Robust Speech Enhancement Based on Singular Value Decomposition and an Efficient Threshold

AJGOU Riadh [1,2]
[1]Department of electrical engineering, faculty of technology, El-oued University. Algeria
[2]department of Electrical engineering, LESIA Laboratory, Med Khider University of Biskra .Algeria. Riadh-ajgou@univ-

SBAA Salim
department of Electrical engineering, LESIA Laboratory, Med Khider University of Biskra .Algeria
s_sbaa@yahoo.fr

GUENDIR Said [1,2]
[1]Department of electrical engineering, faculty of technology, El-oued University. Algeria)
[2]department of Electrical engineering, LESIA Laboratory, Med Khider University of Biskra .Algeria
said-guendir@univ-eloued.dz

GUETTAL Lemya
department of Electrical engineering, LESIA Laboratory, Med Khider University of Biskra .Algeria
ajgousari@yahoo.ca

Abstract— **In this paper, an approved method is introduced for speech enhancement corrupted by additive white Gaussian noise. Our approach is based on filtering the singular value which is obtained from singular value decomposition (SVD) using HANKEL matrix . The efficiency of the proposed methods is its efficient threshold algorithm that maintains the singular value corresponding to original speech signal and also its intelligibility enhancement by using Savitzky-Golay (SG) filter that reduce musical noise, which is usually present after noise reduction. At first, the algorithm extracts the maximum of singular value (SV) of each HANKEL matrix, so that a SV vector will be obtained representing speech signal energy. We extract the maximum and the minimum of SV vector to estimate the threshold. After filtering the SV, the inverse operation is performed to reconstruct the new Hankel matrices and the enhanced signal. This method will be evaluated in segmental SNR and Perceptual Evaluation of Speech Quality scores measure (PESQ), which led to remarkable results in comparing with traditional methods.**

Keywords—SVD; HANKEL matrix; subspace; Savitzky-Golay.

I. INTRODUCTION

The signal subspace algorithm was originally developed by Ephraim and Van Trees [1] for white input noise and was then extended to deal with colored noise by Hu and Loizou [2].

The underlying principle of the subspace algorithm is to divide the space generated by a noisy signal into two subspaces, one relating to the signal and the other to the additive white noise [3].

Signal subspace approach of speech enhancement used by Ephraim and Van trees employed Eigen Value Decomposition (EVD) and used Karhunen-Loeve Transform (KLT) to project clean speech into signal plus noise subspace called the signal subspace and removed noise which falls in the orthogonal noise subspace.

In the literature, numerous attempts have been reported for the techniques that have been used for the noise reduction in speech signals using subspace approaches based SVD [4,5,6, 7]. Moreover, the most speech enhancement techniques may suffer from how to know the noise variance, spectrum, and prior SNR estimation [6, 7]. The other controversial subjects affecting improvement performance are the thresholding function that filtering the singular value which is obtained from SVD [5, 8]. The suitable threshold value will induce to better assessment result, especially in assessment and SNR optimization and PESQ (Perceptual Evaluation of Speech Quality scores measure) [9].

The main objective of this work is to propose an efficient algorithm of speech enhancement based SVD which does not need any information about noise variance or spectrum.

In this algorithm, a thresholding function is used in the singular value (SV) filtering. Also, our proposed approach has the capability to reduce the musical noise (the high frequencies are accentuated), which is usually present after signal denoising in any noise reduction.

Although, Savitzky-Golay (SG) filters are more effective at preserving the pertinent high frequency components of the signal, thus in this paper we have used SG filter [10, 11]. SG smoothing filters are typically used to "smooth out" a noisy signal whose frequency span (without noise) is large. In this type of application, SG smoothing filters perform much better than standard averaging FIR filters, which tend to filter out a significant portion of the signal's high frequency content along with the noise [10]. SG filters are optimal in the sense that they minimize the least-squares error in fitting a polynomial to frames of noisy data [10].

In this paper, we have described the subspace principle based on the decomposition of the noisy speech in (SVD), and their applications in denoising of speech signals. In the last section, we will discuss the results of speech signal enhancement and the performance of our method. The results of the proposed algorithm to a noisy speech signals in different SNR levels corrupted by additive white Gaussian noise are reported is comparing with some speech enhancement algorithms.

II. OUR APPROACH

In this section we propose an enhancement approach based on the decomposition in singular values SVD whose enhancement is made frame by frame. The enhanced signal is reconstituted by grouping all the enhanced frames. The enhanced signal passed through a smooth filter that is Savitzky-Golay filter that is essentially used to smooth out the noisy data [12]. Also, describing the subspace principle based on the decomposition of the noisy speech in singular value decomposition (SVD).

A. Segmenting speech and enhancement frame by frame

The noisy signal is segmented into frames of 34 ms with overlapping manner. In this way the signal can be regarded as quasi-stationary. To avoid edge effects during the reconstruction of the signal, we have used 50% overlap between frames using Hanning window. The implementation of this approach is carried out by making a matrix comprising one frame per line of which each line representing a frame and to be enhanced. Figure 1 presents the procedure for creating the frame matrix from a speech signal. The enhancement method proposed in the subject is based on the decomposition in subspace. The principle is to dividing the speech signal into two subspaces vectors, one relating to the signal, the other to additive noise. Either a random signal $s(k)$:

$$s(k) = A * e^{jw_0 k} \tag{1}$$

This signal is assumed stationary, consisting of N samples. A is a random variable, w_0 is a constant. We suppose that this signal is noised by Gaussian white noise $b(k)$, with zero mean and variance σ^2. It is assumed that the signal and the noise are uncorrelated. $y(k)$ is the noisy signal, so that:

