Effect of SVD Based Processing on the Perception of Voiced and Unvoiced Consonants

Balvinder kour, Randhir Singh, Parveen Lehana

Abstract— Speech is a biomedical signal used by the human beings to communicate. It is generated by exciting the vocal tract from the impulses of the air coming from the lungs through the vocal cords. Sometimes, the speech generated may not be adequate for understanding or transmission. In that case, it is modified using the concepts of speech processing. In this paper the singular value decomposition (SVD) technique is used to process and the output are evaluated using informal listening tests for investigating its effect on perception. This technique may have applications in speech compression, speech enhancement, speech recognition, and speech synthesis. The speech signal in the form of vowels-consonant-vowel (VCV) was recorded for the six speakers (3 males and 3 females). These VCVs were analyzed using SVD based technique and the effect of the reduction in singular values was investigated on the perception of the resynthesized VCVs using reduced singular values. Investigations have shown that the number of singular values can be drastically reduced without significantly affecting the perception of the VCVs.

Keywords— Speech signal, Speech generation, Speech processing, Speech compression, Singular value decomposition.

I. INTRODUCTION

Speech is the most important medium of communication between humans. Speech may contain message or any information. In terms of the signal speech may be defined as the signal carrying the message as the information in the acoustic wave form [1]. Speech signals may also consist of periodic sounds containing noise and some shirt silences. The vocal organs in human are in motion continuously, so the signal which is generated is not stationary [2]. The vocal cords are the primary source for speech production in humans. The process of human speech generation can be subdivided into three parts; the lungs, the vocal folds and the articulators. The vocal cords are vibrated when the adequate amount of air pressure is generated. The vocal cords are vibrated when the adequate amount of air pressure is generated by the lungs which act as a sound source and a set of filters that modifies the sound [3], [4].

Manuscript received on December, 2012.

Balvinder Kour, M.tech student, Electronics and Communication Engineering Department, Sri Sai College of Engineering and Technology, Badhani, Pathankot, Punjab, India.

Randhir Singh, Assistant Professor & Head, Electronics and Communication Engineering Department, Sri Sai College of Engineering and Technology, Badhani, Pathankot, Punjab, India.

Parveen Lehana, Associate Professor, Physics and Electronics Department, Jammu University, J&K, India.

The vocal cords are vibrating valve that caused due to the airflow from the lungs generating the audible pulses that form the laryngeal sound source which adjust the length and tension of the vocal cords for the fine pitch and tone. The articulators articulate and filter the sound from the larynx and also strengthen or weaken the larynx as a sound source. The combination of vocal cords with the articulators is capable of generating highly intricate arrays of sound. Normally the speech frequency ranges from 60-7000 Hz but in males it is 60-180 Hz and 160-180 Hz in females [3].The human speech production is shown in Fig.1.

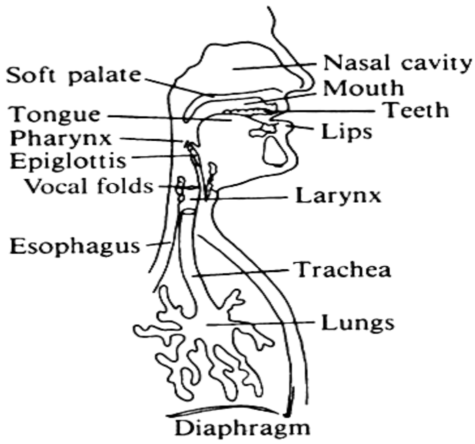

Fig. 1 Human speech production system [5].

In order to understand the impact of speech signal information, the speech recognition phenomenon is used. The speech recognition involves the process of generating the sequence of words which are almost equal to the input speech signal [6]. The signal models are very important to provide the information regarding signal processing system for processing the signal to obtain desire output and information regarding the source signal without its availability [7].Speech compression is the process of obtaining the compact speech signals for the efficient transmission or storage [1]. Due to complexity in the signals possessing varying nature, the signals are being compressed. This will also reduce the computational complexity of the signals [8]. For the compression of the signal, the information contained in it should be represented in small number of coefficient without losing its quality [9]. SVD is one of the speech compression techniques. It is used for reducing the complexity for computational process and noise reduction in the speech signal [10]. The speech consists of vowels and consonants. The vowels are produced when the vocal cords are closed, so all the vowels produced are voiced in nature [1].

