An unsupervised approach for improving speech enhancement using wavelet packet transform and adaptive thresholding

Mohammadali Shafieian[a,*], Mojdeh Rahmanian[b]

[a]*School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran*
[b]*Department of Computer Engineering, Jahrom University, Jahrom, Iran*

Abstract.- In this article is proposed a method for improving speech enhancement techniques that use wavelet packet transform by applying adaptive thresholds on wavelet packet coefficients and using voice activity detection as well as applying spectral subtraction technique. The adaptive thresholds are determined according to the level of noise in the noisy speech signal. Furthermore, principal component analysis method is used as a powerful statistical method and linear transform technique in analyzing wavelet packet coefficients. An advantage of the proposed methods is that unlike other algorithms based on wavelet packet transform in which detection of unvoiced part of speech signal affects the performance of the algorithms considerably, proposed methods don't require any tool to detect voice or unvoiced part of speech signal. The voice activity detection utilized is able to update noise statistics which is beneficial for the colored and non-stationary noises. The proposed methods were evaluated for speech signals containing 30 sentences in NOIZEUS database for 5 different noise types. Simulation results show that using wavelet packet transform combined with adaptive thresholding in our proposed methods outperform similar methods and can significantly enhance the quality of noisy speech for different types of noises. Eventually, evaluation of performance criteria such as SDR, SAR, SIR and SegSNR confirm the ability of the method for speech enhancement.

Keywords: wavelet packet analysis; spectral subtraction; adaptive threshold; speech denoising; voice activity detection.

Una estrategia no supervisada para mejorar el habla utilizando una transformada del paquete wavelet y umbrales adaptativos

Resumen.- En este artículo se proponen métodos para el procesamiento del habla que usan una transformada del paquete wavelet aplicando umbrales adaptativos a sus coeficientes, así como la técnica de sustracción espectral usada para la detección de actividad por voz. Los umbrales adaptativos se determinan de acuerdo con el nivel de ruido en la señal del habla. Además, los principales métodos de análisis de componentes son utilizados por su poder estadístico así como también la técnica de transformación lineal en el análisis de los coeficientes del paquete wavelet. Una ventaja de los métodos propuestos es que a diferencia de otros algoritmos basados en la transformación del paquete wavelet, no requieren ninguna herramienta para detectar la voz o parte no sonora de la señal. La detección de voz utilizada es capaz de actualizar las estadísticas de ruido, lo cual es beneficioso para el ruido de color y no estacionario. Los métodos propuestos fueron evaluados para señales de voz que contienen 30 oraciones en la base de datos NOIZEUS para 5 tipos de ruidos diferentes. Los resultados de la simulación muestran que el uso de la transformación de paquetes wavelet combinado con el umbral adaptativo en los métodos propuestos superan a otros similares y pueden mejorar significativamente la calidad del habla para diferentes tipos de ruidos. Eventualmente, la evaluación de los criterios de desempeño como SDR, SAR, SIR y SegSNR confirman la capacidad del método para mejora del habla.

Palabras clave: análisis del paquete wavelet; sustracción espectral; umbrales adaptativos; supresión de ruido de voz;detección de actividad de voz.

Received: August 28, 2019.
Accepted: October 21, 2019.

1. Introduction

Speech enhancement focuses on improving the quality of speech by utilizing different algorithms. At first, it may seem simple however, its clarity, intelligibility and compatibility with other speech processing algorithms are important issues.

* Correspondence author:
e-mail: shafieian1381@gmail.com (M. Shafieian)

However, note that criteria such as intelligibility and pleasantness are qualitative and can not be measured mathematically. As the background noise is suppressed, the crucial issue is that the speech signal should not harm or garble. Another important issue to be noted is that some algorithms add unnatural twisted noise to speech signal which sounds more uncomfortable than quiet natural background noise.

However, if the goal of speech enhancement algorithm is that the speech signal is driven for example to a speech recognizer not to be listened by humans, so the comfortless of the speech signal is not an important issue. Thus, keeping the background noise in a low level is crucial. The speech signal for which the background noise is suppressed may be used in many applications, i.e. one apparent application is using telephone in a noisy environment such as streets, car and factories or sending speech from the cockpit of an aircraft to the ground or to the cabin. Moreover, enhancing speech for coding and recognition purposes is also a good idea. Furthermore, enhanced speech is able to be compressed in fewer number of bits than non-enhanced speech [1, 2, 3].

Typically, the algorithms for speech denoising or speech enhancement can be classified in four general classes which are spectral subtraction algorithms, wavelet transform, subspace algorithms and the algorithms based on statistical model of the speech signal [4].

In order to remove noise from speech signal, the frequency spectrum of the signal can be modified so that background noise will be removed from speech signal or suppressed after recovering the signal in time domain. To do so, first speech signal should be transformed to frequency domain by using a transform such as Fourier transform, then required modifications should be applied on the frequency spectrum of the signal in such a way that noise will be suppressed. One of well-known approaches for this purpose is Spectral subtraction of power spectrum [5] as well as applying Wiener filter on the spectrum.

While these methods had been successful and their implementation are very simple, they have some challenges too. One of deficiencies of these approaches is the distortion caused on the desired signal by using them. Different algorithms have been proposed to overcome this phenomenon which consist of perceptually motivated techniques [6] and human auditory system aspects [7]; however, it is not clear how much they are optimal in the concept of linear estimation [4]. Another disadvantage of them is creation a type of artificial background noise which has been known as music noise. Furthermore, due to nature of Fourier transform and nonstationary nature of speech signal, spectral subtraction and filtering should be applied on the windowed spectrum with finite length, so in general a window with small size leads to limitation in resolution of frequency spectrum of speech signal, thus the performance of spectrum modification or filtering will be degraded [8].

The algorithms that are based on statistical models are used most common for speech enhancement [4]. To recover coefficients of transform in clean speech or their magnitudes, they are modeled by a problem known as Bayesian estimation in which statistics of speech and noise are known. Thus, under different assumptions for distributions of speech signal and noise, many estimators can be derived. Weiner filter algorithms are demonstrated as optimal filters in estimating noisy spectrum of the signal by minimizing Mean Square Error (MSE) [4, 9]. Minimum Mean Square Error (MMSE) estimator is utilized for evaluating the magnitude of short-time spectral amplitude (STSA) according to a priori signal-to-noise ratio (SNR) estimation and Gaussian statistics [10].

Other methods for speech enhancement are using wavelet transform. The most advantage of wavelet transform is utilizing time windows with variable lengths for various frequency bands. Thus, using wavelet transform allows us to achieve high frequency resolution in lower frequency bands while maintaining accuracy in time resolutions. Thus, wavelet transform has not the limitation of using window with finite length in Fourier transform.

The most useful method for speech enhancement by using wavelet transform is applying threshold on wavelet transform coefficients. This approach acts on this fact that in speech signals, like many other

signals, the concentration of energy is on a few numbers of wavelet coefficients. These coefficients are greater than other wavelet coefficients for the signal itself or wavelet coefficients of any signal especially noise in which the energy is spread on many numbers of coefficients. So, by setting smaller coefficients equal to zero, it is possible to limit the noise as well as preserving vital information in the original signal.

According to this feature, wavelet coefficients of the signal are compared with a threshold and the coefficients which are smaller than the threshold set to zero. This process can be considered as applying filter on wavelet coefficients too. In addition to directly removing noise wavelet coefficients, thresholding methods have been utilized for estimating frequency spectrum of speech signal too [11, 12, 13]. The procedures based on adaptive wavelet thresholding are proposed in [13] and considered as universal threshold. A strategy known as Stein's Unbiased Risk Estimate (SURE) is explained in [14], Bayes Shrink in which a Bayesian estimate is exploited detailed in [15]. The main disadvantage of the wavelet transform is the limited number of frequency bands. Furthermore, it has proven that the unvoiced frames in noisy speech may be a challenging issue in the sense of wavelet shrinking.

In [16], a new wavelet thresholding method for speech enhancement is introduced in which adaptive thresholding on wavelet coefficients and modified thresholding functions are proposed for improvement the performance of speech enhancement as well as the accuracy of automatic speech recognition. In the proposed speech enhancement system, a new voice activity detector (VAD) was represented to update noise statistics in the situations of facing to colored and non-stationary noises. In our propose methods we have utilized the solution proposed in [16].