$$y(k) = s(k) + b(k) \tag{2}$$

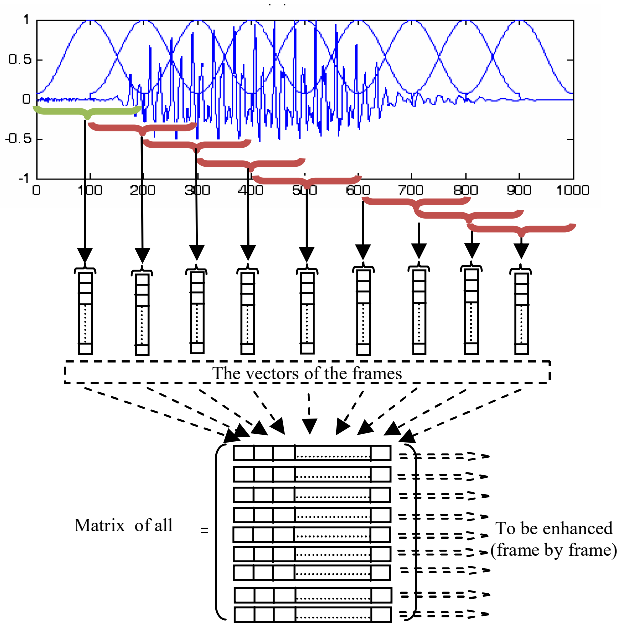

Fig. 1. Decomposition of a speech signal into frames by a Hanning window and a 50% overlap and construction of frame matrix.

From equation (2) we have:

$$Y = \begin{bmatrix} y(1) \\ . \\ . \\ . \\ y(N-1) \end{bmatrix}, \quad S = \begin{bmatrix} 1 \\ . \\ . \\ . \\ e^{jw_0(N-1)} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b(1) \\ . \\ . \\ . \\ b(N-1) \end{bmatrix} \tag{3}$$

Then: $Y = A * S + B$ (4)

then calculate the autocorrelation matrix of $y(k)$:

$$R_y = E\left[\left(Y.Y^{T*}\right)\right] \tag{5}$$

from equation (5), we have:

$$R_y = E\left[|A|^2\right] S * S^{T*} + Rb \tag{6}$$

where $Rb = \sigma^2.I$ (7)

and: $R_y = \mu.S.S^{T*} + Rb$ (8)

we note $\Psi_0,...,\Psi_{N-1}$ as the eigenvectors of R_y matrix associated with p eigenvalues $(\lambda_0,...,\lambda_{N-1})$.

then: $R_y * S = (\mu.S.S^{T*} + Rb).S = \left(\mu.N + \sigma^2\right).S$ (9)

As a result, S is an eigenvector of the autocorrelation matrix R_y, associated with the eigenvalue $\mu.N + \sigma^2$ (strictly positive).

The remaining (N-1) eigenvalues are equal to σ^2.

Thus the study of the eigenvalues and eigenvectors of R_y give us information on the original signal and the noise:

- the minimum va-lues give the value of the noise variance.
- eigenvalue (whose value is strictly greater than σ^2) allows characterize $A.\left(\mu = E\left[|A|^2\right]\right)$
- the eigenvector S associated with the maximum eigenvalue is the samples vector of the original vector.

B. Subspace decomposition method

In practice, the autocorrelation of the noisy signal matrix analysis is impossible since we don't have an infinite number of samples. We will therefore using Hankel matrix [13, 14] of the noisy signal assuming that the original signal and the noise are uncorrelated. We create Hankel matrix and perform SVD, and then we filter singular values using a threshold that is developed. The algorithm operates in three stages:

1) Construction of the Hankel matrix

This denoising strategy begins with the creation of a Hankel data matrix by dividing a noisy speech signal into partially overlapping segments and rearranging them into the following array [15]:

$$H_y = \begin{bmatrix} y(1) & y(2) & ... & y(M) \\ y(2) & : & & : \\ : & : & & : \\ y(L-1) & y(L) & ... & y(N-1) \\ y(L) & y(L+1) & ... & y(N) \end{bmatrix} \tag{10}$$

The Hankel matrix is constructed for each frame of N samples whose column number is equal to M and the number of lines is equal to L, where:

with: $N + 1 = M + L$ (11)

Where: $\left[y(1)..........y(N)\right]$ are the frame samples.

2) Estimate in the sense least square of the Hankel matrix by extraction of K dominant singular values of matrix H_y .

Hankel matrix decomposition is performed as follows [15]:

$$H_y = U \Sigma V^T = \left[U_{1,K} \quad U_{K+1,M}\right] \begin{bmatrix} \Sigma_{1,K}^{signal} & 0 \\ 0 & \Sigma_{K+1,M}^{bruit} \end{bmatrix} \begin{bmatrix} V_{1,K}^T \\ V_{K+1,M}^T \end{bmatrix} \quad (12)$$

There are three matrices, *U,* Σ and *V*. *V* contains the singular vectors of the decomposition and Σ contains the singular values of H_y. Among these singular values, there are a part $\left(\Sigma_{1,K}\right)$ corresponding to the signal and a part $\left(\Sigma_{K+1,M}\right)$ corresponding to the noise.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & .. & .. & .. & 0 \\ 0 & .. & .. & & & : \\ : & .. & \sigma_K & & & : \\ : & & & \sigma_{K+1} & & : \\ : & & & & & 0 \\ 0 & .. & .. & .. & 0 & \sigma_M \end{bmatrix} \quad (13)$$

The flanked part corresponds to the dominant SV and representative of the de-noised speech signal. However, the SV of the Hankel matrix H_y are the square roots of the eigenvalues of $H_y^{T*}.H_y$, where:

$$H_y^{T*}.H_y = \begin{bmatrix} \sum_{K=1}^{L} y(K)^2 & \sum_{K=1}^{L} y(K).y(K+1) & .... & \sum_{K=1}^{L} y(K).y(K+M-1) \\ \sum_{K=1}^{L} y(K+1).y(K) & & & : \\ : & & & : \\ \sum_{K=1}^{L} y(K+M-1).y(K) & .... & .... & \sum_{K=1}^{L} y(K+M-1)^2 \end{bmatrix}$$

(14)

It is found that: $$\tilde{R}_y = \frac{1}{L}.H_y^{T*}.H_y \quad (15)$$

$$\tilde{R}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-(m+1)} x(n+m)x^*(n) \quad (16)$$

$\tilde{R}_{xx}$ represents the biased estimator of the autocorrelation matrix of the noisy observations (assuming ergodicity is assumed). We deduce that the eigenvalues of $H_y^{T*}.H_y$ are proportional to eigenvalues of $\tilde{R}_y$, that is means the singular values of H_y are proportional to $\tilde{R}_y$ eigenvalues.

