EFFECT OF SINGULAR VALUE DECOMPOSITION BASED PROCESSING ON SPEECH PERCEPTION

[1]Balvinder kour, [1]Randhir Singh, [2]Parveen Lehana* and [2]Padmini Rajput

[1]ECE Department, Shri Sai College of Engineering and Technology, Pathankot, Punjab
[2]Dept. of Physics and Electronics, University of Jammu, Jammu, India

*pklehanajournals@gmail.com

ABSTRACT

Speech is an important biological signal for primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. Speech processing is the most important aspect in signal processing. In this paper the theory of linear algebra called singular value decomposition (SVD) is applied to the speech signal. SVD is a technique for deriving important parameters of a signal. The parameters derived using SVD may further be reduced by perceptual evaluation of the synthesized speech using only perceptually important parameters, where the speech signal can be compressed so that the information can be transformed into compressed form without losing its quality. This technique finds wide applications in speech compression, speech recognition, and speech synthesis. The objective of this paper is to investigate the effect of SVD based feature selection of the input speech on the perception of the processed speech signal. The speech signal which is in the form of vowels \a\, \e\, \u\ were recorded from each of the six speakers (3 males and 3 females). The vowels for the six speakers were analyzed using SVD based processing and the effect of the reduction in singular values was investigated on the perception of the resynthesized vowels using reduced singular values. Investigations have shown that the number of singular values can be drastically reduced without significantly affecting the perception of the vowels.

KEYWORDS

Speech signal, Speech generation, Speech processing, Speech compression, Singular value decomposition

1. INTRODUCTION

The information technology is having an important impact on many aspects of our lives. Communication between human beings and information processing devices is also equally important. Up to now, such communication is entirely through the use of keyboards and screens, but speech is most widely used, because it is the fastest and most natural means of communication for the human beings [1]. Speech is basically a significant biomedical signal for transmission on the electrical channels. The artificial speech that can be generated in an automated way has been a dream of humankind for many centuries [2]. Modern communication techniques use digital transmission of signals. The phenomenon known as speech recognition is used for converting speech signal to a sequence of words by means of algorithms implemented as a computer program [3]. During speech production, the air flows from lungs and passes from the glottis, throat, and the vocal tract. Depending on which speech sound we articulate, the vocal tract can be excited in three possible ways which are voiced, unvoiced, and transient excitation. In voiced excitation, the air pressure makes the glottis to open and close periodically which generates a triangle shaped periodic pulse train. This fundamental frequency of the excitation lies in the range from 80 Hz to 350 Hz. Unvoiced excitation are generated when the glottis is open and the air pressure creates the

DOI : 10.5121/ijnlc.2013.2102

narrow passage in the throat or mouth. This results in a turbulence which generates a noise signal. The spectral shape of the noise is determined by the location of the narrowness. In transient excitation a closure in the throat or mouth will raise the air pressure. By suddenly opening the closure, the air pressure drops down immediately involving explosive burst [4]. The various signal models have been developed which can provide the basis for theoretical description of signal processing system used to provide desired output and also information regarding signal source without the availability of source. The acoustic models are used to capture speech feature statistics in a parametric way. The dominant technique for acoustic model is the hidden Markov model (HMM). HMM is designed to capture time varying signal's statistics and can be considered as a generalization of the Gaussian mixture model (GMM). As it is difficult to formulate a continuously time varying model, the HMM models it by a state to state transition, hence this approach can be considered as the discretization of the continuous varying case [5]. Compression of signals is based on removing the redundancy between neighbouring samples and between the adjacent cycles. Lossy compression scheme is mostly used to compress information such as speech signals. Discrete wavelets transform (DWT) lossy compression techniques are used to solve the limited bandwidth problem. The performance of the DWT for speech compression is very good compared with other techniques such as μ-law speech coder. It is very essential to represent data by as small as possible number of coefficients within an acceptable loss of visual quality in a data compression of a signal.

