a critical step in the proposed speech enhancement technique since an improper selection may result in an insufficient noise reduction or even an excessive noise removal. In such situations, both of the subjective and objective measurements may be disappointing. The next sub-section provides a brief review of the existing threshold point estimation algorithms and in the

fourth section, a novel technique will be presented to find an optimized point.

As discussed in [15], 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

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

B. An Introduction to the Threshold Point Estimation(TPE) Techniques

As stated in the previous subsection, a precise threshold point must be defined in the singular values matrix of the noisy signal for a proper subspace division. The researchers have developed some methods to calculate this point accurately [18, 19]. 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 must be 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 belong to the noise subspace and must be filtered.

Least Squares Approximation Method (LSA); In this method, the noise variance is supposed to be calculated from the non-speech frames of the signal and then an approximation for the original signal matrix can be obtained.

Minimum Variance Approximation Method (MVA); In this approach before reproducing the reduced rank data matrix, the singular values are transformed using a diagonal matrix. In comparison with the LSA approach, using minimum variance approximation method often leads to a better speech recognition performance.

Maximum Changes in the Slope of Curve (MCSC); In our previous article [17], we proposed calculating the maximum changes in the slope of the singular values curve to obtain the threshold point. The MCSC method utilizes a nearly uncomplicated algorithm which is able to find the threshold point properly and quickly [17]. In the next section, a novel speech enhancement approach is presented which proposes good solution for finding the seemingly more optimized threshold point as well as some other crucial parameters used for speech enhancement.

IV. THE PROPOSED GA-SVD METHOD

Regarding the basic theories concerning the subspace-based signal enhancement, when a speech signal is infected with an additive noise, its singular values are affected and changed. On the other hand, after precisely evaluating the effects of additive noise on various speech signals and executing many experiments, it is deduced that by reducing the noise from the singular values per se, some noisy data will be still available in the structure of the signal. Indeed the noise causes the singular vectors (which can be supposed as the span bases of the signal) vary randomly. Thus, in addition to the singular values enhancement, the singular vectors can be also filtered for further noise reduction.

A. Enhancing the Singular Vectors

To reduce the effect of noise from Singular Vectors (SVs) which are treated as time-series, we utilize the Savitzky-Golay filter [20]. In the Savitzky-Golay approach, each value of the series is replaced with a new value which is obtained from a polynomial fit to $2k' + 1$ neighboring points. The parameter k' is equal to, or larger than the order of the polynomial. 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 polynomial is denoted by $S.G_{deg}$ and the number of consecutive samples (or the window length of the Savitzky-Golay filter) is shown by $S.G_{win}$. Filtered SVs can be then obtained as follows

$$U_e^i = F(\bar{U}_s^i), i = 1, \dots, P \quad (12)$$

$$V_e^i = F(\bar{V}_s^i), i = 1, \dots, Q \quad (13)$$

where $F(\cdot)$ denotes the Savitzky-Golay filter function, \bar{U}_s and \bar{V}_s are the singular vectors correspond to the signal subspace (refer to Equation 7), U_e^i and V_e^i are the enhanced singular vectors after applying the Savitzky-Golay filter, and the integer variable i is the sample index.

B. Enhancing the Singular Values

In previous subsections, some of the most common techniques for subspace division were introduced briefly. In the presented paper, we propose a novel technique for finding the seemingly 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 procedure (GA-TE) will be explained in the following subsection.

C. Applying GA for Parameter Setting

The previous subsections introduced some crucial parameters affecting the performance of the proposed speech enhancement method. They include the number of rows in the Hankel data matrix "l", the optimum threshold point needed for space subdivision " P_{thr} ", the degree of polynomial " $S.G_{deg}$ ", and the window size of the Savitzky-Golay filter " $S.G_{win}$ " used for filtering the singular vectors. To optimally setting the mentioned parameters, we specify a well-defined cost function (Equation 14) and then use the genetic algorithm to minimize this function. The GA is an iterative algorithm which randomly chooses some values from a search space in each repetition [21]. Hence we define our proposed cost function as below

$$\begin{aligned} Cost(l, P_{thr}, S.G_{deg}, S.G_{win}) &= (1 - \alpha) \left(\sum_t |X_e(i) - X_n(i)| \right) \\ &+ \alpha \sum_t |X_e(i+1) - X_e(i)| \end{aligned} \quad (14)$$

