International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-1, Issue-3, July 2011

Optimization of Artificial Neural Network for Speaker Recognition using Particle Swarm Optimization

Rita Yadav, Danvir Mandal

***Abstract*— this paper proposes a particle swarm optimization (PSO) based optimization technique for Artificial Neural Network weights optimization for speaker recognition. PSO is a search algorithm, in which each potential solution is seen as a particle with a certain velocity flying through the problem space. The particle swarms find optimal regions of the complex search space through the interaction of individuals in the population. PSO is attractive for optimization in that particle swarms will discover best optimized value as they fly within the subset space. Combining the ANN and PSO algorithms improves the performance as compared to that ANN alone.**

***Index Terms*— Artificial Neural Network (ANN), Feature Extraction, Matlab, Mel Frequency Cepstral Coefficient, Particle Swarm Optimization, Speaker Recognition.**

I. INTRODUCTION

The speaker recognition systems fall into two main categories, namely: speaker identification systems and speaker verification systems. In speaker identification, the goal is to identify an unknown voice from a set of known voices. Whereas, the objective of speaker verification is to verify whether an unknown voice matches the voice of a speaker whose identity is being claimed [1,2].

Speaker identification systems are mainly used in criminal investigation while speaker verification systems are used in security access control. Speaker identification systems can be closed-set or open-set. Closed-set speaker identification refers to the case where the speaker is known member of a set of speakers. Open-set speaker identification includes the additional possibility where the speaker may not be member of the set of speakers [3,4].

Another distinguishing feature of speaker recognition systems is whether they are text-dependent or text independent [2]. Text-dependent speaker recognition systems require that the speaker utter specific phrase or password.

Manuscript received Jun 30, 2011.

Rita Yadav, Student (M. Tech, Electronics and Communication Engineering), Punjab Technical University/ Institute of Engineering & Technology,Bhadda (Ropar), Punjab, India, 9560396935, (e-mail: ritayadav226@gamil.com).

Danvir Mandal, Assistant Professor, Electronics and Communication Engineering, Punjab Technical University/ Institute of Engineering & Technology,Bhadda (Ropar), Punjab, India, 9888081214, (e-mail: danvir_mandal@rediffmail.com).

Text-independent speaker recognition systems identify the speaker regardless of his utterance [2]. This paper focuses on the closed-set text-independent speaker identification.

In speaker recognition system, first step is feature extraction. There are so many methods developed so far for feature extraction. The most commonly used feature extraction method is Mel Frequency Cepstral Coefficient (MFCC) which is based on frequency domain of Mel scale for human ear scale [5]. The main steps involved in MFCC feature extraction method are preprocessing, framing, windowing, DFT, Mel filter bank, Logarithm and Inverse DFT. The feature vectors thus obtained are given as input to Artificial Neural Network (ANN).

There has been a great deal of interest in application of (ANN) for pattern matching problems. An ANN consists of simple processing units that are interconnected by different weights [6,7]. The interconnecting topology between the units and the weights of the connections defines the operation of the network. In feed forward network there is a set of units that are designated as input units through which input features fed to the network. Then there are layers of hidden units that extract the features from the inputs and then there are units of output units where in classification task each corresponds to a class. Recently there have been significant research efforts to apply evolutionary computation (EC) techniques for the purposes of evolving one or more aspects of artificial neural networks [8]. Evolutionary computation methodologies have been applied to three main attributes of neural networks: network connection weights, network architecture (network topology, transfer function), and network learning algorithms. Most of the work involving the evolution of ANN has focused on the network weights and topological structure

From the beginning of 90's, new optimization technique researches using analogy of swarm behavior of natural creatures have been started. Particle swarm optimization, a new branch of the soft computing paradigms called evolutionary algorithms (EA), was developed by Kennedy and Eberhart (1995) [9]. It is a group-based stochastic optimization technique for continuous nonlinear functions. It is a simple concept adapted from natural decentralized and self-organized systems where all the particles move to get better results.

In this paper ANN is used for pattern matching. The network connection weights are then optimized using PSO and comparison in then made on weights of ANN.

In next section, we will introduce the Artificial Neural Network. Section 3 describes Particle Swarm Optimization (PSO) method. Section 4 describes the experimental evaluation on speaker recognition. Finally, section 5 summarizes the conclusion drawn from this study.