Compression techniques can be divided into two main categories: lossless and lossy. Compression methods can also be divided into three functional categories: Direct method, Transformation method and Parameter extraction method. In direct methods the samples of the signals are directly provided for the compression. While transformation methods include Fourier transform (FT), wavelets transform (WT), and discrete cosines transform (DCT). In parameter extraction method, pre-processor is employed to extract some features which can be used to reconstruct the signal. The compression of speech signals has many practical applications. It can be used in digital cellular technology where the same frequency bandwidth can be shared between many users. In the compression many users are allowed to the system. In the data compression technology the information can be represented with lowest number of bits i.e. minimum size. This technology is required in the field of speech which can be used to satisfy transfer requirements of huge speech signals via communication companies and Internet. With the increasing technology the demand for digital information is increasing over the past decades [6].In speech compression it is very difficult to achieve a low bit rate in digital representation of input signal with negligible loss of signal quality and its quality can be assessed by a human which is an ultimate receiver. The perceptual coding based method has also been developed which is based on the concept of masking the distortion [7].

SVD can also be used as a one of the data compression method. It has been already used for RADAR signals. In this technique SVD is used to extract the prototype pulse from a matrix of aligned pulses [8]. The objective of this paper is to investigate the effect of SVD based feature selection of the input speech on the perception of the processed speech signal. The scope of this paper is limited to only the processing and evaluation of cardinal vowels. The detail of SVD is given in the following section. The methodology of the investigations is presented in Section III. The results and conclusions are presented in Section IV.

2. SINGULAR VALUE DECOMPOSITION

Singular value decomposition (SVD) is a technique for deriving the important parameters of a given signal. In linear algebra, the singular value decomposition means the factorization of a real or complex matrix, with many useful applications in signal processing and statistics [9].Singular value decomposition has been proved to be an efficient tool for signal processing techniques such

as image coding, signal enhancement, and image filtering. SVD based speech enhancement technique has recently been proposed in the literature. This technique assumes that the noise in the speech is additive and uncorrelated with the pure speech signal. The SVD technique enhances the noisy signal by retaining a few of the singular values from the decomposition of an over-determined, over-extended data matrix. The singular values that are ignored are associated with the noisy part of the signal. The signal reconstructed from the reduced rank matrix is the enhanced speech signal [10]. The applications employing the SVD include computing the pseudo inverse, least squares fitting of data, matrix approximation, and determining the rank, range and null space of a matrix [9]. SVD process has also found application in square matrices which was further converted to rectangular matrices [11]. A few years ago, singular value decomposition was also explored for water marking. For data processing and dimension reduction of the signal Principal component analysis (PCA), a variable reduction procedure is used [12]. PCA has found numerous applications such as handwritten zip code classification and human face recognition [13]. The SVD method has also been employed on the eigenvectors corresponding to the largest singular values contained signal information, while the eigenvectors corresponding to the smallest singular values containing noise information. The enhanced signal was reconstructed using only the information associated with the largest singular values [14]. The spoken language interface relies on speech synthesis or text to speech system to provide range of capability for having machine communicate to the user. The various speech coding technique were developed for both efficient transmission and storage of speech to conserve bandwidth or bit rate maintaining voice quality [15]. The combination with column pivoting for SVD and QR decomposition column pivoting (QRcp) have been used to select the potential set of features. The idea is to select those features that can explain different dimensions showing minimal similarities among them in an orthogonal sense. This helps in the improvement of quality of selection and enhanced identification rate. SVD QRcp has been found useful for selecting the subset of data in a heart sound classification problem using artificial neural network [16]. The optimized singular vector denoising approach for speech enhancement explains about a new algorithm for speech enhancement. In this approach, the effect of noise is reduced from both singular values and singular vectors [17].