The algorithms based on speech subspace project the noisy speech segments of the speech signal onto orthogonal subspaces. The speech subspace is composed of the vectors with high-energy in the basis of segment's principal component (PC). The first algorithm based on speech subspace was presented in [17] where the authors used the Singular Value Decomposition (SVD) technique for eliminating the noise subspace in order to achieve speech denoising. Thus, in Ephraim [18] have used the fast Fourier transform (FFT) for approximating PC basis. In these methods the "musical noise" artifacts have removed, on the other hand, the subspace approach improves perceived speech quality but does not increase speech clarity.

A famous method represented in [19] digitalize the noise and clean speech covariance matrices jointly which leads to the optimal estimators. Unfortunately, a challenging task for the approaches based on speech is their efficient implementation with an optimal choice of parameters especially when facing colored and babble noise. To deal with this constraint, many solutions based on the speech separation have been presented such as K-SVD, for enhancing the perceptual clarity of the speech signal which is degraded by additive background noise [20] and nonnegative matrix factorization (NMF) [21, 22].

However, these solutions always have requirements such as prior training in supervised separation, experimental parameters, or special features. Consequently, some researchers have been studying on using principal component analysis (PCA) in which the goal is to find a set of orthogonal factors to describe the observations variance and track the new factors to determine the essential features without need to prior training.

Utilizing *PCA* for speech enhancement has a widespread use as a classical multivariate tool for speech processing [4, 23]. For speech separation, classical PCA is extended robustly by generalizing an eigenvalue decomposition for a pair of covariance matrices which is introduced in [24]. This can be used for speech denoising [25], speech identification [26], and speech recognition [27].

In [4], a speech enhancement method for a single-channel speech is proposed which combines the wavelet packet transform and an improved version of the principal component analysis (PCA), so that the capability of PCA for de-correlating the coefficients in which a linear relationship is extracted, accompanied with deriving feature vectors from wavelet packet analysis for speech

enhancement.

Consequently, the enhanced speech achieved by the inverse wavelet packet transform is decomposed into three subspaces. Our proposed methods are very close to the method introduced in [4], but the contribution of our methods is applying adaptive thresholding on wavelet packet coefficients and using voice activity detection as well as applying spectral subtraction technique which has not been utilized simultaneously by other researchers. Simulation results show that our proposed methods outperform the method introduced in [4].

In this paper, is introduced a method for improving speech enhancement methods that use wavelet packet transform by applying adaptive thresholds on the coefficients.

The remainder of this paper is organized as follows: in section 2 wavelet packet transform have been discussed. Different thresholds that used in our proposed methods such as soft and hard thresholds, adaptive threshold and Minimum Description Length (MDL) thresholding are investigated in section 3; in section 4 principal component analysis (PCA) technique has been discussed. It is followed by explaining Improved PCA (IPCA) by details in section 5 which is proposed in [4]; in section 6, Spectral Subtraction (SS) technique is introduced. Proposed methods are presented in section 7 and simulation results as well as comparison between proposed methods and other methods are shown in section 8. Finally, it is followed by conclusion in section 9.

2. Wavelet packet transform

In Fourier transform a fixed resolution is used so that we can not have an accurate time-frequency representation of the signal, i.e. it is not possible to determine which frequencies are presented in each time samples, however it is possible to determine which frequency band has been exist in each time interval. This is directly related to concept of resolution. Although the challenges for resolution of time and frequency is a result of a physical phenomenon and does not depend on type of transform, but an alternative approach for signals analysis can be utilized which is called multi-resolution analysis. The concept of multi-resolution analysis is the basis of wavelet transform [28].

Wavelet Packet Transform (WPT) is a generalization of decomposing process that provides various facilities for signal analysis. In wavelet Transform (WT) analysis, the signal is decomposed to approximation and detail subbands. This approximation subband then is decomposed to second level of approximation and detail subbands and this process continues [16]. However, in wavelet packet analysis, detail subbands may be decomposed too. Figure 1 illustrates a four-level decomposed tree for wavelet packet transform and Figure 2 represents spectral features for three-level decomposition tree.

Wavelet packets are constructed from wavelet decomposition tree which is able to represent desired frequency resolution. Mathematical basis for wavelet packets was first established by Coifman and Wickerhauser in 1992. The main advantage of wavelet packet transform is its comprehensiveness in matching of the transform to a signal regardless of its statistical properties i.e. wavelet packet tree provides much more basis than wavelet transform so that it can provide more proper representation for the signal.

The sub-tree with best basis for representation of the signal is called wavelet packet tree with optimal basis. Criteria and methods for selection of tree with optimal basis is investigated in [29].

Thus, in methods based on wavelet packet transform for denoising, much more steps are required than denoising methods based on discrete wavelet transform, i.e. more steps means more computations in achieving the tree with optimal basis. After construction of best tree, thresholding should be applied to coefficients so that reconstruction of the signal should be done based on these modified coefficients in each node of the tree [30].

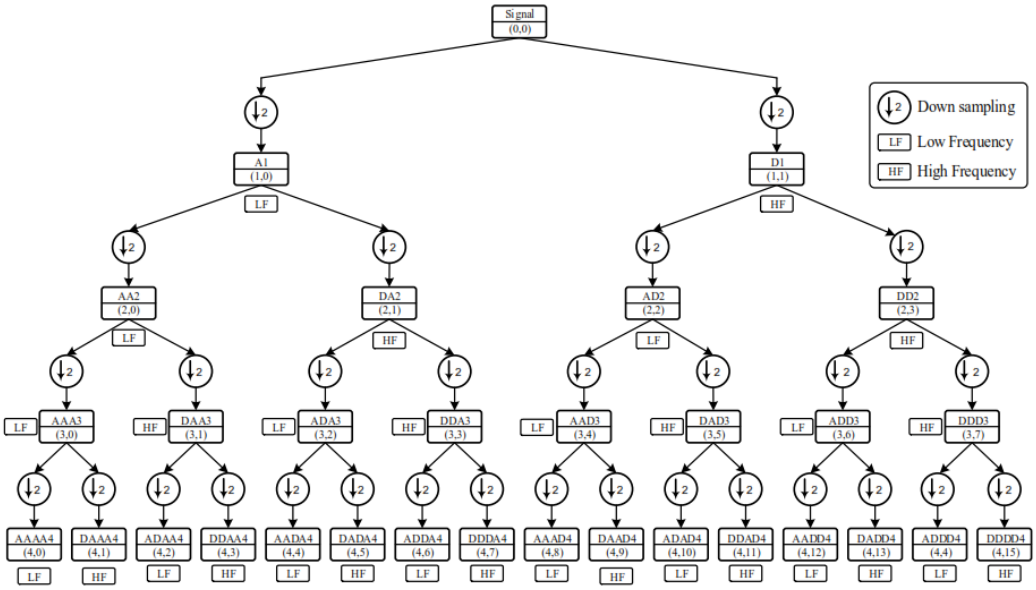


Figure 1: A four-level decomposition tree for wavelet packet transform.

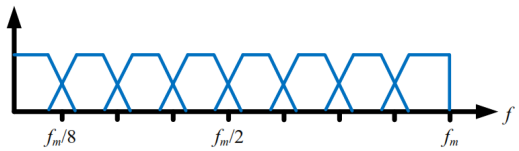


Figure 2: Spectral features for three-level decomposition tree.