II. Artificial Neural Network (Back Propagation)

Speech recognition is a multileveled pattern recognition task, in which acoustical signals are examined and structured into a hierarchy of sub word units (e.g., phonemes), words, phrases, and sentences [10]. Artificial neural networks have emerged as a promising approach to the problem of speech recognition [11,12]. ANN can be most adequately characterized as computational models' with particular properties such as the ability to adapt or learn, to generalize, or to cluster or organize data, and which operation is based on parallel processing which are the requirements for speech and speaker recognition. ANN can learn complex features from the data, due to the non-linear structure of artificial neuron [11, 13]. Various ANN training algorithms such as BPA, Radial Basis Function, Recurrent networks etc., are being used for training purpose.

We have used Back propagation Neural Network [14] for the recognition system. It has been successfully applied to many pattern classification problems including speaker recognition [13].Back propagation is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions [15]. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as required.

Properly trained back propagation networks tend to give reasonable answers when presented with inputs that they have never seen. Typically, a new input leads to an output similar to the correct output for input vectors used in training that are similar to the new input being presented [16]. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs [13].

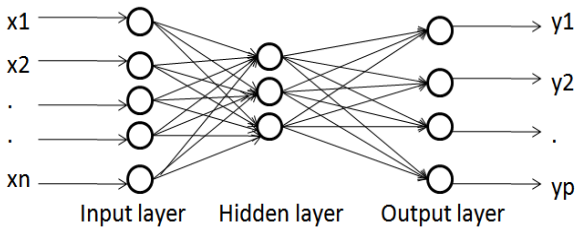

Fig. 1 Architecture of Feed Forward Neural Network

III. Particle Swarm Optimization

Swarm intelligence (SI) is an innovative distributed intelligence paradigm for solving optimization problems that originally took its inspiration from the biological phenomena of swarming, flocking, and herding [17]. Particle swarm optimization (PSO) incorporates group behaviors observed in flocks of birds, schools of fish, swarms of bees, and even human social behavior, from which the idea emerged. PSO is a population based optimization tool which can be easily implemented and applied to solve various function optimization problems, or the problems that can be transformed into function optimization problems. As an algorithm, the main strength of PSO is its fast convergence, which compares favorably with many global optimizations. PSO algorithms are especially useful for parameter optimization in continuous, multi-dimensional search spaces.

In PSO, each particle keeps track of its coordinates in the solution space which are associated with the best solution (fitness) that has achieved so far by that particle. This value is called personal best, pbest [18].

Another best value that is tracked by the PSO is the best value obtained so far by any particle in the neighborhood of that particle. This value is called gbest [18].

In PSO each particle tries to modify its position using the following information:

- the current positions,
- the current velocities,
- the distance between the current position and pbest,
- the distance between the current position and the gbest.

The modification of the particle's position can be mathematically modeled according the following equation:

$$V_i^{k+1} = wV_i^k + c_1 * rand_1(\,) * (pbest_i - s_i^k) + c_2 * rand_2(\,) * (gbest - s_i^k) \quad (1)$$

where, V_i^k : velocity of agent i at iteration k,
w: weighting function,
c_j : weighting factor,
rand: uniformly distributed random number between 0 and 1,
s_i^k : current position of agent i at iteration k,
$pbest_i$: pbest of agent I,
gbest: gbest of the group.

The following weighting function is utilized in (1).

$$w = wMax - [(wMax - wMin)iter] / maxIter \quad (2)$$

where wMax= initial weight,
wMin = final weight,
maxIter = maximum iteration number,
iter = current iteration number.

The current position (searching point in the solution space) can be modified by the following equation:

$$S_i^{k+1} = s_i^k + V_i^{k+1} \quad (3)$$

Fig. 2 shows the general flow chart of PSO, which can be described as follows:

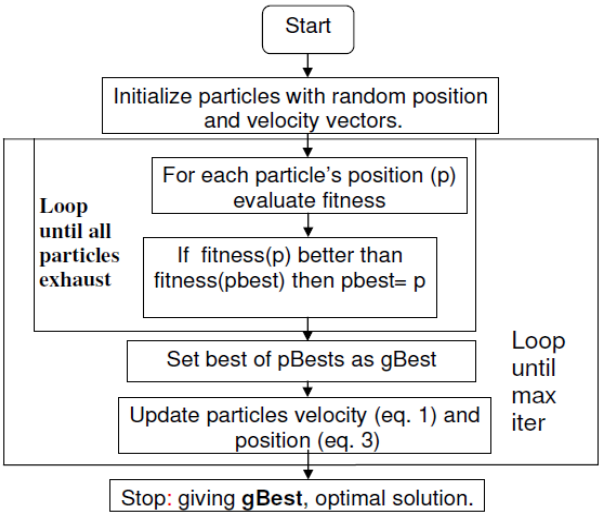

Fig. 2: A general flow chart of PSO

PSO Algorithm:

Step-1: Initialize the population - position and velocities
Step-2: Evaluate the fitness of the individual particle (pBest)
Step-3: Keep track of the individuals highest fitness (gBest)
Step-4: Modify velocities based on pBest and gBest position.
Step-5: Update the particles position
Step-6: Terminate if the condition is met
Step-7: Go to Step 2.