3. METHODOLOGY

For the analysis of the speech signal let $S(\omega,t)$ be any singular matrix, representing the spectrogram of the given speech signal $s(t)$. The short time Fourier transform of input speech signal may be written as

$$S(\omega,t) = U(w,t)\, E(w,t)\, V(w,t)^T$$

where $V(\omega,t)$ is $n \times k$ unitary matrix, $E(\omega)$ is an $k \times k$ rectangular diagonal matrix having singular values, and $V(\omega,t)^T$ is conjugate transpose of $V(\omega,t)$ is an $k \times n$ unitary matrix [18]. The recording of three male and three female speakers having the age between 20-25 years at the sampling frequency of 16 kHz has been taken. The recorded speech was segmented into three different vowels to investigate the effect of SVD based processing of input speech on the perception of speech signal. Spectrograms of the segmented vowels were obtained. SVD of the spectrograms was taken and the singular values less than a threshold were retained. The synthesis was carried out with the reduced singular values. The perceptual evaluation of the synthesized speech was carried out.

4. RESULTS AND CONCLUSIONS

The investigations were carried out using perceptual evaluation and spectrograms of the synthesized speech. Table I shows the minimum singular values used for the synthesis for obtaining perceptually satisfactory quality. The mean of the singular values for three vowels is also plotted as histograms in Fig. 1. It is clear from the histograms that the minimum singular values required for the satisfactory synthesis quality is a small fraction of the maximum singular values in the recorded speech.

The spectrogram of three cardinal vowels for six speakers is shown in Fig. 2 to Fig. 7. The first column shows the spectrograms of the recorded vowels and the second column shows the spectrograms of the synthesized vowels at different singular values. The spectrograms of the synthesized vowels are smooth and visually similar to the original recorded vowels. It was observed that first four or five values of the singular matrix are enough to synthesize the vowels with satisfactory quality.

Table 1 Maximum and minimum SVD values in the three vowels for the recorded and synthesized vowels, respectively.

Speakers	**\a**		**\e**		**\u**	
	Max. SVD in recorded	**Min SVD in synth.**	**Max. SVD in recorded**	**Min SVD in synth.**	**Max. SVD in recorded**	**Min SVD in synth.**
Sp1	0.6224, 1	0.0302, 7	0.7208, 1	0.0090, 12	0.7211, 1	0.0317, 4
Sp2	0.9648, 1	0.0626, 7	0.7921, 1	0.0180, 7	0.6968, 1	0.0400, 4
Sp3	0.7887, 1	0.1638, 4	1.0600, 1	0.0576, 4	1.8170, 1	0.0940, 5
Sp4	1.2191, 1	0.457, 10	1.1056,1	0.0427, 5	1.5483, 1	0.1464, 4
Sp5	1.1856, 1	0.1686, 4	1.0470, 1	0.0912, 4	1.2232, 1	0.0959, 5
Sp6	0.6249, 1	0.0488, 6	1.0031, 1	0.0379, 5	1.0135,1	0.0782, 5
Mean	0.9009166	0.1551	0.95476	0.04274	1.16998	0.08013

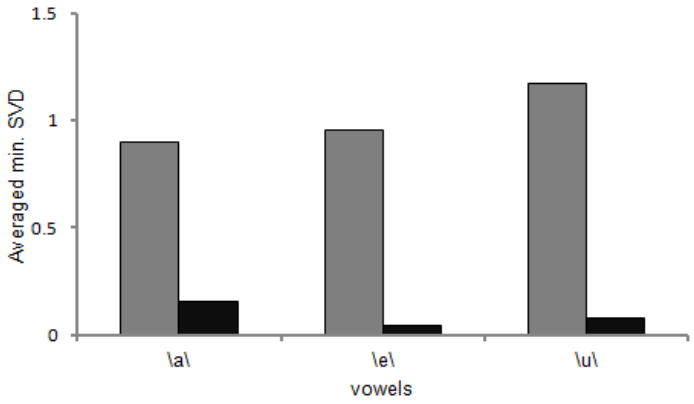

Fig. 1 Averaged maximum and minimum SVD values in the three vowels for the recorded and synthesized vowels, respectively.