Consonants are basically defined as the sounds which are articulated by temporary obstruction in the air stream passing through the mouth. The articulators cause the obstruction which may be total, intermittent, partial or narrowing sufficient to cause friction. Almost all the articulators are involved in the articulation of consonants. The consonants may be divided into unvoiced and voiced consonants. Voiced consonants are those which are articulated with the vibration of the vocal cords and the unvoiced consonants are articulated without vibration of vocal cords. During the production of unvoiced consonants the vocal cords are kept apart [11].The objective of this paper is to investigate the effect of voiced and unvoiced consonant 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 voiced and unvoiced consonants. 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.

II. SINGULAR VALUE DECOMPOSITION

The important parameters of a given speech signal can be derived by using the Singular value decomposition (SVD). SVD is very important for signal processing techniques such as image coding, signal enhancement and image filtering. SVD technique has been used to remove additive noise by varying the singular spectrum of the observed matrix of the speech signal [12]. In linear algebra, SVD is very important for mathematical aspects, calculating rank of the matrix, also the approximation of the parent matrix can be done by SVD. The SVD can be used in many other applications such as least square approximations for solving linear equations of the given system [13]. SVD has also find application in the speech enhancement of a given input signal. In this technique there were two SVD used and were applied to the overextended and over determined matrix which were obtained from the signal containing noise. Here SVD has been used as a pre-processor for recognising the speech signal [14]. For the enhancement of a given signal, robust speech recognition technique using SVD has been proposed to remove the noise in the signal due to environment which was increasing due to varying noise source [12]. Dendrinos et al. has also proposed the method for enhancing the speech signal using SVD which was based upon the idea that the given speech signal contains some of the Eigen vectors and its corresponding values and values contain the signal containing noise [15]. SVD in time domain has also been used for the enhancement of speech signal by separating the signal and noise from the input speech signal. In this technique the initial parameters has been estimated for enhancing the signal resulting in the improved signal to noise ratio of the signal [16].

III.METHODOLOGY

For analysing 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 [17].The recording of the six speakers including three males and three females having age between 20-25 years has been taken at the sampling frequency of 16 kHz. The recorded speech was segmented into different vowel consonant vowel (VCV) to investigate the effect of voiced and unvoiced consonants on the perception of the input speech signal. Various spectrograms of the segmented VCVs were obtained. The SVD of the spectrogram was taken for a particular band of value. The perceptual evaluation of the synthesized speech was carried out.

Table 1 Maximum and minimum SVD ratio in the unvoiced and voiced consonants for the synthesized and noise signal, respectively.

Speakers	Unvoiced Consonant		Voiced Consonant	
	Max.SVD synth ratio	Min SVD Noise ratio	Max.SVD synth ratio	Min SVD Noise ratio
Sp1	120:129	1:10	120:129	1:10
Sp2	120:129	1:10	120:129	1:10
Sp3	120:129	1:10	120:129	1:10
Sp4	120:129	1:10	120:129	1:10
Sp5	120:129	1:10	120:129	1:10
Sp6	120:129	10:20	120:129	1:10

IV.RESULTS AND CONCLUSIONS

The investigations were carried out using perceptual evaluation and spectrograms of the synthesized speech. Table I shows the maximum and minimum SVD ratio in the unvoiced and voiced consonants for the synthesized and noise signal, respectively. The spectrogram of two VCVs for six speakers is shown in Fig. 2 to Fig. 13.

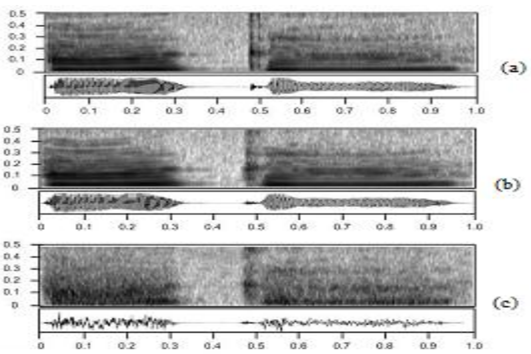

Fig. 2 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp1.