3) We derive an estimate of H_{signal}

In the least squares sense we deduce an estimate of H_{signal} , , denoted H_{signal}^{LS} , then to restore the structure of the Hankel matrix of data signal, and estimate the original speech signal, where the average value of the antidiagonals is assigned to each coefficient of the antidiagonal of the matrix H_{signal}^{LS} . Then we obtain an estimate of the signal by selecting the first column and the last line of this matrix.

C. Threshold function for selecting dominant SV

Based on equation (9), we find that $\left(\mu.N+\sigma^2\right)$ is the eigenvalue associated with eigenvector containing the samples of the original signal. The other values are equal to σ^2 .

According to the discrete Karhunen Loeve [16, 17], the average energy of the error is minimized by minimizing the function: $$E = \sum_{i=K+1}^{N} \lambda(i) \quad (17)$$

$\lambda(i)$ are the eigenvalues of R_y. Since we are working with an estimator $H_y^{T*}.H_y$ of the autocorrelation matrix so the $\lambda(i)$ are the singular values of H_y . (Singular value of H_y = square root of $H_y^{T*}.H_y$ eigenvalue). Therefore, it appears legitimate to select the dominant SV whose value is strictly greater than a threshold. To establish an adaptive threshold, the diagonal of $\sum$ matrices that represent singular values was analyzed:

$$diag(\Sigma) = [\sigma_1, \sigma_2 ....\sigma_M] \quad (18)$$

After Hankel matrix (H_y) construction, we take $diag(\Sigma)$ matrices that represent the SV for silence zones (represent noise) and speech zones. Figure 2 plots the SV of $diag(\Sigma)$ of four frames that represent the silence zone. Figure 3 plots the results of $diag(\Sigma)$ matrices of four frames from speech zone. From figures 2 and 3, we conclude than the maximum of the SV of speech zones frames are greater than those of silence zone, where the maximum singular values of silence zones frames are : 0.45, 0.54, 0.57, 0.135 and : 2.585, 4.49, 6.6, 7.10, for speech frames zone.

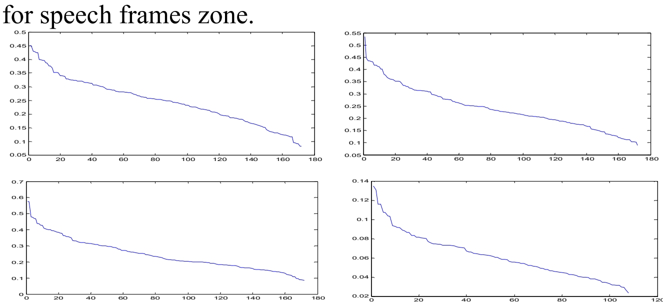

Fig. 2 Singular values of four frames of the silence zone. The abscissa axis represents SVD index and ordinate axis represents SVD.

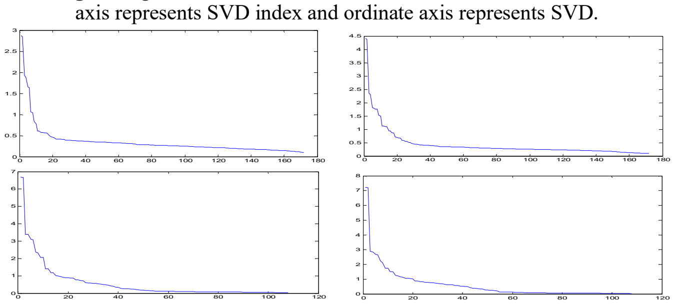

Fig.3. Singular values of four frames of the speech zone. The abscissa axis represents SVD index and ordinate axis represents SVD

In conclusion, we can assume that the singular values with low values correspond to the noise. The noise reduction is achieved by removing SV of low-energy.. This leads us to establish a threshold that should be adaptive and effective to filter SV that really represent the original signal from SVD.

In this paper we have developed an effective algorithm of threshold estimation. After signal windowing (Hanning) and 50% overlap. The algorithm creates the Hankel matrix and perform SVD, then extracting its maximum singular value for all frames. The figure 5 illustrates a speech signal and its maximum singular values contour. Where the first remark is that the maximum of singular values represent the energy of the speech signal.

The threshold is estimated by extracting the maximum SV of each HANKEL matrix ($\max(diag(\Sigma_j))$) so that a maximum singular values vector will be obtained. We extract the maximum ($\max\{\max(diag(\Sigma_j))\}$) and the minimum ($\min\{\max(diag(\Sigma_j))\}$) of the maximum SV to estimate our selection threshold. Based on equation (13) and figure 5, the proposed thresholding function is determined as:

$$threshold = \min\{\max(diag(\Sigma_j))\} + \alpha.delta \qquad (19)$$

$$j = 1,2.....v \quad (j \text{ and } v \text{ are the index and the frame length respectively})$$

$$delta = \max\{\max(diag(\Sigma_j))\} - \min\{\max(diag(\Sigma_j))\} \qquad (20)$$

where α : is a real number in the interval of] 0, 1 [depends on noise level. So that:

$$\begin{aligned} &if\ \sigma_i > threshold \\ &\quad \sigma_i \text{ accepted} \\ &otherwise \qquad (21) \\ &\quad \sigma_i \text{ rejected} \\ &\qquad i = 1,2...M \quad (i \text{ is the singular value index}) \end{aligned}$$

For the effect of α coefficient its value should be around 0.1, but for low SNR (under 0dB) and for more performance α value should be around 0.4. So, to enhance the speech signal, we suppose to keep singular values which satisfy the equation (19). Figure 5 and 6 illustrates the principle of threshold estimation of the dominant SV and the threshold estimation algorithm.