Original speech signal Synthesized speech signal

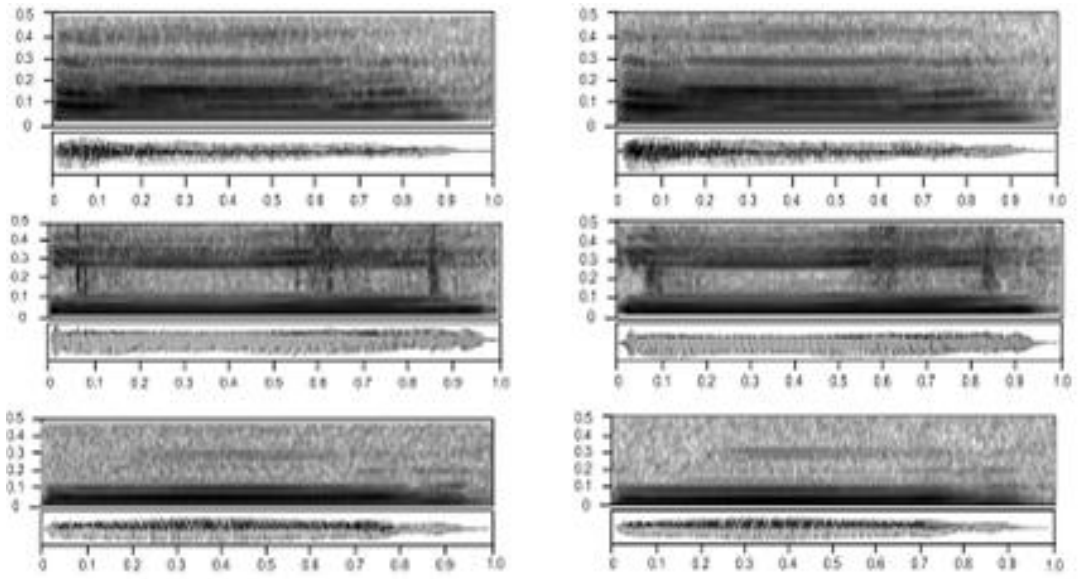

Fig. 2 Spectrogram of original and synthesized speech for vowels \a\,\e\, \u\ for Sp1.

Original speech signal Synthesized speech signal

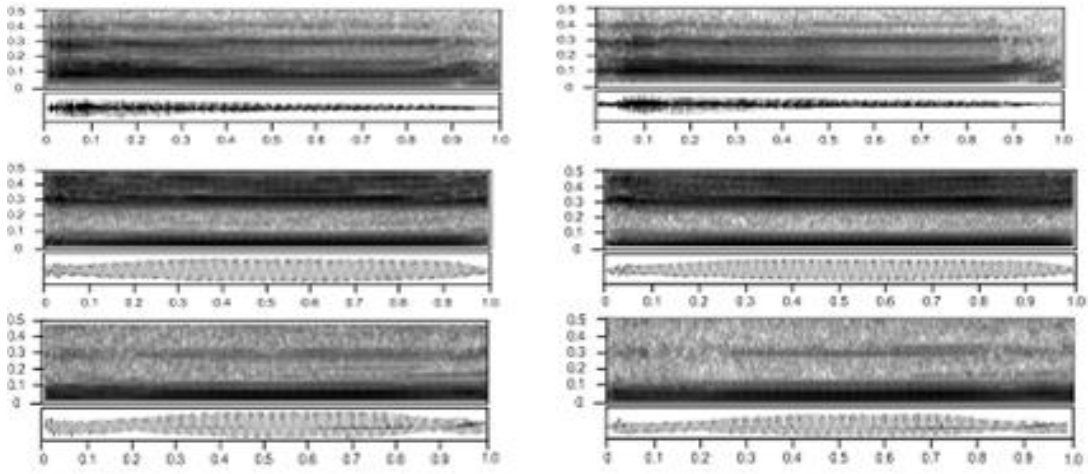

Fig. 3 Spectrogram of original and synthesized speech for vowels \a\, \e\, \u\ for Sp2.