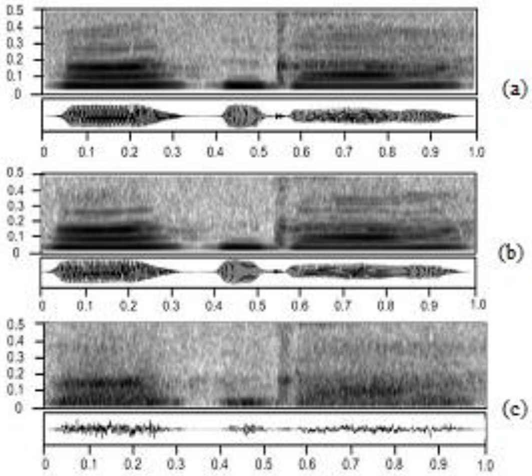

Fig. 3 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp1.

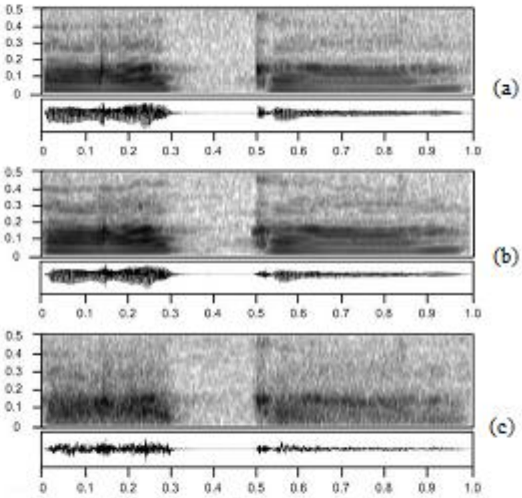

Fig. 4 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp2.

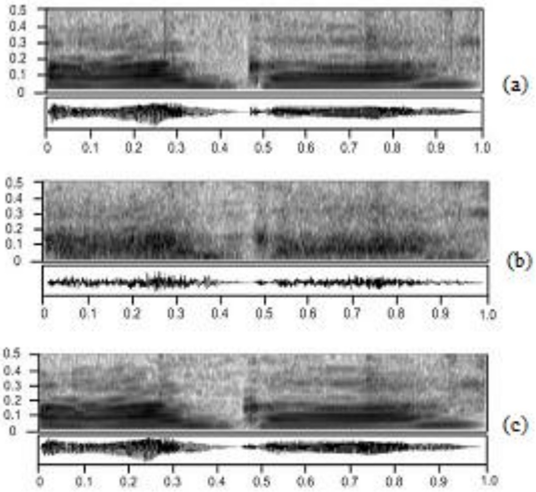

Fig. 5 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp2.

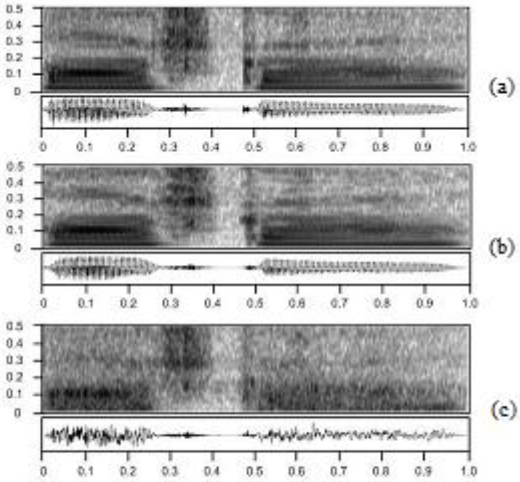

Fig. 6 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp3.