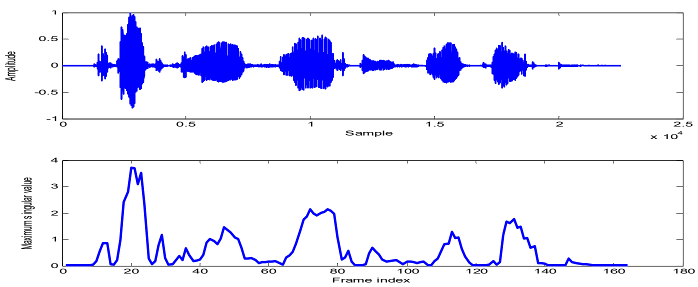

Figure 4 Speech signal and its contour of maximum singular values for each frame (singular values calculated using Hankel matrix)

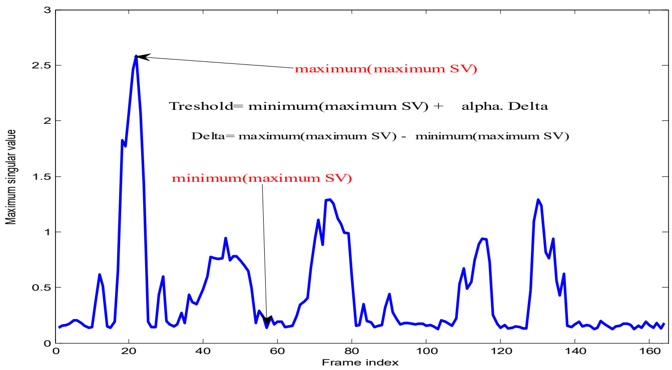

Fig. 5. The maximum SV of each frame versus frame index.

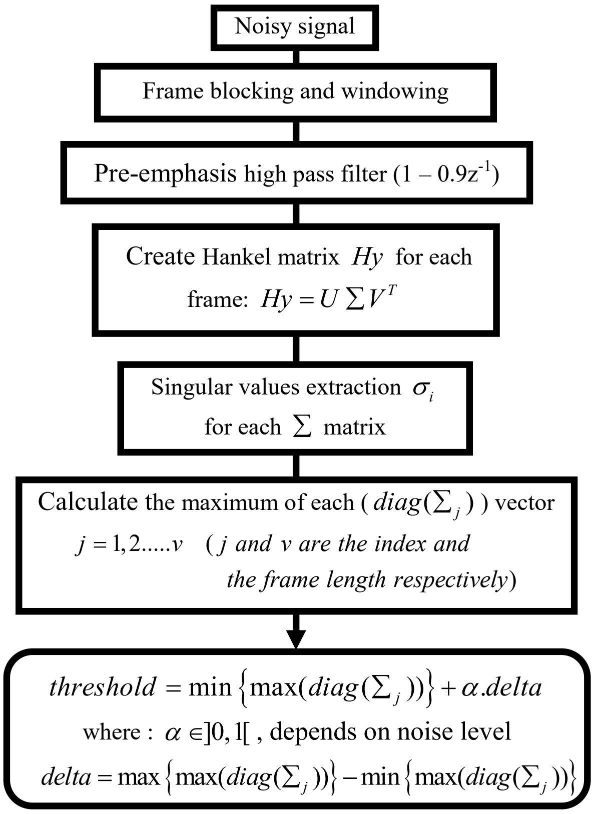

Fig. 6. Algorithm of threshold estimation of singular values.

D. Intelligibility enhancement

In this step we should improve the speech intelligibility and speech quality. Thus in this part we enhance the singular values belonging to a speech frames selected by our threshold and reduce the other SV of non-speech frames.

During the speech frames, we enhance the singular values as:

$$\tilde{\Sigma}_j = \tilde{\Sigma}_j \times (1+\beta) \qquad \beta \in]0,1[\qquad (22)$$

j : is the frame index

as default value we set $\beta = 0.32$.

where the singular values that are maintained reconstruct a new Σ matrix denotes $\tilde{\Sigma}$ matrices.

During the silence segments, we enhance the singular values as:

$$\tilde{\Sigma}_j = \tilde{\Sigma}_j \times (1-\gamma) \qquad \gamma \in]0,1[\tag{23}$$

As default value we set $\gamma = 0.9$.

We resume this section by:

- Extracting the dominant singular values by our threshold (equation 19).
- detecting the speech / non-speech frames by a voice activity detection (VAD).
- we obtain a new Σ matrix named $\tilde{\Sigma}$.
- the singular values that are maintained and constructing $\tilde{\Sigma}$ matrices were processed again by equation (22) and (23).

Therefore, we are in need for voice activity detection (VAD) algorithm which should be robust to the noisy context. Many different algorithms of voice activity detection have been proposed for several applications [18] real-time speech transmission on the internet [19] and noise reduction for digital hearing aids [20]. In this work we have chosen a VAD developed by [21].

E. Reconstruct the new Hankel matrix

The new Hankel matrix named $\tilde{H}y$ must be reconstructed, where from equation (12), the new $\tilde{H}_y$ is as follows:

$$\tilde{H}_y = \tilde{U}\tilde{\Sigma}\tilde{V}^T = \left[U_{1,K}\right]\left[\Sigma_{1,K}^{signal}\right]\left[V_{1,K}^T\right] \tag{24}$$

Where: Adaptation of the size of U and V to the reduced number of selected σ_i , so that we have $\tilde{U}$ and $\tilde{V}$.

After Hankel matrices reconstruction ($\tilde{H}_y$) based on the new matrices $\tilde{U}$, $\tilde{V}$ and $\tilde{\Sigma}$, we obtain an estimate of the enhanced frames by selecting the first column and the last line of each matrix ($\tilde{H}_y$).