Original speech signal Synthesized speech signal

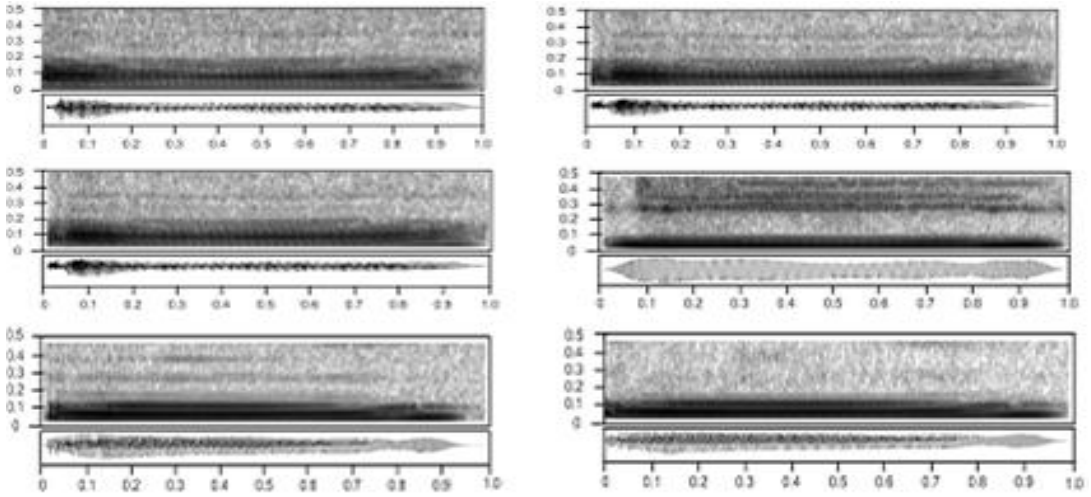

Fig. 4 Spectrogram of original and synthesized speech for vowels \a\, \e\, \u\ for Sp3.

Original speech signal Synthesized speech signal

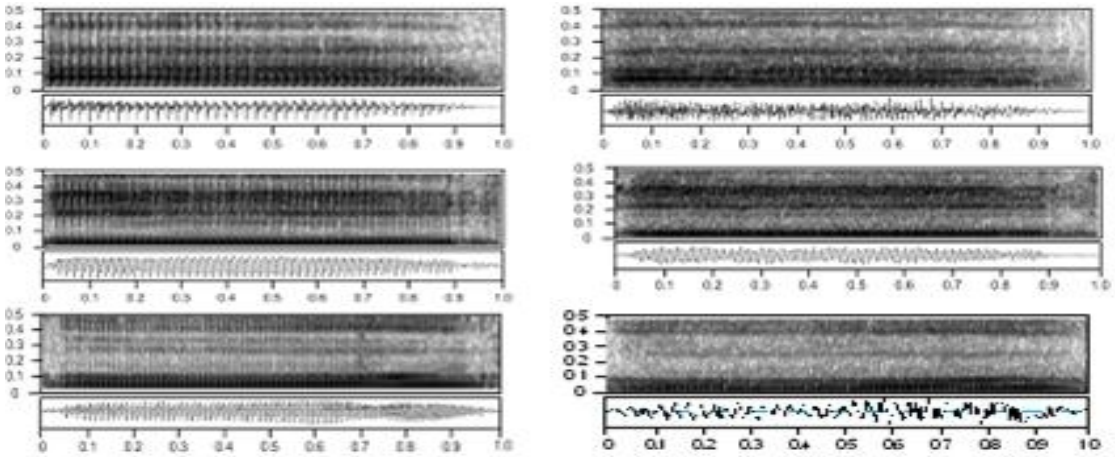

Fig. 5 Spectrogram of original and synthesized speech for vowels \a\, \e\, \u\ for Sp4.