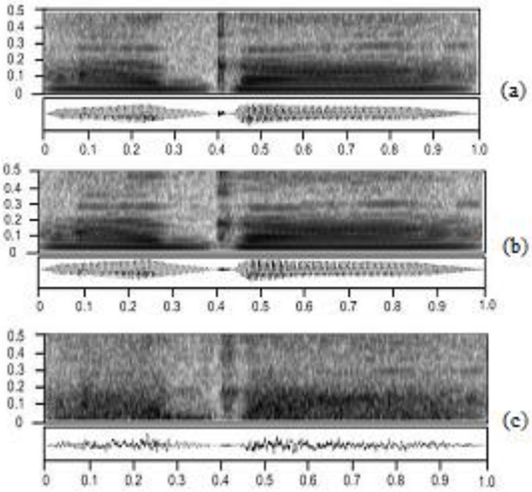

Fig. 7 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp3.

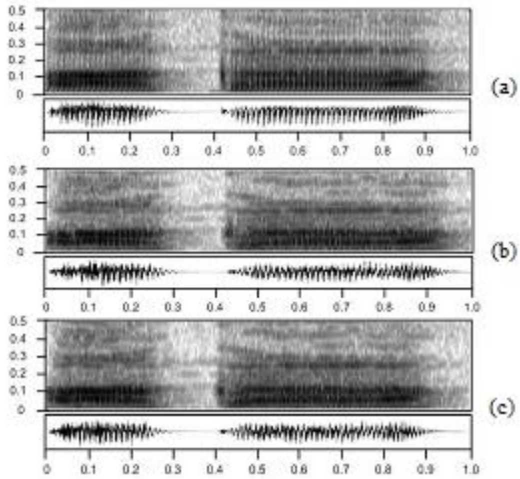

Fig. 8 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp4.

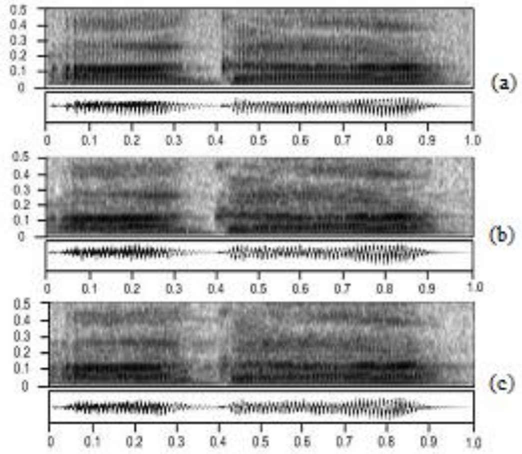

Fig. 9 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp4.

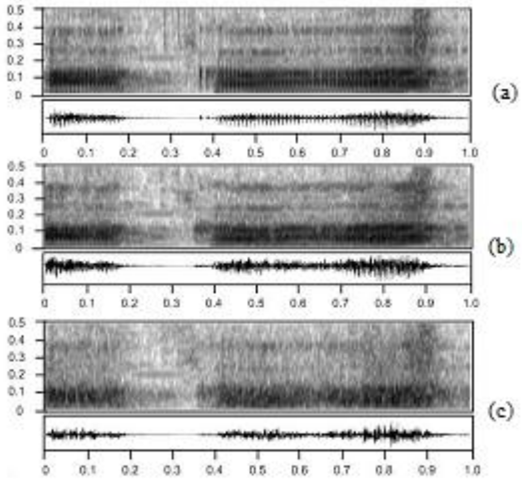

Fig. 10 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp5.

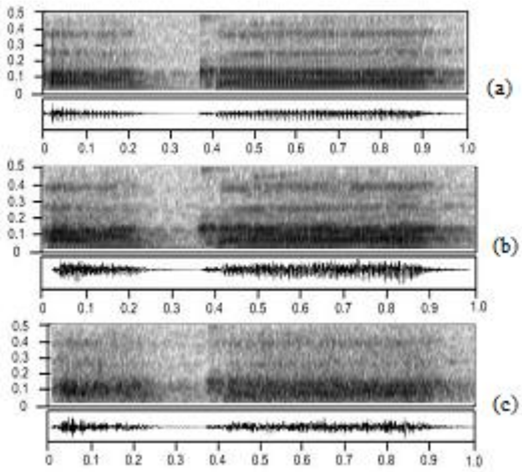

Fig. 11 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp5.