3. Thresholding

3.1. *Denoising by hard and soft thresholding on wavelet packet coefficients*

Denoising by applying threshold on wavelet coefficients is according to this fact that in most signals such as speech, concentration of energy occurs mostly in a small number of wavelet dimensions [31]. These coefficients are relatively large compared to other dimensions or to any other signal specially noise in which the energy has a wide spectrum over a large number of coefficients. Thus, by setting smaller coefficients equal to zero, denoising can be achieved nearly optimally while important information of the original signal will be preserved [31]. The main algorithm for denoising can be summarized as follows:

Suppose that s is a clean speech signal with finite length and y is a noisy speech signal which is corrupted by a white Gaussian noise with variance σ^2, represented by equation (1):

$$y = s + noise \tag{1}$$

If W is assumed to be wavelet transform matrix, so in wavelet domain, equation (1) can be written as equation (2):

$$Y = S + NOISE \tag{2}$$

in which, $Y = W{\cdot}y$, $S = W{\cdot}s$ and $NOISE = W{\cdot}noise$. Thus, wavelet coefficients for estimated speech signal $\widehat{S}$ can be achieved from wavelet coefficients of noisy speech signal, according to equation (3):

$$\widehat{S} = \mathrm{THR}\,(Y, T) \tag{3}$$

in which, $THR\,(\cdot,\cdot)$ denotes thresholding function and T is assumed a scalar value as the threshold. Standard thresholding functions which are used in the methods based on wavelet are soft and hard thresholding functions which are determined using equation (4) and (5), respectively.

$$THR_H\,(Y,T) = \begin{cases} Y & |Y| \geq T \\ 0 & |Y| < T \end{cases} \tag{4}$$

$$THR_S(Y,T) = \begin{cases} \text{sign}(Y)(|Y| - T) & |Y| \geq T \\ 0 & |Y| < T \end{cases} \quad (5)$$

Furthermore, some other functions for thresholding are introduced in [31, 32]. The appropriate value for threshold can be defined by many methods. A threshold called universal threshold is proposed by Donoho in [12] for Fast Wavelet Transform (FWT) which can be defined by equation (6):

$$T = \widehat{\sigma}\sqrt{2\ln(N)}, \quad (6)$$

and for wavelet packet transform, the value for the threshold can be determined by equation (7):

$$T = \widehat{\sigma}\sqrt{2\ln(N\log_2(N))}, \quad (7)$$

in which, N is assumed as the length of noisy signal, Y, and $\widehat{\sigma}$ is standard deviation for additive white Gaussian noise with zero mean that is estimated by Donoho and Johnston according equation (8) [33]:

$$\widehat{\sigma} = \frac{MAD}{0{,}6745} = \frac{Median(|c|)}{0{,}6745}, \quad (8)$$

in which, c is the sequence of coefficients for wavelet transform of the noise and MAD is the median absolute deviation. When facing with correlated noise, a level dependent threshold was represented by Johnston and Silverman [34] using equation (9):

$$T_j = \widehat{\sigma}_j\sqrt{2\ln(N_j)}, \quad (9)$$

in which N_j is the number of samples in the scale j and $\widehat{\sigma} = MAD_j/0{,}6745$ in which MAD_j denotes the median absolute deviation estimated on the scale j.

3.2. *Denoising by applying adaptive threshold on wavelet packet coefficients*

Basic thresholding on wavelet coefficients that is discussed in previous section has some defects which occurs when facing to a noisy speech signal corrupted by real-life noises. This challenge has two aspects. First, in basic method it is assumed that the noise has a white spectrum. However, in most practical systems we are faced with colored noises and white noise does not exist in the real environment. Thus, the basic wavelet shrinkage function does not lead to a good speech quality and it is not able to remove the non-stationary noises [16].

Another challenge is the shrinkage of the segments of the speech signal that do not contain speech components and considered as unvoiced speech and contain many components of the signal that are called as noise-like speech. This results in degradation of the quality in the enhanced speech. Furthermore, using a single threshold for all wavelet packet bands is not suitable and utilizing classic thresholding functions such as Hard and Soft thresholding functions often leads to time-frequency discontinuities. Therefore, in [16] some modifications is proposed to solve these problems. Block diagram of this proposed system is illustrated in Figure 3.

According to Figure 3, wavelet packet transform is applied to each input frame and the coefficients are processed for Denoising. Finally, the inverse wavelet packet transform (IWPT) is applied to achieve the enhanced speech. Basic blocks of the processing part of this diagram is described in [16] completely.

In the block diagram shown in Figure 3, a modified version of the hard thresholding function is used for thresholding function instead of standard form of thresholding function in which the wavelet coefficients lower than the threshold value set to zero which leads to time–frequency discontinuities in the spectrum of enhanced speech. Thus, a nonlinear function is applied to the threshold value. The proposed thresholding function is defined by equation (10) [16]:

$$THR(Y,T_{j,k}) = \begin{cases} Y & |Y| \geq T_{j,k} \\ sign(Y)\cdot\dfrac{|Y|^{\gamma}}{T_{j,k}^{\gamma-1}} & |Y| < T_{j,k} \end{cases} \quad (10)$$

in which, $T_{j,k}$ is adaptive node-dependent threshold value defined in [16] and Y is noisy speech signal.

Voice activity detector allows discrimination between the speech and the non-speech parts of

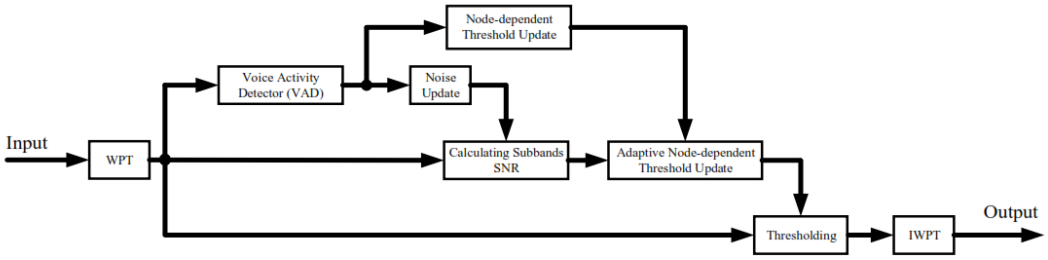


Figure 3: Block diagram for thresholding proposed in [16].

signal. So, VAD should be robust to the noisy background [16]. The proposed VAD algorithm in [16] is based on the energy of wavelet packets. For convenience, the first input frame is assumed silence frame. Note that, if this assumption does not hold, the detector will converge slower and also misclassify a few frames. The VAD algorithm is described by details in [16].

3.3. MDL Tresholding

When we face to data which result from finite number of observations, making decision for selecting among them leads to a problem called model selection. This problem is one of the most important problems in the field of statistical deduction. Minimum Description Length (MDL) method is one of deduction methods which relatively new and provides a general solution for model selection problem.

MDL method is based on the viewpoint that any regularity in data can be used for data compression so that, data can be described by fewer number of samples rather than what is needed for describing real data. Thus, more regulations in data leads to much more compression. According to what is mentioned above, by assuming the concept of learning equal to the concept of finding regularities in data, it may be concluded that as the data can be compressed more, so much more learning is accomplished from the data. MDL method defines that for hypotheses, H, and data sets D, we try to find a hypothesis or a set of hypotheses in H which is able to compress data sets D as much as possible [35].

For selecting coefficients of a transform like wavelet transform or wavelet packet transform, MDL method can also be used so that by finding regularities in coefficient's sets, compression of coefficients can be accomplished [36]. MDL criterion to select coefficients was developed by Rissanen in [37] and is utilized independently by Saito [38] and Pesquet [39] for enhancement of a speech signal in additive white noise. MDL criterion is an informatics and theoretical criterion which is used in many applications for estimating the order of parametric models [36]. Primary observations show that MDL method acts well in situation that signals are added by additive white Gaussian noise. However, limiting MDL algorithm to reduction of white noise, restricts its appropriateness in many practical applications [36].

4. Principal component analysis (PCA)

Principal Component Analysis (PCA) is a powerful statistical method and also is an advantageous linear transform technique for data analysis which can be used in various fields such as feature extraction, data compression and redundancy removal. Furthermore, it is a useful technique for pattern recognition in data. This method acts such that similarities and differences in the data can be determined. In fact, principal component analysis is a method to determine correlations between data variables. If the data have high correlation like MRI images, principal component analysis can be utilized as a powerful tool to transform data representation domain to features domain. PCA is classified as an unsupervised method and consists of eigenvalue decomposition for covariance matrix [40, 41].

In many applications in which matrix is used, matrices can be summarized to smaller matrices

with lower dimensions that are very similar to original matrix in some cases. These smaller sized matrices are called narrow matrices that have only a small number of rows or a small number of columns, and therefore can be used much more efficiently than can the original matrix with large size. The procedure of finding these narrow matrices is called dimensionality reduction [42].

In classification applications the data which contain information are considered as inputs of decision system. Ideally there should be no need for selection or feature extraction process as a separate process so the classification system should be able to use necessary data and eliminate irrelevant data. However, there are some reasons for applying dimensionality reduction as a separate process [40] such as achieving to lower usage of memory space and lower computations, more robustness for data sets with simpler models, possibility of knowledge extraction from data and providing visual display and analysis for the data with lower dimensions.