Original speech signal Synthesized speech signal

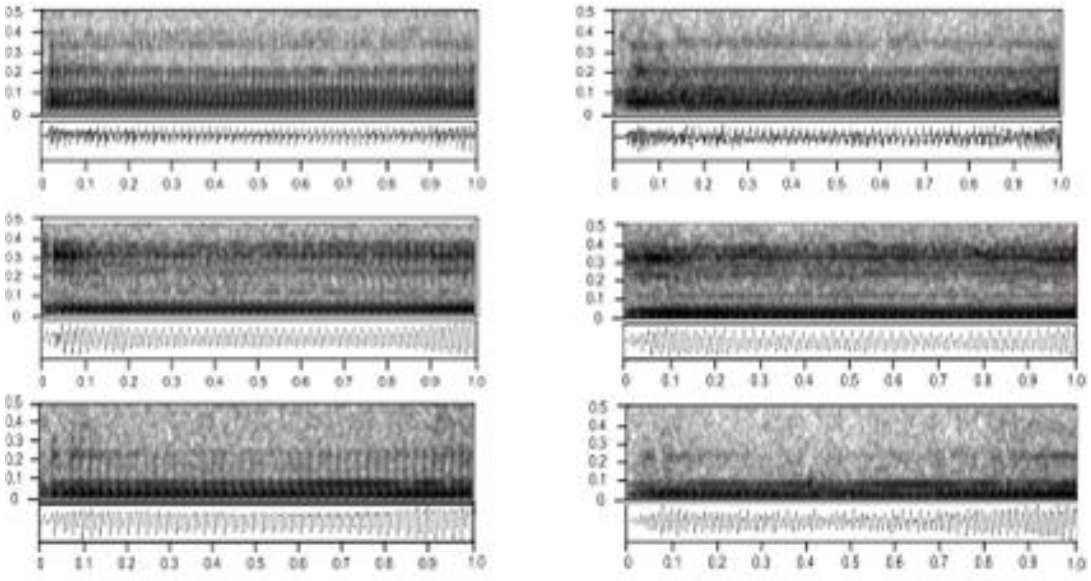

Fig. 6 Spectrogram of original and synthesized speech for vowels \a\, \e\, \u\ for Sp5.

Original speech signal Synthesized speech signal

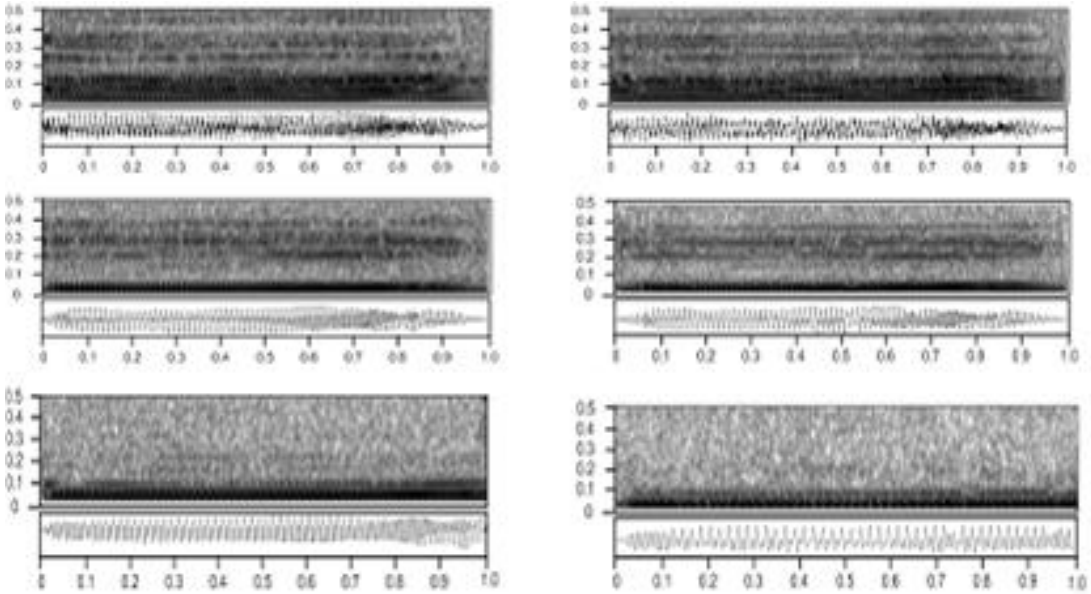

Fig. 7 Spectrogram of original and synthesized speech for vowels \a\, \e\, \u\ for Sp6