Fig. 3 shows the Block Diagram for Speaker Recognition.

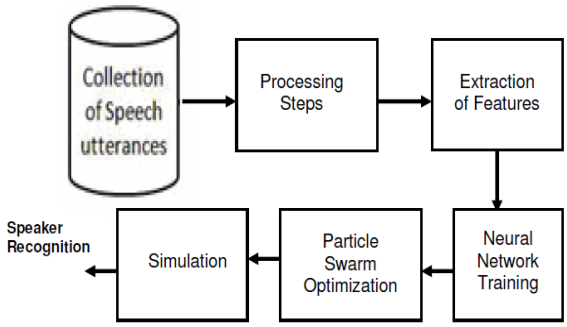

Fig. 3: Block Diagram for Speaker Recognition

IV. Experimental Evaluation

A. Experimental Setup

There are two speaker databases in the simulation Experiment, as Training and test. In each speaker database there are 11 speakers. The experiment is tested in Matlab 7.5. The sound signal is first preprocessed by removing the silence zones, which are removed by means of short time energy calculation. Segments of 15ms are chosen for this purpose. A segment whose energy is less than some threshold relative to the average energy of the entire signal is discarded. A high emphasis filter H(z) = (1−0.95z−1) is applied to the speech signal. The speech signal is then divided into analysis frames where the signal can be assumed to be stationary. For accurate estimation of voiced speech, the sampling frequency should be more than 6 kHz to include a speech bandwidth of at least 3 kHz. The underlying platform is a Pentium 4 PC that takes the speech signals from a speaker, do Fast Fourier Transform (FFT) to extract features, feed the features to a trained classifier and return the identity of the speaker in a real-time fashion. The feature set used for encoding the speech signal has a form of cepstral coefficients computed as follows:

1. Partition the speech signals into disjoint half-frames of 128 points. Neighboring half frames are jointed to form a complete frame of 256 points. Since the sampling frequency is 8 kHz, each frame lasts 256/8 = 30msec.

2. Window each frame using hamming window of 15msec.

3. Compute the cepstral coefficients of each frame using equation:

$$\text{cepstrum}\{\text{frame}\} = \text{FFT}^{-1}(\log|\text{FFT}\{\text{frame}\}|) \quad (3)$$

4. The first 12 coefficients of the cepstrum is taken as the feature vector of the frame.

The feature vector obtained is first trained with ANN and then optimized using PSO for SR

B. Results

The experimental results are illustrated in Fig.4-10.

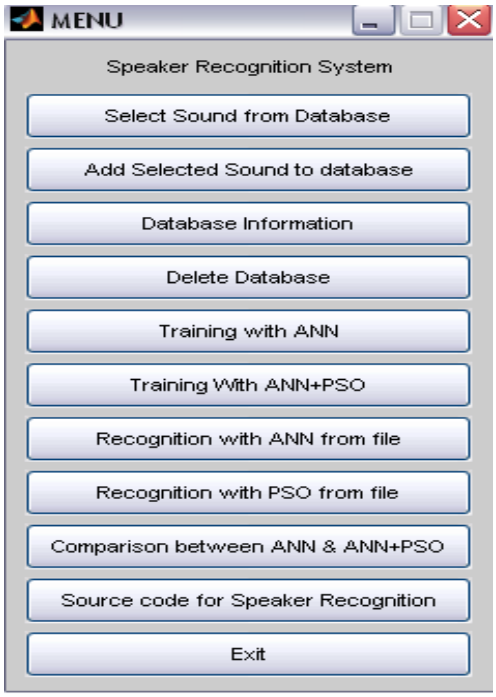

Fig. 4: Menu for Speaker Recognition.