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 (14), the first term indicates the distance between the enhanced speech and the noisy speech. This distance must be tuned intelligently due to this fact that the enhanced signal should still be similar to the noisy signal after filtering since this is the only thing we know about the shape and structure of the original signal. The second term also indicates the smoothness of the enhanced speech signal. The parameter α is the smoothing factor and must be 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 $\alpha = 0.5$) may be useful. It must 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 seems necessary to manage precisely the smoothness of the final enhanced signal.

D. Performance Comparison of the TPE Techniques

To calculate the performance of the five pre-mentioned threshold point estimation techniques, all of them are implemented. In this experiment, ten random noisy speech signals are provided using the AURORA database and then the additive white Gaussian noise is added to the signals at 0, +2, +5, and +10dB SNR levels. Table 1 presents the averaged SNR improvement after applying the five algorithms to the ten noisy speech signals. The results of this experiment can reasonably convince us to apply the GA-TE method for choosing the appropriate threshold point and filtering the singular values.

E. The Relationship Between Noise Reduction and Speech Quality

There are two important goals often interested in speech enhancement applications: reducing the undesired noise from the speech and improving the perceptual quality and audibility of the noisy speech signal. In this subsection we discuss on the two parameters which may affect the relationship between the noise reduction and speech quality in our proposed GA-SVD method.

α Effect; As discussed before, α is a factor determining the smoothness of the enhanced signal and must be chosen between 0 and 1. Selecting the smoothness factor α depends on the signal type and the application, so can be accomplished experimentally. In speech enhancement applications, the smoothness factor is better to be determined as a balanced value ($\alpha = 0.5$), whereas the characteristics of the speech signals may vary more randomly.

K_{red} Effect; In the presented article, since the noisy signals are supposed to be speech and it is important to preserve the details of the signal, we propose to reduce the noise subspace's singular values by a proper reduction factor (instead of setting them to zero) and hence try to retain the quality of the speech as well as improving its signal to noise ratio. Therefore, the enhanced singular value matrix can be achieved by

$$\Sigma_e = \begin{bmatrix} \hat{\Sigma}_s & 0 \\ 0 & \hat{\Sigma}_n * K_{red} \end{bmatrix} \quad (15)$$

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.

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

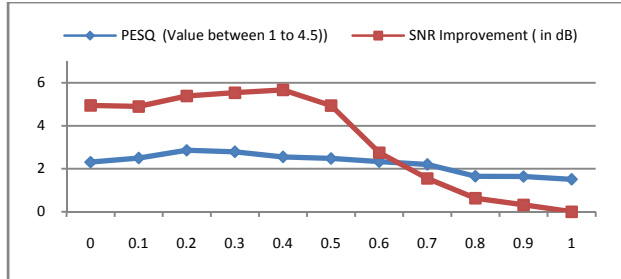


Figure 1. Plot of PESQ and SNR improvement versus reduction factor K_{red} , for a given noise-reduced speech.

Fig. 1 presents the PESQ and SNR improvement for a noise-reduced speech as an example in the case of white additive 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. In this experiment we may choose $K_{red} = 0.2$ to obtain the most desired results.

F. Reconstruction of the Noise-Reduced Speech Signal

The enhanced data matrix is given by

$$H_e = U_e^i \Sigma_e V_e^{iT} \quad (16)$$

Where the orthogonal matrices U_e^i and V_e^i are the enhanced versions of the left and right singular vectors and Σ_e represents the enhanced singular values matrix. The noise-reduced signal X_e is then extracted as follows

$$X_e = [H_e(1,1) \dots H_e(1, Q), H_e(2, Q) \dots H_e(P, Q)] \quad (17)$$

G. Experimental Result

In this sub-section, the five well-known speech enhancement approaches and the proposed GA-SVD method are implemented to evaluate their performance in reducing the effect of additive white noise from the speech signals. The methods applied in the experiments 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 (it is called as Pure SVD method in this section), the Spectral Subtraction approach and its improved version called as Spectral Over-Subtraction, the Bionic Wavelet Transform (BWT) and the proposed GA-SVD method. Note this fact that all of the methods are first precisely optimized for this speech enhancement application and their performance are then compared with together.