REFERENCES

[1] Ziolko B (2009) "Speech Recognition of Highly Inflective Languages", Ph.D. Thesis, University of York, Artificial Intelligence Group Pattern Recognition and Computer Vision Group, Department of Computer Science, United Kingdom, pp 1-122.

[2] Pantazis I (2006) "Detection of discontinuities in concatenative speech synthesis", Msc thesis ,University of Crete, Department of Computer Science, Heraklion, Fall.

[3] Gaikwad S K, Marathwada B A , Gawali B W, & Yannawar P (2010) "A review on speech recognition technique", Research Student Department of CS& IT, *International Journal of Computer Applications*, Nov., Vol. 10, No.3, pp 1-24.

[4] Plannerer B (2005) "An introduction to speech recognition", March, pp 1-68.

[5] Rabiner L R (1989*)* "A tutorial on hidden markov models and selected applications in speech recognition", in *Proc. of the IEEE,* Vol.77, No. 2.

[6] Elaydi H**,** Jaber M I**,** Tanboura M B "Speech compression using wavelets**"**, Electrical & Computer Engineering Department, Islamic University of Gaza,Gaza, Palestine.

[7] Jayant N ,fellow, Johnston J, & Safranek R (1993) " Signal compression based on models of human perception", *in proc. of IEEE signal processing research department* , oct,Vol. 81, No. 10 , pp 1385-1422.

[8] ZHOU Z (2001) " Data compression for radar signals an SVD based approach" , Msc thesis, state university of new york at binghamton, may, pp 1-57.

[9] Wall M E., Rechtsteiner A, & Rocha L M (2003) "Singular value decomposition and principal component analysis".

[10] *Lilly B **T** & Paliwal K K (1997)"* Robust speech recognition using singular value decomposition based speech enhancement" , *IEEE Tencon Speech and Image Technologies for Computing and Telecommunications*, Signal Processing Laboratory School of Microelectronic Engineering Griffith University.

[11] Akritas A G & Malaschonok G I (2002) "Applications of singular value decomposition (SVD)", Department of Computer and Communication Engineering, *University of Thessaly*, Greece, pp 1-15.

[12] Jolliffe, I. T (1986) " *Principal Component Analysis*" ,Springer-Verlag ,pp. 487.

[13] Zou H, Hastiey T and Tibshiraniz R (2004), " Sparse Principal Component Analysis", *Department of Health Research & Policy and Department of Statistics*, Stanford University, Stanford, April 26, pp 1-30.

[14] Dendrinos *et al*,& Loizou P C (2003) "A generalized subspace approach for enhancing speech corrupted by colored noise" ,Yi Hu, *Student Member, IEEE.*

[15] Kamm C, Walker M, & Rabiner L "The role of speech processing in human-computer intelligent communication", *Speech and Image Processing Services Research Laboratory AT&T Labs Research*, Florham Park, pp 1-26.

[16] Chakroborty S, & Saha G (2010) "Feature selection using singular value decomposition and QR factorization with column pivoting for text-independent speaker identification", *Speech Communication*, pp 693–709.

[17] Zehtabian A, & Hassanpour H "Optimized Singular vector denoising approach for speech enhancement", *Iranica Journal of Energy & Environment, IJEE an Official Peer Reviewed Journal* , Babol Noshirvani University of Technology, pp 166-180.

[18] Cao L (2007) "Singular Value Decomposition Applied to Digital Image Processing", Division of computing studies, Arizona state university polytechnic campus mesa, May, pp 1-16.