F. Reconstruction of the denoised signal

Where $y(k)$ is the sample of the noisy signal, $y(k1)$ and $y(k2)$ were noted as is the sample placed in frame 1 and frame 2 respectively. w1 is the weight induced by Hanning window weighting $y(k1)$. The weight w2 weighting $y(k2)$. $y(k1)$, $y(k2)$ samples are noisy samples, figure 7 illustrate the reconstruction principle of denoised signal. $s(k)$ is the sample of the signal we want to find (denoised).

$$s(k) = \frac{y(k1).w1 + y(k2).w2}{w1 + w2} \tag{25}$$

By this method we will recover the original signal.

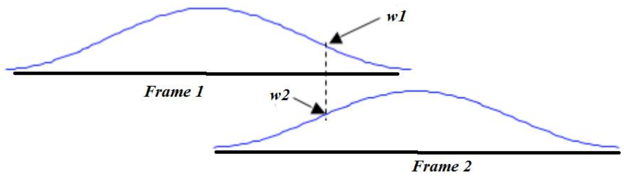

Fig.7. Reconstruction principle of the denoised signal.

G. Savitzky-Golay filter

Due to a musical noise generated after denoising where the high frequencies are accentuated. SG filters are more effective to resolve this problem preserving the pertinent high frequency components [22]. More detailed about SG filter is reported in [12]. This filter used by several researchers as in [23, 24], witch is a least-squares polynomial algorithm for smoothing of data [15]. The SG filter is mainly used to smooth noisy data by performing a local polynomial regression of degree d_{SG} over a succession of values of at least $1+d_{SG}$ points which are treated as being uniformly spaced in succession. SG Polynomial order is specified by its degree indicating up to the fitting point of each frame. The size of the frame specifies the number of samples used to perform the smoothing task for each data point.

SG filter typically requires three inputs: the noisy signal $y(k)$, the order of the polynomial d_{sg} and its frame size M_{sg} .

Filtered $\tilde{y}(k)$ (enhanced speech) can be obtained as follows:

$$\hat{y}(k) = F(\tilde{y}(k), d_{SG}, M_{SG}) \tag{26}$$

where $F(.)$ denotes the SG filter function, $\hat{y}(k)$ denoted the enhanced speech, after passing $\tilde{y}(k)$ through the SG filter.

A SG filter typically requires pre-determined values of order and frame size for its application [25]. To assess the denoising performance under high noise conditions, the presence of noise requires choosing a filter with a different impulse response length at each reconstruction instant [25]. So that, we should use the suitable values of design parameters. In the literature there some works about adaptive SG filter like [25, 26, 27]. However, to be more efficient, in this paper a simulation study was done (we have fixed the frame size of the filter and we have varied the order and select the order that corresponds to the best PESQ, SegSNR , SNR in different SNR levels).

The basic denoising algorithm based singular value decomposition can be summarized in Figure 8.

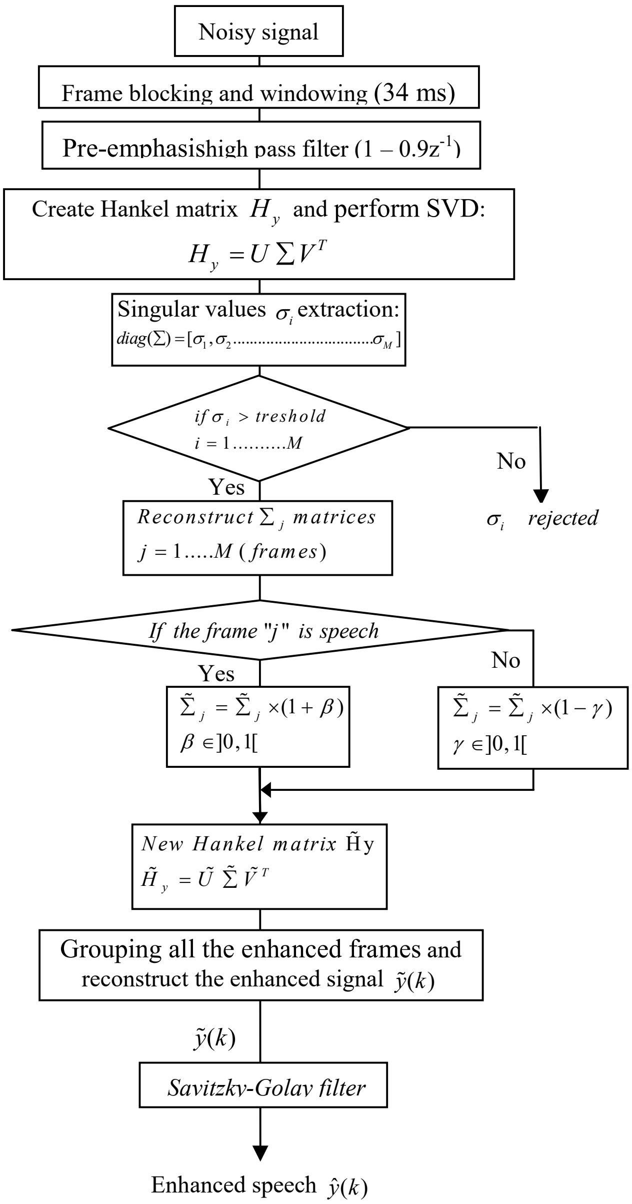

Fig.8. The process of the proposed speech enhancement algorithm based on adaptive threshold and SG filter.

III. RESULTS AND DISCUSSION

In this part we present our results of simulation and possible discussion. The proposed speech enhancement approach has been evaluated on the spoken English sentence chosen from the NOIZEUS corpus developed in the Hu and Loizou laboratory [9] which is adapted to the evaluation of the enhancement algorithms [8]. The speech database corpus including 30 cleaned speech signals and sampled at 8 kHz and quantized linearly using 16 bits resolution. For example the sentence that we have used for performance measure of our approach is spoken by a male speaker that was "*The birch canoe slid on the smooth planks*" is with 2.8161 sec duration (22529 samples). This signal named "sp01.wav".