The speech signals used in these experiments are taken from the AURORA database. 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.

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).

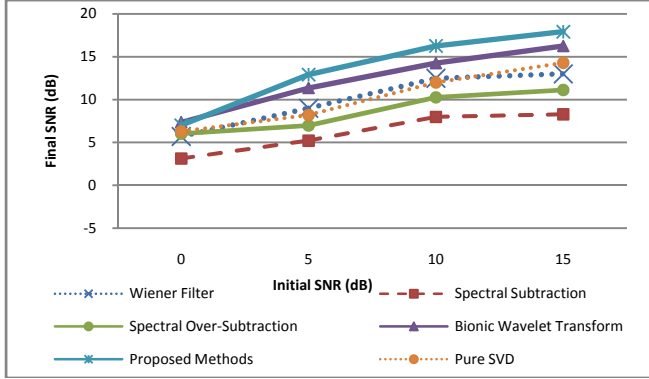


Figure 2. SNR results for white Gaussian noise case at varying SNR levels (0, +5, +10 and +15 dB).

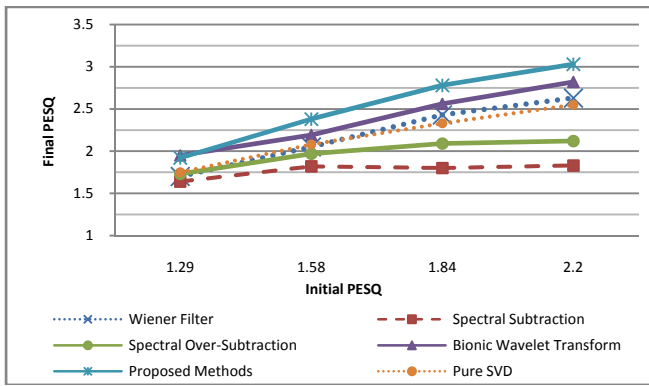


Figure 3. PESQ results for white Gaussian noise at varying SNR levels (0, +5, +10 and +15 dB).

The six speech enhancement algorithms are then implemented on each noisy speech and consequently the averaged SNR and PESQ results are drawn in Fig. 2 and Fig. 3, respectively. Note that in Fig. 3, each initial PESQ level is determined at the corresponding initial SNR value of the noisy speech.

V. EXTENSION TO THE COLORED NOISES

In the preceding section, the performance of the novel GA-SVD speech enhancement method in reducing the additive white noise was described clearly and its considerable prominence compared to the other methods was demonstrated. In this section, the performance of the proposed method is evaluated at the presence of colored noise process. The colored 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 colored noisy signals are relatively different from that of the white noise case, it may be more difficult to discriminate the principal values and vectors associated to the signal from those associated to noise. Solving this problem, we apply the GSVD (Generalized Singular Value Decomposition) algorithm which has a well-defined implicit whitening level interiorly.

Indeed, the GSVD concept is an extension of the truncated Quotient SVD (QSVD) theory, which is clearly described in [22] and its effectiveness in reducing the colored noise is well proved [23]. Utilizing the GSVD, the novel speech enhancement procedure described in the previous sections can be modified and easily

extended to reduce the effect of colored noise from the speech. The results of applying the proposed method to the speech signals infected by colored noises are described in the following.

A. Babble Noise Condition at 10 dB SNR

The proposed GA-GSVD method is now applied to an arbitrary speech signal which is corrupted by a 10 dB Babble noise process. Babble noise is considered as one of the most well-known colored noises. The time domain representations and the time-frequency spectrums of the speech signals are provided in Fig. 4 and Fig. 5. The SNR result of the enhanced speech (illustrated in Fig. 4-c) proves a considerable enhancement in the signal-to-noise ratio.