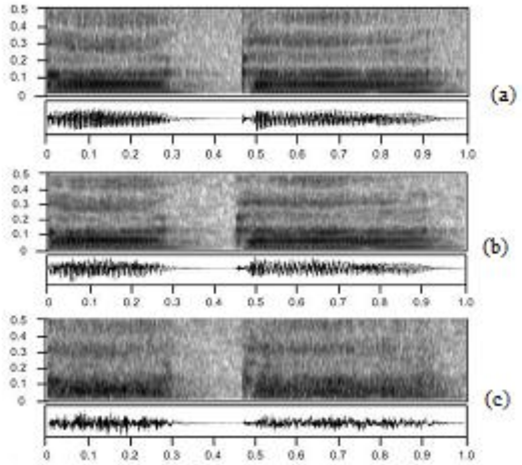

Fig. 12 Spectrogram of unvoiced consonant vkdk (a) Original speech signal (b) Synthesized speech signal (c) Noise speech signal for Sp6.

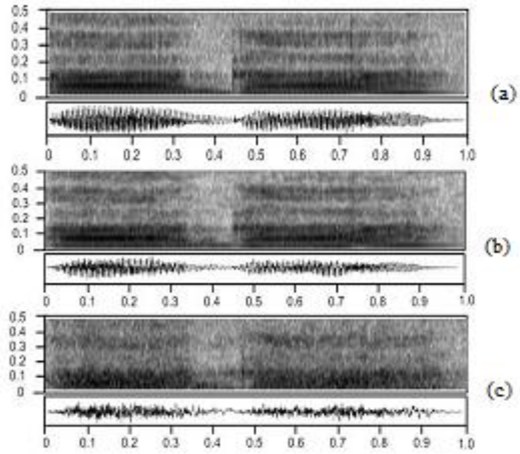

Fig. 13 Spectrogram of voiced consonant vkxk (a) Original speech signal (b)Synthesized speech signal (c) Noise speech signal for Sp6.

The first column shows the spectrograms of the recorded VCVs and the second column shows the spectrograms of the synthesized VCVs and third column shows the spectrogram of VCVs having noise in the signal at different singular values. The spectrograms of the synthesized VCV are smooth and visually similar to the original recorded VCVs. It was observed that first four or five band of values of the singular matrix are enough to synthesize the VCVs with satisfactory quality. It was also observed that there was some effect on the identity of the speaker when the important singular values from the beginning of the singular matrix were made zero.

ACKNOWLEDGMENT

First and above all, I would like to praise God, the almighty for providing me such opportunity and the hardworking capabilities to proceed successfully. I would like to pay a special thanks to my parents without them everything wouldn't be possible. It is honour to express my heartfelt gratitude to those who contributed towards the preparation on this research. I am indebted to my guides Prof. P.K Lehana and Prof. Randhir Singh whose invaluable guidance and timely suggestions inspired me to complete the research work.