5. Improved PCA (IPCA)

In the context of Improved PCA (IPCA) which is introduced in [4], the obtained eigenspace is decomposed to three subspaces in order to achieve the most reduction of noise and to guarantee minimum distortion in the signal. In this method, FFT is applied on noisy signal and its spectrum is calculated to obtain X' matrix. The FFT of the signal $x'(t)$ can be given by equation (11):

$$X'(i,n) = \sum_{k=-\infty}^{+\infty} x'(k)\, w(i-k)\, e^{-j2\pi kn/L_f}, \quad (11)$$

where i denotes the index of the time-frame, n denotes the index of the discrete frequency, $w(i)$ is a window function for analysis, and L_f is the frequency analysis length. The spectrum magnitude $\left|X^{'}(i,n)\right|$ of speech signal should be smoothen at each frame in the frequency domain, then these frames are accumulated as column vectors to achieve the matrix representation as X'.

Consequently, proposed improved principal component analysis (IPCA) technique in [4] is applied to obtain three matrices such as Sparse (S), Remainder (Re) and Low rank (Lo) matrices from the matrix X'. Partition to these subspaces is based on the hypothesis that the noise spectrum always represents an iterative pattern and has a limited variation whereas the speech signal has more alteration and is relatively sparse within the noise. So, the new formulation for IPCA in the frequency domain can be described by the equation (12) [4]:

$$X' = S + Re + Lo, \quad (12)$$

where, X' is the input coefficients matrix, S, Re and Lo denote to sparse matrix, remainder matrix and low-rank matrix respectively. Since the sparse matrix S represents the matrix for speech structure and low-rank matrix Lo denotes the structure of noise matrix, so the goal is to recover these two matrices from the input matrix under the disturbance of the remainder matrix Re in which the distribution of inputs is zero-mean Gaussian distribution. Thus, the background noise is can be denoted as the sum of the remainder and the low-rank components. To model the noise alterations, the remainder noise matrix is applied whereas the low-rank matrix is exploited to describe the stable statistics of noise. This assumption leads to more improvement. Now, to solve the following convex optimization problem, the algorithm proposed in [43] is utilized the equation (13):

$$min \left\|X' - S - Lo\right\|_* + \alpha \left\|S\right\|_1, \quad (13)$$

where, $\|\ \ \|_*$ denotes the nuclear norm of Lo, where $\|Lo\|_*$ is defined as the sum of the singular values in Lo and indicates minimizing the rank of Lo.

The sparsity of S can be measured by the L1-norm $\|\ \ \|_1$, that represents the summation of absolute value of elements in the matrix. A trade-off between the sparsity of S, and the rank of Lo can be represented by $\alpha = 1/\sqrt{max(m1, m2)}$. Finally, by applying the Inexact Augmented Lagrange Multiplier (IALM) in [44], the enhanced matrix can be obtained by equation (14) [4]:

$$\begin{cases} X' = S + Re + Lo \\ g\left(S, Lo\right) = Re = X' - S - Lo \end{cases} \quad (14)$$

The remainder matrix Re in terms of the sparse and low components are described by the function g. As a result, the IALM function is given by equation (15) in [4]:

$$\begin{aligned} I\left(S, Lo, Re, \beta\right) = & \left\|Lo\right\|_* - \alpha \left\|S\right\|_1 \\ & + \left\langle X'.Re \right\rangle \\ & + \frac{\beta}{2} \left\|X' - S - Lo\right\|_F^2 \end{aligned} \quad (15)$$

where $I(\quad)$ is the IALM function, and β is a positive scalar. Then, the best solution of improved PCA is (S^*, Lo^*, Re^*) of $\left(S_i^*, Lo_i^*, Re_i^*\right)$ and the convergence rate is at least $O\left(\beta_i^{-1}\right)$. It is obvious that proper selection of β_i is necessary to obtain a minimum number of SVD. So, the sub-problem $\left(S_{i+1}^*, Lo_{i+1}^*, Re_{i+1}^*\right) = \underset{Lo,S,Re}{argmin} I\left(S, Lo, Re^*, \beta_i\right)$ can be solved inexactly by the IALM technique. The process is presented in the algorithm below in details [4]:

Algorithm for the IALM model.

Input: data matrix X', and the parameter α.
Initialization: $Lo = 0; S = 0; Re = random;$ $\beta > 0; \omega > 1; i = 0;$
while not converged do
Update S:
$S_{i+1} = \underset{S,Re}{argmin} I\left(S, Lo, Re^*.\beta_i\right)$
Update Lo:
$Lo_{i+1} = \underset{Lo,Re}{argmin} I\left(S_i, Lo, Re_i, \beta_i\right)$
Update Re:
$Re_{i+1} = Re_i + \beta_i \left(X' - S_{i+1} - Lo_{i+1}\right)$
$\beta_{i+1} = \omega \beta_i$
$i + 1$
end while
Output: $S_i; Lo; Re_i$

To avoid fitting the background noise with concurrent speech is introduced the remainder matrix in the subspaces.

6. Spectral subtraction

Spectral Subtraction (SS) algorithm is a primary algorithm which has been using in speech enhancement problem. This algorithm is simple but leads to distortion in the signal and causes additional noise known as musical noise which is annoying. To solve this problem, many algorithms have proposed such as cognitive techniques [6] and methods based on characteristics of human hearing system [7]. However, the optimality of these algorithms in the aspect of linear estimation is not so apparent [4].

Spectral subtraction is used to recover power or magnitude spectrum of a signal which is corrupted by additive noise by estimating mean of noise spectrum from noisy signal. Typically, estimation and updating of the noise spectrum is done when the signal is not present and only the noise is existed. The noise is assumed to be a stationary process or a process with slight variations and spectrum of noise has not significant variations between updating intervals. In order to recover signal in time domain, a combination of instantaneous magnitude spectrum estimation and the phase of noisy signal is exploited, then it is transformed to time domain by applying inverse discrete Fourier transform. In the aspect of computational complexity, spectral subtraction method is relatively efficient. However, due to stochastic variation in the noise, this method may lead to negative estimations in short–time magnitude spectrum or power spectrum. Magnitude and power spectrums have non-negative values and any negative estimations for them should be mapped to non-negative values. This nonlinear rectifying process results in distortion of recovered signal distribution. The distortion caused by processing becomes more important when signal to noise ratio decrease [45]. Spectral subtraction technique is discussed by details in [45].

7. Proposed methods

According to what is mentioned in previous sections, now we will introduce our proposed

methods for speech enhancement. In these methods, we have utilized the tools and techniques which was introduced in previous sections.

In all three proposed methods, wavelet packet transform is utilized so that the signal is decomposed to approximation and detail subbands. Then approximation subband is decomposed to the second level of approximation and detail subbands by using quadrature mirror filter. As mentioned previously, wavelet packet transform is a generalization of wavelet transform. In wavelet packet transform, filtering is applied to both high frequency and low frequency subbands, illustrates a wavelet packet decomposition tree in which each node is denoted by (E, n) where E refers to decomposition level and n refers to node label in the subband. The root of this tree i.e. $(E, n) = (0, 0)$ determines the original signal. Left hand and right-hand branches of the tree denotes highpass and lowpass filtering with sampling rate of 2:1, respectively. All speech enhancement methods proposed here are based on calculation of the spectra from adjacent frames in speech signal. Practically, the adjacent frames overlap a little and the frame length is in the order of milliseconds. The speech frames obtained by windowing are padded with zeros to make their length equal to the nearest power of two.

7.1. Proposed method 1

In the first proposed method whose block diagram is illustrated in Figure 4, the following will be done for speech enhancement:

Step 1: The noisy speech is divided into F frames with the length of $N = 512$ samples which have half-length overlap on their adjacent frames. *So*, a matrix $X = \{x_1, x_2, \cdots, x_F\}$ with dimension NF is obtained. Thus, the result of this speech framing step is a matrix whose columns are the segments noisy speech which have a partially overlapping.

Step 2: Each column in the data matrix X for noisy speech that is obtained in the previous step is decomposed to wavelet packet decomposition (WPD) coefficients matrixes $\left\{W_{E,0}, W_{E,1}, \cdots, W_{E,2^E-1}\right\}$ by selecting wavelet packet (WP) function using Daubechies wavelet with level E which is assumed to be 4 in our simulation.