B. Monte-Carlo Simulation

In this section, the six speech enhancement methods are applied to a variety of speech signals which are infected by three famous sorts of the colored noises; the Pink, the Factory and the Babble noise processes. The clean speech signals and the additive noises are respectively taken from the AURORA and NATO-RSG10 databases. In the presented experiment, each method is implemented by ten times on the signals and the gained results are then averaged and drawn in Table 2.

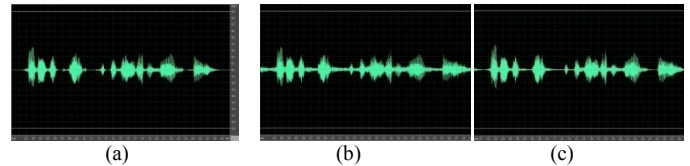


Figure 4. The time domain representation of (a) an arbitrary clean speech signal, (b) the speech signal which is corrupted by a 10 dB Babble noise process, (c) the noise reduced speech with SNR= 13.7 dB

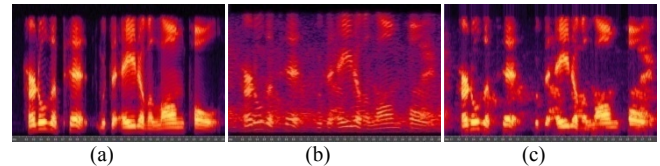


Figure 5. The Time-Frequency representation of (a) an arbitrary clean speech signal, (b) the speech signal which is corrupted by a 10 dB Babble noise process, (c) the noise reduced speech

TABLE 2. SNR IMPROVEMENT RESULTS FOR COLORED NOISE CASE AT VARYING SNR LEVELS (0, +5 AND +10 DB).

Methods	SNR Improvement (in dB)								
	Pink Noise			Factory Noise			Babble Noise		
	0 dB	5 dB	10dB	0 dB	5 dB	10dB	0 dB	5 dB	10dB
Iterative Wiener	2.40	1.30	0.88	1.97	1.04	0.80	2.25	1.05	0.64
Pure 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

VI. DISCUSSION

From the figures and tables, the proposed speech enhancement technique and the BWT method have the best performances in reducing the effect of additive noises from the speech signals. In lower initial SNR values, the performance of BWT methods is close to or even better than that of the proposed method. But while the SNR increases, the novel SVD-based method excels the BWT. The performance of the two Spectral-based methods is also heavily dependent on the initial SNR level 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. Note that in the case of the Wiener filtering method, the parameters of the approach are optimally tuned to gain a balance between the noise reduction and quality improvement [6]. But the results are not still satisfying compared to the proposed method. In the traditional SVD and GSVD approaches, the singular values of the data matrix are filtered for speech enhancement, while the singular vectors of the noisy data matrix are not enhanced. Finding the optimum crucial parameters, utilizing a proper reduction factor and reducing the effect of noise from the noisy singular vectors, result in a meaningful distance between the results achieved by the traditional methods and the novel proposed method.

VII. CONCLUSION

In this paper a new algorithm for speech enhancement is presented. In the proposed approach, the effect of noise is reduced from the singular values as well as the singular vectors. We utilize the Genetic Algorithm for optimally setting the parameters needed for our proposed speech enhancement process. In the case that the additive noise does not have the white noise characteristics, the GSVD operation is used for subspace division. The results indicate the better performance of our proposed method in comparison with other well-known speech enhancement techniques.