REFERENCES

1. Marwan Al-Akaidi, “Excerpt.introduction to speech processing,” *Fractal Speech Processing* , De Montfort University, Leicester, 2004, pp 224.
2. J G Proakis and D G Monolakis, “Digital signal processing,” Fourth edition, pearson prentice hall, 2007.
3. I R Titze ,” Principles of Voice Production,” Prentice Hall,1994.
4. M. Dobrovolsky , “ Phonetics: The Sounds of Language,” Francis katamba, Heavenly labials in a world of gutturals, Wallace Stevens, pp 16 -58.
5. P.Palo,” A review of articulatory speech synthesis,” Master’s Thesis, Helsinki university of technology, Department of Electrical and Communications Engineering, *Laboratory of Acoustics and Audio Signal Processing*, Espoo, June 5, 2006, pp 1-126.
6. S.K Gaikwad, B.A Marathwada and P Yannawar , “A review on speech recognition technique,” *International Journal of Computer Applications* , Department of CS& IT, University Aurangabad, Vol. 10,No.3, November 2010, pp 16-24.
7. L.R Rabiner “A tutorial on hidden markov models and selected applications in speech recognition,” in *Proc. of the IEEE*,1989, Vol.77, No. 2, pp 257-286.
8. M G Christenseny, Jan ostergaardz, and S H Jensenz, “On compressed sensing and its application to speech and audio signals,” Dept. of Media Technology, Aalborg University, Denmark.
9. Elaydi H**,** Jaber M I**,** Tanboura M B “Speech compression using wavelets,**”** Electrical & Computer Engineering Department, Islamic University of Gaza, Palestine.
10. F. Khakpoor and G. Ardeshir, “Using PCA and SVD to improve wavelet-based method for detection of voice and silence in speech,” *European Journal of Scientific Research* , Faculty of Electrical & Computer Engineering, Babol Noushirvani University of Technology ,Babol, Iran , Vol.37 , No.4, 2009, pp 641-648.
11. T McCormick, B Langford and P Pikkert etal , “Phonetics made easy a manual of language acquisition for cross cultural effectiveness compiled and adapted by various individuals,” Summer Institute of Linguistics, LACE Version ,pp 2-46.
12. K Hermus, I Dologlou, PP Wambacq and D V Compernollel , “Fully adaptive svd-based noise removal for robust speech recognition,” Katholieke Universiteit Leuven, Belgium .
13. Bethany Adams and Nina Manual, “ Using the Singular Value Decomposition Particularly for the Compression of Color Images,” November 13, 2005.
14. B T Lilly and K K Paliwal ” Robust speech recognition using singular value decomposition based speech enhancement,” *IEEE Tencon Speech and Image Technologies for Computing and Telecommunications*, Signal Processing Laboratory School of Microelectronic Engineering Griffith University,1997, pp 257-260.
15. Y Hu “Subspace and multitaper methods for speech enhancement,” Phd Thesis, the university of texas at dallas,doctor of philosophy in electrical engineering,december 2003,pp 1-138.
16. B Nazari, S Sarkarni and P Karimi, “A method for noise reduction in speech signal based on singular value decomposition and genetic algorithm,” *IEEE Confrence publications Eurocon*, pp 102 -107, 2009.
17. L Cao, “Singular Value Decomposition Applied to Digital Image Processing,” Division of computing studies, Arizona state university polytechnic campus mesa, 2007, pp 1-16.

AUTHORS PROFILE

Er. Balvinder Kour, resident of Jammu (J&K), received her B.Tech degree in Electronics and Communication Engineering from M.B.S College of Engineering and Technology (MBSCET) in the year 2010.She is now pursuing M.Tech in Electronics and Communication Engineering.

Er. R.S Andotra, (Assistant Professor and Head) received his M.Tech. Degree in Electronics and Communication Engineering from Beant College of Engineering and Technology, Gurdaspur (Punjab) affiliated to Punjab Technical University, Jalandhar (Punjab). He is presently working as Assistant Professor (H.O.D), Electronics and Communicaton Engineering Department, Sri Sai College of Engineering and Technology, Pathankot (Punjab) and pursuing in PhD. Electronics and Communication from Punjab Technical University. His research interests include Speech signal processing, Digital signal Processing, Image processing, Analog and Digital Communication, Electronics and control systems, etc. and having more than 10 publications in national/international conferences and journals and a lot of experience in guiding M.Tech students.

Dr. P.K Lehana, (Associate Professor) received his Master’s degree in Electronics from Kurushetra University in 1992. He worked as lecturer in Guru Nanak Khalsa College, Yamuna nagar, Haryana for next two years. He qualified NET-JRF in Physical science in 1994 and got selected as permanent lecturer in A. B. College, Pathankot, where he worked for one year. He also qualified NET-JRF in Electronic Science and presently working as Associate Professor in Physics and Electronics Department, University of Jammu and did Ph.D. degree from IIT, Bombay in Speaker Transformation.. He also invited for conducting workshops on MATLAB/simulinks in different esteemed institutions/colleges. His research interests include Speech recognition, Speaker transformation, Signal processing, Speech signal processing, Analog and Digital signal processing, Nanowires characterization, Robotics, Image processing, Analog communication, Digital communication, Microwaves and Antennas, Electronics and control systems, Instrumentation, Electronics system designing, etc. and having more than 100 publications in national/international conferences and journals. He has a lot of experience in guiding M.Tech, M.Phil, Ph.D. students and other researchers also.