A. Processing chain and simulation parameters

The frame length N usually 20 to 34 milliseconds the speech signal can be assumed stationary. Hence, the value of N is typically between 165 and 272 samples for 8 kHz data. In our simulation The frame size is chosen to be 34 ms. Furthermore, the overlap is 50% (17ms, 136 samples). Using the Matlab software, we will first add a WGN to our cleaned signal with desired SNRs. Then, we will separate the signal into "frames" which will first be windowed using Hanning window before being treated separately by our subspace method based SVD. Finally, the frames will be grouped together to reconstruct the enhanced signal.

The noise samples used are of zero-mean, the noisy signal is normalized to unity.

Figure 9 presents the processing chain used in our simulation, of which we take a cleaned signal from NOIZEUS corpus database, we add a White Gaussian Noise (WGN) with desired SNRs.

The parameters of the Hankel matrix (equation 10, 11) : $L = 54$ samples, $L = 219$, of which $N + 1 = M + L$, N: sample number per frame.

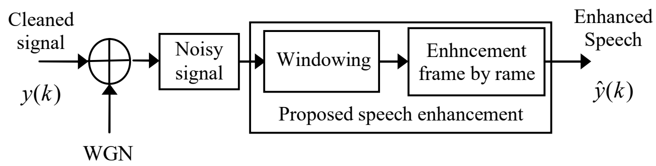

Fig. 9. Processing chain.

B. Threshold estimation at varying SNR levels

In this step, the estimated threshold values are investigated during enhancement of corrupted speech signal of different SNR levels. After several experiments on the corpus [9], as default value we set: $\alpha = 0.1$ (equation 19) for $SNR \geq 0.dB$ and $\alpha = 0.3$ for $SNR < 0.dB$, $\beta = 0.32$ (equation 22), and $\gamma = 0.9$ (equation 23). The simulation results are shown in figure 10 which reports the threshold evaluation measures obtained for speech signal "*The birch canoe slid on the smooth planks*" at varying SNR levels. From this figure it can be observed that the threshold value decreases regularly until it becomes almost zero for a high SNR (low noise), which means the effectiveness of the threshold estimation algorithm whose threshold values to a tendency to accept all singular values.

C. Performance evaluation versus SG filter coefficients

The proposed method is evaluated using Segmental SNR (SegSNR) and PESQ [9]. The PESQ method evaluates the quality of the speech signal by comparing the reference signal with the degraded signal. The PESQ algorithm models the human perception of the speech signal and thus enables the prediction of speech quality comparable to the subjective assessment as it would be performed by the human audience[28]. Of the seven basic objective measures tested by [9], the PESQ measure yielded the highest correlation with overall quality and signal distortion. Otherwise the SegSNR measure, which is widely used for evaluating the performance of speech enhancement algorithms, yielded a very poor correlation coefficient with overall quality [9], thus this measure unsuitable for evaluating the performance of enhancement algorithms without PESQ measure. In this work we have adopted Loizou's PESQ and SegSNR implementation [9]. In this step we have to search the SG filter order that can leads to the best scores of PESQ and SegSNR. We proceed to set the frame length of the SG filter to 31 samples and plot all the PESQ and SagSNR values which correspond to the order of the filter SG from 2 to 28 using the signal "*The birch canoe slid on the smooth planks*" corrupted by WGN for different SNR levels (SNR from -10 to 30dB with step =5dB). Figures 11, 12 represent PESQ and SegSNR versus SG order respectively at SNR=-10 dB. Figure 13, 14 represent PESQ and SegSNR versus SG order respectively at SNR=-5 dB. From figures 11, 12, 13, 14, it can be concluded that when the SG Golay order decreases, the PESQ value increases and vice versa for the case of SagSNR. Although, simulation results of all the best values of PESQ and SegSNR estimated at different SNR levels are repoted in table 1.

From table 1, it is shown that SG filter order isn't the same for the best PESQ and SegSNR. In this case we favorite PESQ value (PESQ measure yielded the highest correlation with overall quality and signal distortion). Furthermore, an improvement of the SNR and the PESQ implies a better intelligibility of speech signal. PESQ provides a score in the range of 1 to 5 wherever 1 is unacceptable, 2 is acceptable and 5 is excellent. Thus, we assume a minimum of accepted PESQ (PESQ=2) through this minimum and search for the SG order of which corresponds to the best SegSNR resulting. SG orders applied in our simulation are reported in table II.

D. *Comparison with different noise reduction techniques*

We also compare the proposed method with five others state of art denoising methods such as Harmonic Regeneration Noise Reduction (HRNR) [29], Speech Enhancement Based on a Priori Signal to Noise Estimation(SEPSNR) [30]. Speech Enhancement Using a Noncausal A PrioriSNR Estimator (SENPE) [31], Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay (UMNPE) [32] and Speech enhancement using a minimum mean-square error log-spectral amplitude estimator (SEMMSE) [33]. The comparison of speech enhancement algorithms including our proposed method in terms of SNRseg and PESQ score and in presence of WGN in different SNR levels, are shown in figure 15 and 16. Where we have used 'sp01.wav' corrupted with WGN according to figure 9. The PESQ of our proposed approach is higher than the overall PESQ for the other techniques at different input SNR levels but for SNR =30dB, HRNR [29] technique becomes the best. It is also concluded from the SegSNR results that our method algorithm represents the best result at input SNR from -10dB to 30dB, but for 25dB our thchnique have the same value of SegSNR as UMNPE [32] method.

It is also concluded from this result that our approach algorithm significantly depending on SG order that we should use to have the best result.

From the results, it is evident that SG filter is ideal for musical noise removal which leads to a good speech intelligibility.

Table 1. Best of PESQ and SegSNR values and corresponding SG filter order.

	PESQ		SegSNR	
Initial SNR	Best PESQ	SG order	Best SegSNR	SG order
-10dB	1.3221	6	0.2514	26
-5dB	1.8925	5	2.0336	26
0dB	2.3996	3	3.9012	27
5dB	2.6829	5	4.6191	26
10dB	3.0503	5	4.6191	26
15dB	3.3241	5	9.5933	27
20dB	3.7217	7	12.6424	27
25dB	3.9235	3	15.8007	27
30dB	4.0873	5	18.9597	27

Table I1. SG orders applied in our simulation.