Step 3: The best wavelet packet basis coefficients are applied to the WPD tree in which we have supposed Shannon entropy to construct optimal wavelet packet decomposition tree.

Step 4: The coefficients are chosen as a column vector to achieve WPD coefficients matrices for all tree nodes $\left\{Y_{E,0}, Y_{E,1}.\cdots, Y_{E,2^E-1}\right\}$, so that the column and row numbers of these matrices are F and $\mathrm{N}/2^{\mathrm{E}}$, respectively. Note that $Y_{E,n}$ means the matrix that contains all coefficients of wavelet packet transform in the node (E, n) of wavelet packet decomposition tree for all frames i.e. by the matrix $Y_{E,0}$ we mean the matrix that contains wavelet packet coefficients in the $(E, 0)$ node of decomposition tree for all frames; since length of each frame is assumed as 512 in step 1, so the dimension of the matrix would be 512×F and if the levels of wavelet packet decomposition is assumed to be 4, then the number of Y matrices will be 30.

Step 5: Adaptive thresholding using VAD is applied so that noise will be removed from wavelet packet coefficients to some extent.

Step 6: Conventional PCA is applied on the obtained coefficients matrices to calculate principal components' score matrices $\left\{O_{E,0}, O_{E,1}, \cdots, O_{E,2^E-1}\right\}$ and loading matrices $\left\{LO_{E,0}, LO_{E,1}, \cdots, LO_{E,2^E-1}\right\}$.

Then reconstruct WP coefficients matrixes $\left\{Y'_{E,0}, Y'_{E,1}.\cdots, Y'_{E,2^E-1}\right\}$ are reconstructed, the corresponding column vectors are combined to obtain $\left\{W'_{E,0}, W'_{E,1}, \cdots, W'_{E,2^E-1}\right\}$. To get the principal components, first the Eigen-decomposition of covariance matrices is computed to obtain eigenvector and eigenvalues. Second, they are arranged in a decreasing order. Finally, by using Kaiser's rule the eigenvalues are selected to determine the number of retained principal components.

Step 7: According to what is shown in [46] by Rissanen the thresholding algorithm proposed by Donoho and Johnstone in [13] results in the elimination of too many coefficients. So, MDL model selection criterion is applied to select correct values of threshold for denoising in our proposed method. More information about MDL model, is

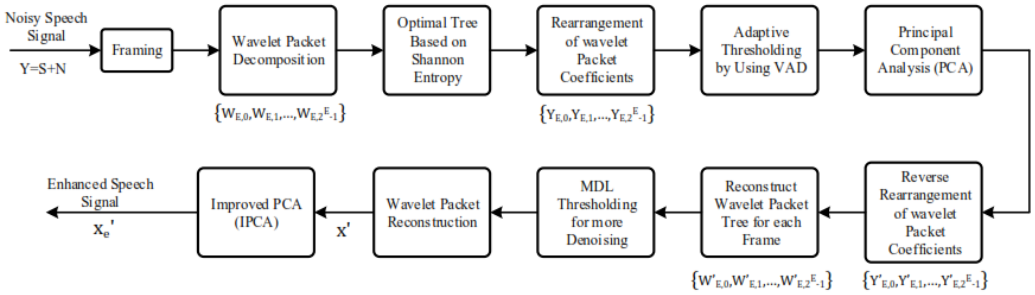


Figure 4: Block diagram of proposed method 1.

presented in [36] by details.

Step 8: By using WP Reconstruction (WPR) method adding overlaps, the matrix $B = \{x'_1, x'_2, \cdots, x'_F\}$ for the frames of enhanced speech signal are reconstructed.

Step 9: The final enhanced speech signal $X_e = \{xe'_1, xe'_2, \cdots, xe'_F\}$ is reconstructed by applying improved PCA (IPCA) proposed in [4] on matrix B that is obtained in the previous step.

Step 10: By transforming matrix X_e to a vector, enhanced speech signal will be obtained which can be displayed and heard.

7.2. Proposed method 2

The second proposed method whose block diagram is illustrated in Figure 5, is similar to the first method however, in the second method we have not used adaptive thresholding by using VAD. Furthermore, in the second method, spectral subtraction technique is utilized instead of applying Improved-PCA in the last step of the block diagram.

7.3. Proposed method 3

The third proposed method whose block diagram is illustrated in Figure 6 is like to two other proposed methods explained previously. Although, in the third method, adaptive thresholding by using VAD is added to the second method and in the last step, spectral subtraction technique is utilized.

8. Simulation results

In this section, simulation results for three proposed methods will be represented. Sampling frequency in all simulations is assumed to be 8 kHz and length of the frames is considered as 512 samples. Moreover, each frame has half overlap on its previous frame. In cases of using FFT transform, its length is assumed to be 512. Note that the simulation and evaluations are based on NOIZEUS speech database (contains 30 sentences) and TIMIT speech database. For evaluation of results, five noise types are considered such as white noise, babble noise, factory noise from the NOISEX-92 database [47] and street noise and car traffic noise from the NOIZEUS database [47]. The four noises i.e. babble, factory, car, and street noises are assumed as non-stationary processes. These simulations are performed in different signal to noise ratios and the results are shown as figures.

Furthermore, in order to compare proposed methods with other similar methods for speech enhancement, four methods are considered. These methods are Wavelet Packet – Improved PCA (WP-IPCA) [4], Robust-PCA [2], geometric approach [7] and on-line semi-supervised method which is based on Nonnegative Matrix Factorization (on-Line-Sup-NMF) method [48]. The figures that illustrated for comparison are the results of 30 simulations for 30 sentences in NOIZEUS speech database and five different types of noise, then simulation results are averaged in each SNR. Note that for comparing results and evaluating performance of each method, four criteria are used such as Signal to Artifact Ratio (SAR), Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR) and Segment SNR (SegSNR).

To measure speech enhancement performance results from our methods, the BSS EVAL toolbox developed by [49] is used. Fundamentally, each

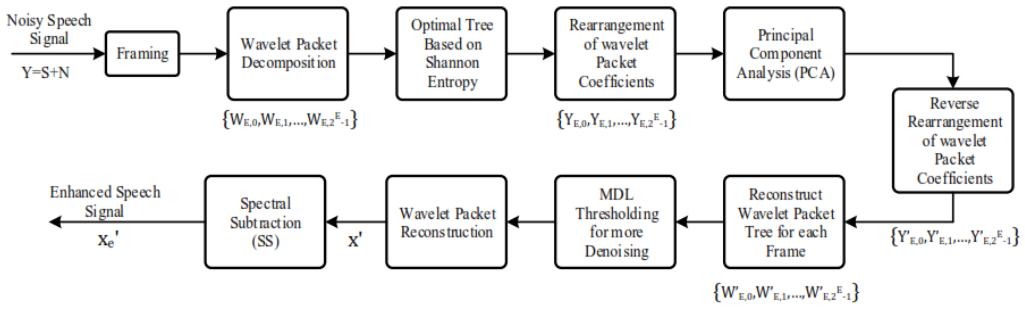


Figure 5: Block diagram of proposed method 2.

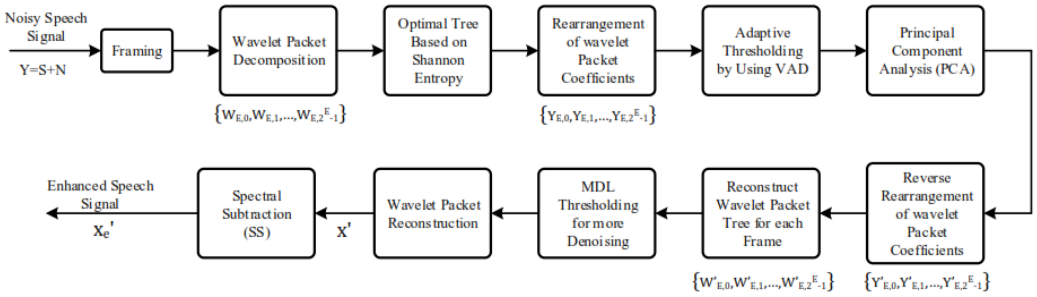


Figure 6: Block diagram of proposed method 3.

estimation of speech signal is decomposed into a true speech and error parts which correspond to interference from the noise and artifacts like musical noise. This toolbox provides three metrics: Speech-to-Distortion Ratio (SDR), Speech-to-Interference Ratio (SIR) and Speech-to-Artifacts Ratio (SAR) [49].