REFERENCES

- [1] S.V Vaseghi , *Advanced Digital Signal Processing and Noise Reduction*, Third Edition. John Wiley & Sons Ltd, 2006.
- [2] J. Btocker, U. Parlitz, M. Ogorzalek, "Nonlinear Noise Reduction," *Proceeding of the IEEE*, vol. 90, NO. 5, MAY 2002.
- [3] K.-C. Lee, J.-S. Ou, and M.-C. Fang, "Application of SVD Noise reduction Technique to PCA-Based Radar Target", *Progress In Electromagnetic Research, PIER* 81, 447–459, 2008.
- [4] T. Athanaselis, S.E. Fotinea, S. Bakamidis, I. Dologlou, and G. Giannopoulos, "Signal Enhancement for Continuous Speech Recognition," *ICANN/ICONIP, LNCS* 2714, pp. 1117–1124, 2003.
- [5] H. Hansanpour, M. Mesbah, B. Boashash, "Time-Frequency Feature Extraction of Newborn EEG Seizure Using SVD-Based Techniques," *EURASIP Journal of Applied Signal Processing*, no. 16, pp. 2544–2554, 2004.
- [6] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New Insights Into the Noise Reduction Wiener Filter", *IEEE Transaction On Audio, Speech, and Language Processing*, vol. 14, no. 4, pp.1218- 1234, 2006.
- [7] M. Bahoura, J. Rouat, "Wavelet speech enhancement based on time-scale adaptation," *Speech Communication*, vol.48. pp 1620–1637, 2006.
- [8] M.T. Johnson, X.Yuan and Y. Ren, " Speech signal enhancement through adaptive wavelet thresholding," *Speech Communication*, vol. 49. pp 123–133, 2007.
- [9] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transaction on Acoustic Speech Signal Processing*. vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.
- [10] K. Yamashita and T. Shimamura, "Nonstationary Noise Estimation Using Low-Frequency Region for Spectral Subtraction," *IEEE Signal processing letters*, vol. 12, NO. 6, June 2005
- [11] J. Yamauchi and T. Shimamura, " Noise estimation using high frequency regions for spectral subtraction," *IEICE Transaction*. vol. E85-A, no. 3, pp. 723–727, Mar. 2002.
- [12] T. Murakami, T. Hoya, Y. Ishida, "Speech Enhancement by Spectral Subtraction Based on Subspace Decomposition," *IEICE Transaction*, vol.E88–A, NO.3 March 2005.
- [13] R. Mihnea Udrea , N. D. Vizireanu, S. Ciochina, "An improved spectral subtraction method for speech enhancement using a perceptual weighting filter," *ELSEVIER, Digital Signal Processing*, 2007
- [14] R. M. Gray, "Toeplitz and Circulant Matrices: A review," *Department of Electrical Engineering, Stanford University, Stanford 94305, USA*, year. 2006.
- [15] A. Zehtabian and H. Hassanpour, "A Non-destructive Approach for Noise Reduction in Time Domain," *World Applied Sciences Journal* 5 (2), 2008.
- [16] P. C. Hansen and S. H. Jensen, "Subspace-Based Noise Reduction for Speech Signals via Diagonal and Triangular Matrix Decompositions: Survey and Analysis," *EURASIP Journal on Advances in Signal Processing*, doi: 10.1155/2007/92953, 2007.
- [17] H. Hassanpour, S.J. Sadati and A. Zehtabian, "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*, 10-20 March 2008.
- [18] S. Van Huffel, "Enhanced resolution based on minimum variance estimation and exponential data modeling," *Signal Processing*, vol. 33, no 3, pp. 333-355, sept. 1993.
- [19] B. T. Lilly and K.K. Paliwal, "Robust Speech Recognition Using Singular Value Decomposition Based Speech Enhancement", *IEEE TENCON - Speech and Image Technologies for Computing and Telecommunications*, pp. 257-260, 1997.
- [20] J. Luo, K. Ying and J. Bai, "Savitzky-Golay smoothing and differentiation filter for even number data", *Signal Processing*, Vol. 85, No. 7, pp. 1429-1434, 2005.
- [21] S.N.Sivanandam, S.N.Deepa., *Introduction to Genetic Algorithms*, Springer, 2008.
- [22] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Transactions on Speech Audio Processing*. vol. 3, no. 6, pp. 439–448, 1995.
- [23] G.-H. Ju and L.-S. Lee, "Speech enhancement based on generalized singular value decomposition approach," in *Proc. ICSLP*, pp. 1801–1804, 2002.