Initial SNR	SG order to be used
-10dB	7
-5dB	6
0dB	14
5dB	10
10dB	13
15dB	21
20dB	25
25dB	23
30dB	27

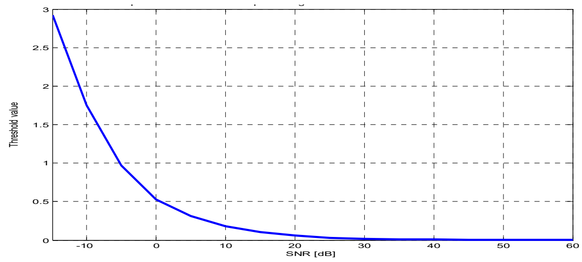

Fig. 10. The threshold value as a function of the input SNR. This threshold is adaptive versus speech signal embedded in the noise of different SNRs

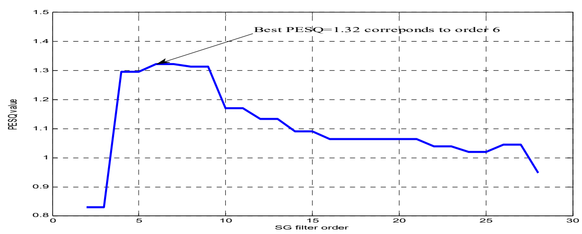

Fig.11. PESQ values versus SG filter order for speech signal at -10dB. Frame length of the filter =31 samples.

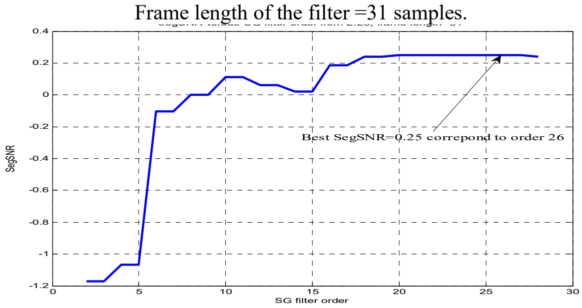

Fig.12. SegSNR values versus SG filter order for speech signal at -10dB.

Frame length of the filter =31 samples.

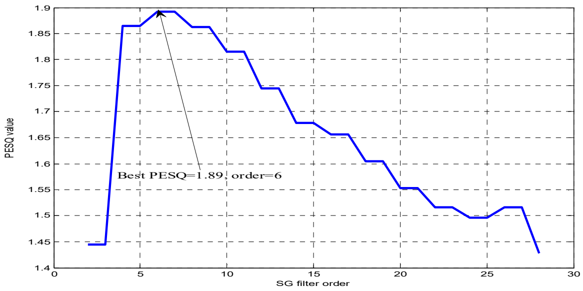

Fig. 13. PESQ values versus SG filter order for speech signal at -5dB. Frame length of the filter =31 samples.

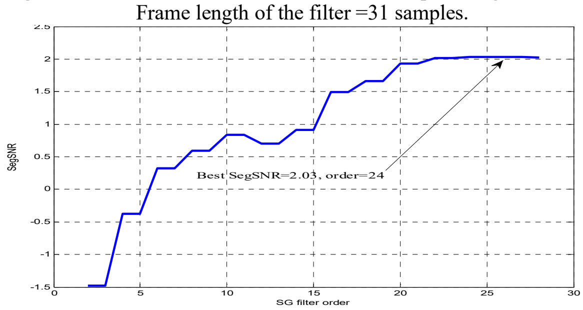

Fig.14. SegSNR values versus SG filter order for speech signal at -5dB. Frame length of the filter =31 samples.

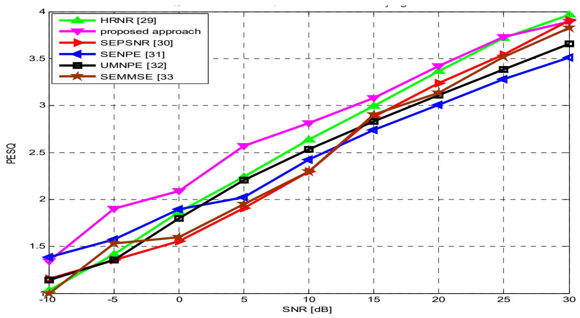

Fig. 15. PESQ assessment results with five other enhancement methods at SNR levels of -10 dB to 30dB.

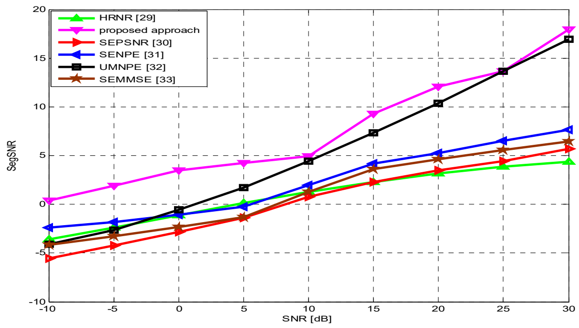

Fig. 16. SegSNR assessment results with five other enhancements methods at SNR levels of -10 dB to 30dB.

IV. CONCLUSION

In this paper, a subspace approach based on SVD performed from Hankal matrix was implemented to deal the enhancement of speech degraded by additive WGN noise. The main contribution is the efficient threshold that we have developed and Savitzky-Golay filter applied to enhance the intelligibility of speech after denoising procedure (eliminate musical noise). SG filter is ideal for musical noise removal which leads to a good speech intelligibility. Our approach has proved a good performance in terms of PESQ and SegSNR compared to five advanced methods. The main problem is how to choose the filter coefficients to have a good yield in speech denoising. Thus we are in need to implement an adaptive Savitzky-Golay filte applied in speech enhancement.

REFERENCES

[1] Ephraim Y, Van Trees HL. “A signal subspace approach for speech enhancement,” IEEE Trans Speech Audio Proc. 1995;3:251–266.