8.1. Simulation results for proposed methods

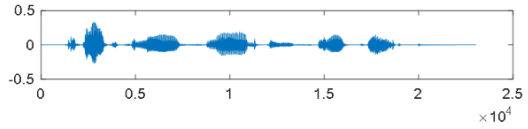
Simulation results for the three proposed methods which is shown in Figure 4, 5 and Figure 6 for one speech signal is shown in Figure 7 and Figure 8.

In these figures, original signal, noisy signal and enhanced speech signal by three proposed methods are shown in both time domain and time–frequency domain as spectrum. These simulations are implemented at $SNR = 5$ dB. Simulation results illustrated in these figures are for "SP01.wav" speech signal from NIOZEUS speech database.

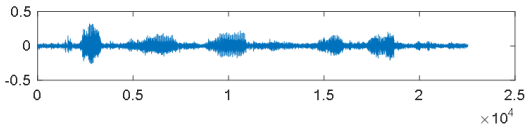
Furthermore, in order to compare our proposed method with other similar methods for speech enhancement, four methods which mentioned at the beginning of this section are considered. The figures illustrated for four performance criteria and are the results of 30 simulations for 30 sentences in NOIZEUS speech database and for five different types of noise, then simulation results are averaged in each SNR.

SDR criterion demonstrates total quality of the speech enhancement method [4]. According to the results for SDRs shown in Figure 9, our first proposed method achieved the best performance among other speech enhancement methods especially in low SNRs. For high input SNRs, the difference between our first proposed method and Robust-PCA is very close. This can be explained by this fact that disturbance of remainder matrix is small at the SNR greater than zero, and the distortion in speech is reduced by introducing new limitations on the sparse matrix. However, our first proposed method outperforms WP-IPCA because we have utilized adaptive thresholding by using VAD and this leads to more noise reduction and better speech enhancement.

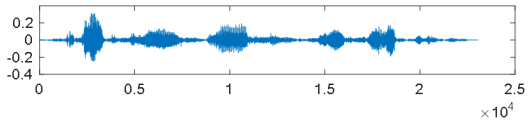
SIR criteria shows noise reduction rate [4]. Figure 10 illustrates that although on-Line-Sup-NMF achieves the best values for SIR between



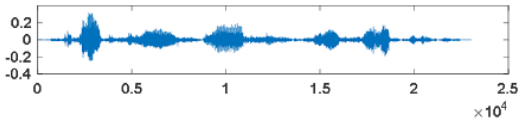
(a) Original signal



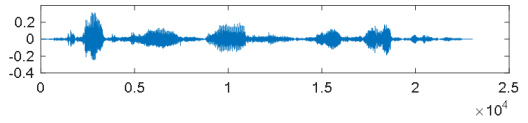
(b) Noisy signal for white noise at SNR=5 dB



(c) Enhanced speech signal by using the first proposed method

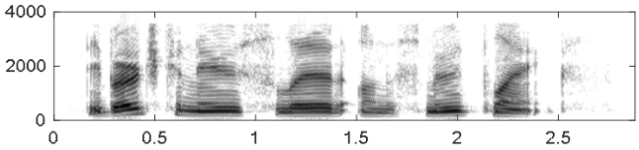


(d) Enhanced speech signal by using the second proposed method

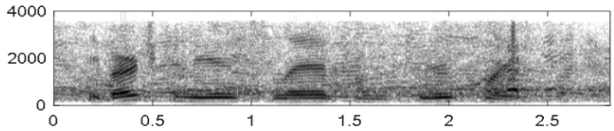


(e) Enhanced speech signal by using the third proposed method

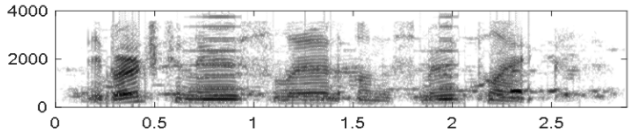
Figure 7: Simulation results for the proposed method in time domain.



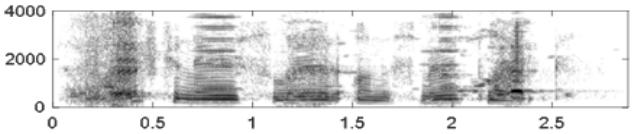
(a) Original signal



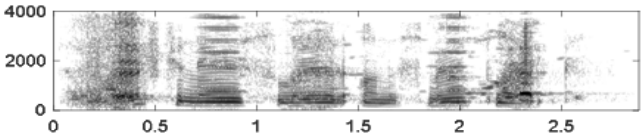
(b) Noisy signal for white noise at SNR=5 dB.



(c) Enhanced speech signal by using the first proposed method



(d)Enhanced speech signal by using the second proposed method



(e)Enhanced speech signal by using the third proposed method

Figure 8: Simulation results for the proposed method in frequency domain.

other methods, but our third proposed method achieves better performance especially at high input SNRs. The reason is that on-Line-Sup-NMF algorithm is obtained in a semi-supervised procedure, but our proposed method for speech enhancement is an unsupervised method. In on-Line-Sup-NMF method, two representations for the noise are learned, so it can exploit more prior information than our method in the speech enhancement process.

Moreover, it can be seen that our proposed method outperforms Robust-PCA and WP-IPCA methods because noisy speech signal is decomposed into two simple subspaces when the sparse matrix includes some part of the low matrix. By contrast, our first proposed method decomposes the eigenspace as the sum of three subspaces to model the variation of noise. In addition, using adaptive thresholding by VAD results in better performance than WP-IPCA.

The SAR metric determines the artifacts introduced by the speech enhancement method [4]. As it can be seen from Figure 11, our proposed method outperforms the others. However, the difference between our first proposed method, WP-IPCA and the Robust-PCA is small for high SNR levels. The On-Line-Sup-NMF method results in the worst performance among several methods. It is because of requiring prior training of noises for this method. According to the three recent figures, it can be concluded that our third method has outperformed all the other approaches, which is close to Robust-PCA at high input SNRs in some performance criteria. This shows the efficiency of the adaptive thresholding accompanied with PCA technique.

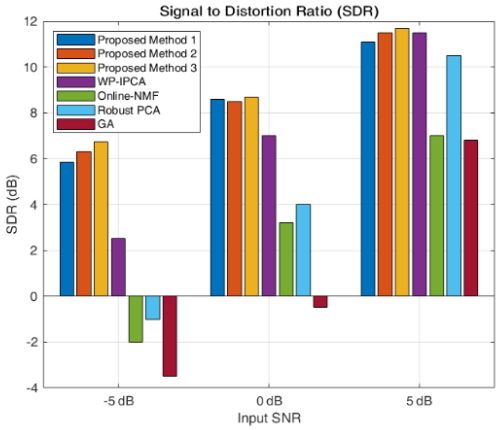


Figure 9: Average of SDRs values for several speech enhancement methods and comparison with our three proposed methods.

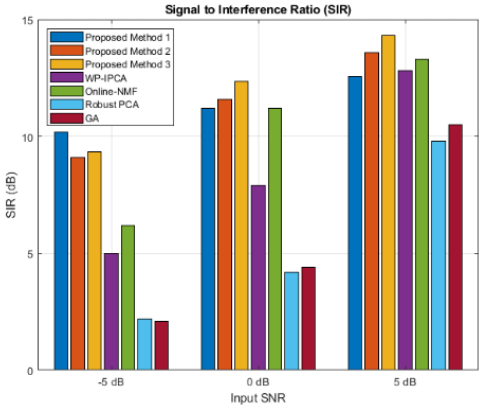


Figure 10: Average of SIRs values for several speech enhancement methods and comparison with our three proposed methods.

8.2. *Comparison between three proposed methods and discussions*

In this section, simulation results for proposed methods for performance criteria are shown in figures for 4 parameters in order to compare results. As mentioned previously, these simulations are done for speech signals containing 30 sentences in NOIZEUS database for 5 different noise types, then the averaged results computed in each SNR. Moreover, simulations done for -5 dB, 0 dB, 5 dB, 10 dB and 15 dB SNRs.