[2] Hu Y, Loizou P. “A subspace approach for enhancing speech corrupted with colored noise,” IEEE Signal Processing Letters. 2002;9(7):204–206.

[3] Loizou, P. C., Lobo, A., & Hu, Y. (2005). Subspace algorithms for noise reduction in cochlear implants. *The Journal of the Acoustical Society of America*, *118*(5), 2791-2793.

[4] Ghasemi, J. (2011, March). A new approach based on SVD for speech enhancement. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on* (pp. 371-376). IEEE.

[5] Doclo, S., & Moonen, M. (2002). GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Transactions on Signal Processing*, *50*(9), 2230-2244.

[6] Hermus, K., & Wambacq, P. (2006). A review of signal subspace speech enhancement and its application to noise robust speech recognition. *EURASIP Journal on Advances in Signal Processing*, *2007*(1), 045821.

[7] Ghasemi, J., & Mollaei, M. K. (2009). A new approach for speech enhancement based on eigenvalue spectral subtraction. *Signal processing: An international journal*, *3*(4), 34-41.

[8] Hsieh, C. T., Huang, P. Y., Chen, T. W., & Chen, Y. H. (2015, November). Speech enhancement based on sparse representation under color noisy environment. In *Intelligent Signal Processing and Communication Systems (ISPACS), 2015 International Symposium on* (pp. 134-138). IEEE.

[9] Yi et LOIZOU, Philipos C. Evaluation of objective quality measures for speech enhancement. Audio, Speech, and Language Processing, IEEE Transactions on, 2008, vol. 16, no 1, p. 229-238.

[10] ORFANIDIS, S. J. Introduction to signal processing, Sophocles J. Orfanidis, Rutgers University, 2010, vol. 13, p. 795.

[11] Orfanidis, S. J. (1995). Introduction to signal processing. Prentice-Hall, Inc.

[12] Harjeet Kaur, Rajni. (2017). A novel approach for denoising electrocardiogram signal using hybrid technique. Journal of Engineering Science and Technology Vol. 12, No. 7 (2017) 1780 - 1791.

[13] Cadzow, J. A. (1988). Signal enhancement-a composite property mapping algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*(1), 49-62.

[14] Lu, L., Xu, W., & Qiao, S. (2015). A fast SVD for multilevel block Hankel matrices with minimal memory storage. *Numerical Algorithms*, *69*(4), 875-891.

[15] Alfred, C. H., & Cobbold, R. S. (2008). Single-ensemble-based eigen-processing methods for color flow imaging-Part I. The Hankel-SVD filter. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, *55*(3).

[16] Hu, Y., & Loizou, P. C. (2003). A generalized subspace approach for enhancing speech corrupted by colored noise. *IEEE Transactions on Speech and Audio Processing*, *11*(4), 334-341.

[17] Rezayee, A., & Gazor, S. (2001). An adaptive KLT approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, *9*(2), 87-95.

[18] Freeman, D. K., Cosier, G., Southcott, C. B., & Boyd, I. (1989, May). The voice activity detector for the Pan-European digital cellular mobile telephone service. In *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on* (pp. 369-372). IEEE.

[19] Sangwan, A., Chiranth, M. C., Jamadagni, H. S., Sah, R., Prasad, R. V., & Gaurav, V. (2002). VAD techniques for real-time speech transmission on the Internet. In *High Speed Networks and Multimedia Communications 5th IEEE International Conference on* (pp. 46-50). IEEE.

[20] Bang, D. H., Kil, S. K., Kang, H. D., Yoon, G. S., & Lee, S. M. (2009). Noise cancellation using microphone array in digital hearing aids. *The Transactions of The Korean Institute of Electrical Engineers*, *58*(4), 857-866.

[21] Pang, J. (2017, January). Spectrum energy based voice activity detection. In *Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual* (pp. 1-5). IEEE.

[22] A. Savitzky and M. J. E. Golay, "Soothing and differentiation of data by simplified least squares procedures," Anal. chem., vol. 36, pp. 1627–1639, 1964. . (ref golay3).

[23] Ronald W. Schafer, "What Is a Savitzky-Golay Filter?", IEEE Signal Processing Magazine, July 2011.

[24] Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote sensing of Environment*, *91*(3), 332-344.

[25] Krishnan, S.R.; and Seelamantula, C.S. (2013). On the selection of optimum Savitzky-Golay filters. IEEE Transactions on Signal Processing, 61(2), 380-391. [ref 15]

[26] Vivó-Truyols, G., & Schoenmakers, P. J. (2006). Automatic Selection of Optimal Savitzky− Golay Smoothing. *Analytical chemistry*, *78*(13), 4598-4608.

[27] Zehtabian, A., Hassanpour, H., Zehtabian, S., & Zarzoso, V. (2010, December). A novel speech enhancement approach based on singular value decomposition and genetic algorithm. In *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of* (pp. 430-435). IEEE.

[28] A.P. Stark, K.K. Wojcicki, J.G. Lyons and K.K. Paliwal, "Noise driven short time phase spectrum compensation procedure for speech enhancement", Proc. INTERSPEECH 2008, Brisbane, Australia, pp. 549-552, Sep. 2008.

[29] Clapous, C.; Marro, C.; Scalart, P., "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, Issue 6, pp. 2098 - 2108, Nov. 2006. .

[30] Scalart, and J. Vieira Filho, "Speech Enhancement Based on a Priori Signal to Noise Estimation," IEEE Intl. Conf. Acoust., Speech, Signal Processing , Atlanta, GA, USA, Vol. 2, pp. 629–632, May 1996. .

[31] I.Cohen ."Speech Enhancement Using a Noncausal A PrioriSNR Estimator". IEEE, signal processing letters, vol. 11, no. 9, september 2004.

[32] Gerkmann, T. & Hendriks, R. C. "Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay". IEEE Trans Audio, Speech, Language Processing, 2012, 20, 1383-1393.

[33] Erkelens Ephraim, Y. & Malah, D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator IEEE Trans Acoustics Speech and Signal Processing, 33(2):443-445, Apr 1985.