According to Figure 12, 13 and Figure 14, it can be seen that the third proposed method achieves the best performance compared with two other proposed methods in all performance criteria. This

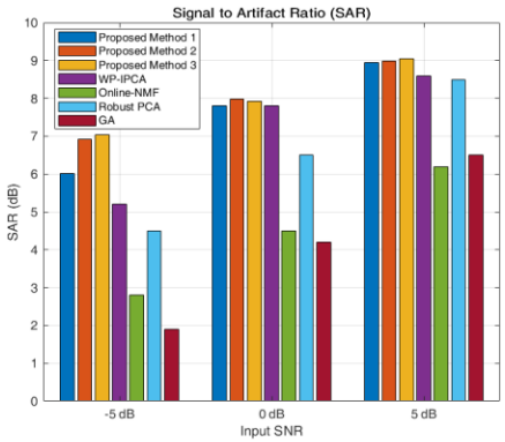


Figure 11: Average of SARs values for several speech enhancement methods and comparison with our three proposed methods.

is because of the fact that both spectral subtraction and adaptive thresholding are used in the third method so that their features in reduction of noise and speech enhancement have been utilized together.

A practical measure for quality of speech is Segmental SNR (SegSNR). It is constructed by averaging the estimation of frame level SNR using equation (16) [4]:

$$\text{SegSNR} = \frac{10}{F}\sum_{k=1}^{F}\log_{10}\left[\frac{\sum_{i=0}^{N-1}x^2(k,i)}{\sum_{i=0}^{N-1}\left[x(k,i)-xe'(k,i)\right]^2}\right], \quad (16)$$

where, N is the length of each frame, and F is the number of frames. Furthermore, xe' is the kth frame for the enhanced speech signal and x is the kth frame for original speech signal.

The results of the SegSNR values are shown in Table 1 for four speech enhancement algorithms and our three proposed methods. Table 1 shows that three proposed methods in different values of SNRs. In addition, Table 1 shows that our third proposed method achieves the best performance, although on-line-Sup-NMF gives a good performance at SNRs equal to −5 and 0 dB in contrast with WP-IPCA and Robust-PCA methods. This table confirms the results obtained for the SIR improvements criteria.

The evaluation achieved by using BSS toolbox, and speech quality measures showed that our noise

Table 1: SegSNR values (dB) of similar speech enhancement methods and our proposed methods in different SNRs.

Noise type	Method	SNR (dB) -5	0	5
White	Proposed Method 1	-0,92	1,12	2,40
	Proposed Method 2	-0,60	2,41	2,65
	Proposed Method 3	**-0,05**	**2,66**	**2,73**
	WP-IPCA	-0,93	1,98	2,34
	Online-NMF	-0,68	2,40	2,68
	Robust-PCA	-1,18	1,86	2,32
	GA	-1,29	-0,95	0,26
Babble	Proposed Method 1	-2,65	-1,15	0,10
	Proposed Method 2	-2,06	-1,00	0,36
	Proposed Method 3	**-2,03**	**-0,97**	**0,39**
	WP-IPCA	-2,84	-1,43	-0,21
	Online-NMF	-2,05	-1,01	0,37
	Robust-PCA	-3,47	-1,61	-0,29
	GA	-4,08	-3,11	-1,23
Car	Proposed Method 1	-2,54	0,53	1,63
	Proposed Method 2	-1,91	0,88	1,71
	Proposed Method 3	**-1,83**	**0,96**	**1,84**
	WP-IPCA	-2,80	0,42	1,49
	Online-NMF	-1,92	0,87	1,68
	Robust-PCA	-3,19	-0,33	1,45
	GA	-3,91	-1,43	0,01
Factory	Proposed Method 1	-1,01	0,98	1,99
	Proposed Method 2	-0,64	1,66	2,13
	Proposed Method 3	**-0,60**	**1,72**	**2,18**
	WP-IPCA	-1,05	0,98	1,94
	Online-NMF	-0,63	1,64	2,07
	Robust-PCA	-1,28	0,93	1,87
	GA	-3,28	-1,89	-0,39
Street	Proposed Method 1	-1,98	0,01	1,86
	Proposed Method 2	-1,41	0,11	1,91
	Proposed Method 3	**-1,45**	**0,21**	**2,01**
	WP-IPCA	-2,16	-0,24	1,82
	Online-NMF	-1,43	0,10	1,89
	Robust-PCA	-2,58	-0,36	1,78
	GA	-3,68	-1,42	-1,37

The best performances of segSNR are denoted by bold values.

enhancement methods using adaptive threshold and spectral features of speech signal accompanied with wavelet packet transform, outperforms in contrast with four methods. So, our methods represent the least distortion in the enhanced speech. Note that, although on-line-Sup-NMF method degrades the interfering noise considerably, but adds artifacts to the enhanced speech signal.

The best performances of segSNR are denoted by bold values.

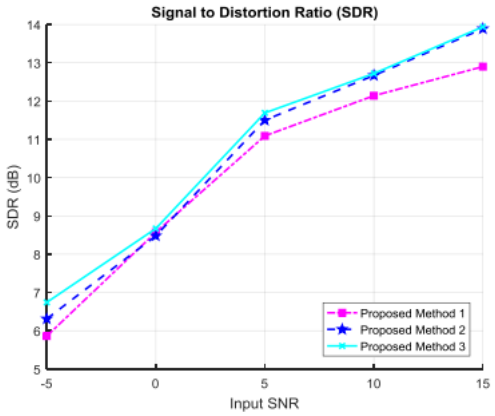


Figure 12: Averaged SDR simulation results for three proposed methods.

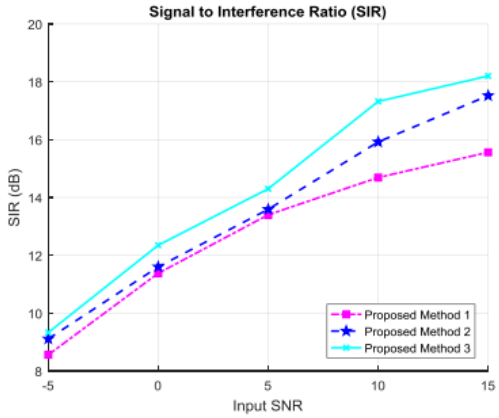


Figure 13: Averaged SIR simulation results for three proposed methods.

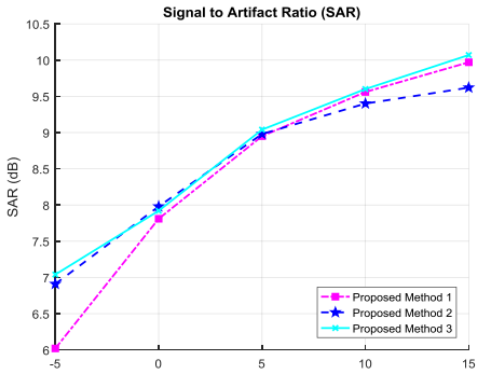


Figure 14: Averaged SAR simulation results for three proposed methods.

9. Conclusion

In this paper, new methods for speech signal decomposition for speech enhancement were

investigated and compared. To do so, new methods for speech denoising based on principal component analysis (PCA) which are along with features of wavelet packet transform and applying adaptive threshold on wavelet packet coefficients are proposed. Furthermore, using an improved version of conventional spectral subtraction method leads to better results in the performance of proposed methods. An advantage of proposed methods for noisy speech enhancement is that there is no need to prerequisite learning or experimental parameters. The main idea of our proposed methods are design and development of filters and powerful thresholding in wavelet packet transform domain. Thus, proposed methods don't require the energy of speech to be compacted into a few principal components however, the distribution of the noise is over all the transformed coefficients which allows a convenient shrinkage function to be applied on these new coefficients and to remove noise without speech degradation. So, a modified version of the principal component analysis called IPCA is applied to decompose the enhanced speech results from inverse WPT into three subspaces.

Simulation results show that using wavelet packet transform combined with adaptive thresholding can significantly enhance the quality of noisy speech for different types of noises. In this method, using signal to noise ratios in next subbands of wavelet packet transform allows us to control thresholds that are applied on wavelet packet coefficients so that more noise components will be removed from the subbands that are more affected by noise. One of advantages of our proposed method is that unlike other algorithms based on wavelet packet transform in which detection of unvoiced part of speech signal affects the performance of the algorithms considerably, our proposed methods don't require any tool to detect voice or unvoiced part of speech signal. Note that although performance of our methods is very close to spectral subtraction technique, but simulation results show that they outperform spectral subtraction method in some performance criteria.

Applying adaptive thresholding algorithm on wavelet packet transform coefficients show that significant enhancement can be achieved for noisy signals combined with white and colored noises. The evaluation of performance criteria such as SDR, SAR, SIR and SegSNR confirm the ability of the method for speech enhancement. Although, the performance of some similar methods is close to our proposed methods, but in general our methods especially the third method outperform other similar methods.

10. Bibliography

[1] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, 27(2):113–120, 1979.

[2] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust Principal Component Analysis? *Journal ACM*, 58(3):11:1–11:37, 2011.

[3] M. Dendrinos, S. Bakamidis, and G. Carayannis. Speech enhancement from noise: A regenerative approach. *Speech Communication*, 10(1):45 – 57, 1991.

[4] M.A. Ben-Messaoud, A. Bouzid, and N. Ellouze. Speech enhancement based on wavelet packet of an improved principal component analysis. *Computer Speech & Language*, 35:58 – 72, 2016.

[5] M.S. Khan, S.M. Naqvi, and J. Chambers. A new cascaded spectral subtraction approach for binaural speech dereverberation and its application in source separation. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6566–6570, Canada, 2013. IEEE.

[6] A.A. Petrovsky, M. Parfieniuk, and A. Borowicz. Warped DFT based perceptual noise reduction system. In *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004.

[7] Y. Lu and P.C. Loizou. A geometric approach to spectral subtraction. *Speech communication*, 50(6):453–466, 2008.

[8] S. Vihari, A. Sreenivasa, P. Soni, and D. Naik. Comparison of Speech Enhancement Algorithms. *Procedia Computer Science*, 89:666 – 676, 2016. Twelfth International Conference on Communication Networks, ICCN 2016, August 19– 21, 2016, Bangalore, India Twelfth International Conference on Data Mining and Warehousing, ICDMW 2016, August 19-21, 2016, Bangalore, India Twelfth International Conference on Image and Signal Processing, ICISP 2016, August 19-21, 2016, Bangalore, India.

[9] V. Sunnydayal and T. Kishore-Kumar. Speech enhancement using sub-band wiener filter with pitch synchronous analysis. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 20–25, Aug 2013.

[10] F. Ykhlef and H. Ykhlef. A smoothed Minimum Mean-Square Error Log-Spectral Amplitude estimator for

speech enhancement. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 246–249, April 2014.

[11] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337, 1995.

[12] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[13] D.L. Donoho and I.M. Johnstone. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995.

[14] Y. Hu and P.C. Loizou. Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE Transactions on Speech and Audio Processing*, 12(1):59–67, Jan 2004.

[15] D. Leporini and J.C. Pesquet. Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal Processing*, 81(1):55 – 67, 2001. Special section on Markov Chain Monte Carlo (MCMC) Methods for Signal Processing.

[16] Y. Ghanbari and M.R. Karami-Mollaei. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Communication*, 48(8):927 – 940, 2006.

[17] R. Bendoumia and M. Djendi. Speech enhancement using backward adaptive filtering algorithm: Variable step-sizes approaches. In *2015 3rd International Conference on Control, Engineering Information Technology (CEIT)*, pages 1–5, Algeria, May 2015.

[18] Y. Ephraim, H. Van-Trees, and S. Soli. Enhancement of noisy speech for the hearing impaired using the signal subspace approach. In *NIH/VA Forum on Hearing Aid Research and Develop-ment*, Bethesda, MD, 1995.

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

[20] C.D. Sigg, T. Dikk, and J.M. Buhmann. Speech enhancement with sparse coding in learned dictionaries. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4758–4761, March 2010.

[21] N. Mohammadiha, T. Gerkmann, and A. Leijon. A new linear MMSE filter for single channel speech enhancement based on Nonnegative Matrix Factorization. In *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 45–48, Oct 2011.

[22] G.J. Mysore and P. Smaragdis. A non-negative approach to semi-supervised separation of speech from noise with the use of temporal dynamics. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 17–20, May 2011.

[23] I.T. Joliffe and B. Morgan. Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research*, 1(1):69–95, 1992. PMID: 1341653.

[24] Y. Benabderrahmane, S.A. Selouani, and D. O'Shaughnessy. Blind speech separation for convolutive mixtures using an oriented principal components analysis method. In *2010 18th European Signal Processing Conference*, pages 1553–1557, Aug 2010.

[25] V.D. Shinde, C.G. Patil, and S.D. Ruikar. Wavelet Based Multi-Scale Principal Component Analysis for Speech Enhancement. *International Journal of Engineering Trends and Technology*, 3(3):397–400, 2012.

[26] S. Bavkar and S. Sahare. PCA based single channel speech enhancement method for highly noisy environment. In *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1103–1107, Aug 2013.

[27] T. Takiguchi and Y. Ariki. PCA-Based Speech Enhancement for Distorted Speech Recognition. *Journal of Multimedia*, 2(5):13–18, 2007.

[28] T. Acharya and A. Ray. *Image Processing: Principles and Applications*. John Wiley & Sons, 2005.

[29] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[30] X. Zhou, C. Zhou, and B.G. Stewart. Comparisons of discrete wavelet transform, wavelet packet transform and stationary wavelet transform in denoising PD measurement data. In *Conference Record of the 2006 IEEE International Symposium on Electrical Insulation*, pages 237–240, June 2006.

[31] H. Sheikhzadeh and H. Abutalebi. An Improved Wavelet-Based Speech Enhancement System. In *7th European Conference on Speech Communication and Technology*, pages 1855–1858, Scandinavia, September 2001.

[32] S. Chang, Y. Kwon, S. Yang, and I. Kim. Speech enhancement for non-stationary noise environment by adaptive wavelet packet. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages I–561–I–564, May 2002.

[33] D. L Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 09 1994.

[34] I.M. Johnstone and B.W. Silverman. Wavelet threshold estimators for data with correlated noise. *Journal of the royal statistical society: series B (statistical methodology)*, 59(2):319–351, 1997.

[35] P.D. Grünwald, I.J. Myung, and M.A. Pitt. *Advances in minimum description length: Theory and applications.* MIT press, 2005.

[36] N.A. Whitmal, J.C. Rutledge, and J. Cohen. Reducing

correlated noise in digital hearing aids. *IEEE Engineering in Medicine and Biology Magazine*, 15(5):88–96, Sep. 1996.

[37] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[38] N. Saito. Simultaneous Noise Suppression and Signal Compression Using a Library of Orthonormal Bases and the Minimum Description Length Criterion. In Efi Foufoula-Georgiou and Praveen Kumar, editors, *Wavelets in Geophysics*, volume 4 of *Wavelet Analysis and Its Applications*, pages 299–324. Academic Press, 1994.

[39] J. Pesquet, H. Krim, H. Carfantan, and J.G. Proakis. Estimation of noisy signals using time-invariant wavelet packets. In *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pages 31–34 vol.1, Nov 1993.

[40] E. Alpaydin. *Introduction to machine learning*. MIT Press, 2009.

[41] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, February 1981.

[42] J. Leskovec, A. Rajaraman, and J.D. Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2014.

[43] T. Zhou and D. Tao. Godec: randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 33–40. Omnipress, 2011.

[44] L. Zhouchen, C. Minming, L. Wu and Y. Ma. The aumented Lagrange multiplier method for exact recovery of corrupted low–rank matrices Technical Report. In *UILU-ENG-09-2215*, UIUC, 2009.

[45] S.V. Vaseghi. *Advanced digital signal processing and noise reduction*. John Wiley & Sons, 2008.

[46] J. Rissanen. MDL denoising. *IEEE Transactions on Information Theory*, 46(7):2537–2543, Nov 2000.

[47] Y. Hu and P.C. Loizou. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, 49(7):588 – 601, 2007. Speech Enhancement.

[48] C. Joder, F. Weninger, F. Eyben, D. Virette, and B. Schuller. Real-Time Speech Separation by Semi-supervised Nonnegative Matrix Factorization. In Fabian Theis, Andrzej Cichocki, Arie Yeredor, and Michael Zibulevsky, editors, *Latent Variable Analysis and Signal Separation*, pages 322–329, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[49] E. Vincent, R. Gribonval, and C. Fevotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1462–1469, July 2006.