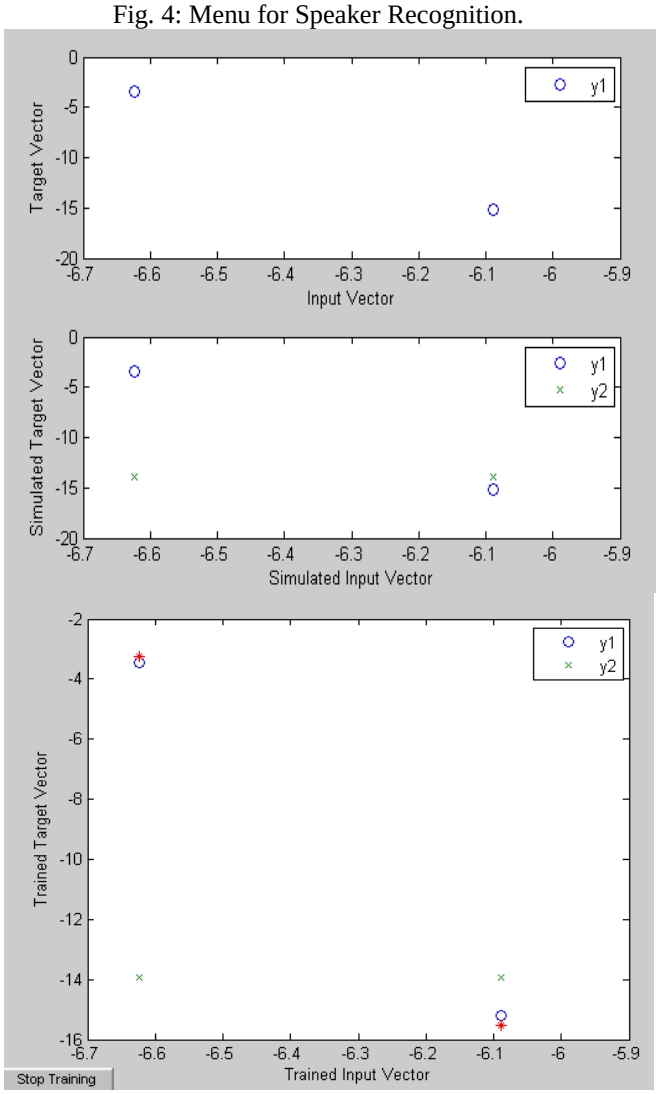

Fig. 5: ANN Training.

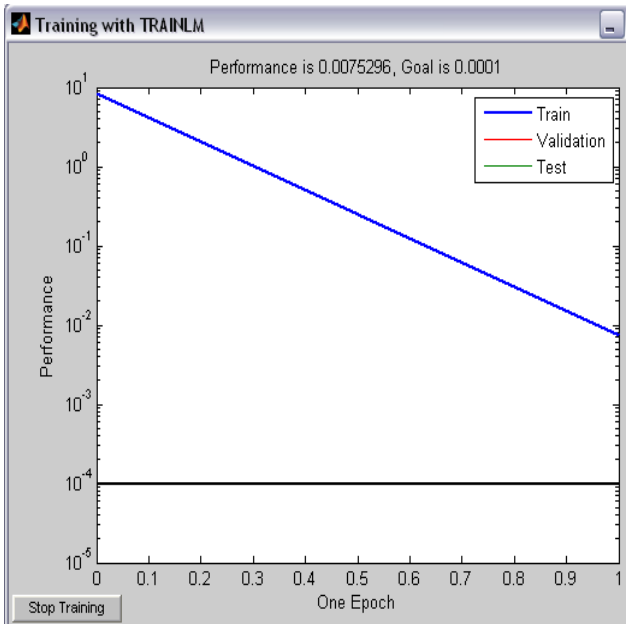

Fig. 6: ANN Training.

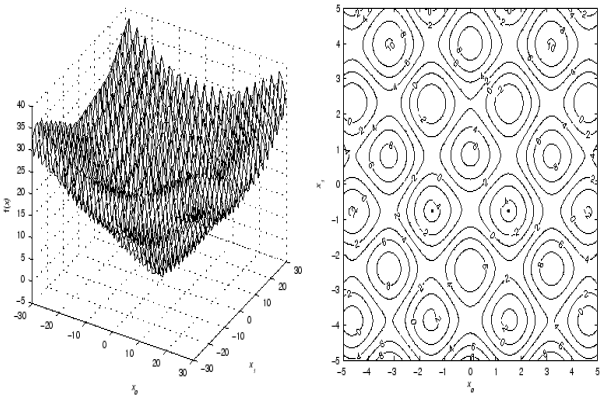

Figure 7: Plot of Ackley Function.

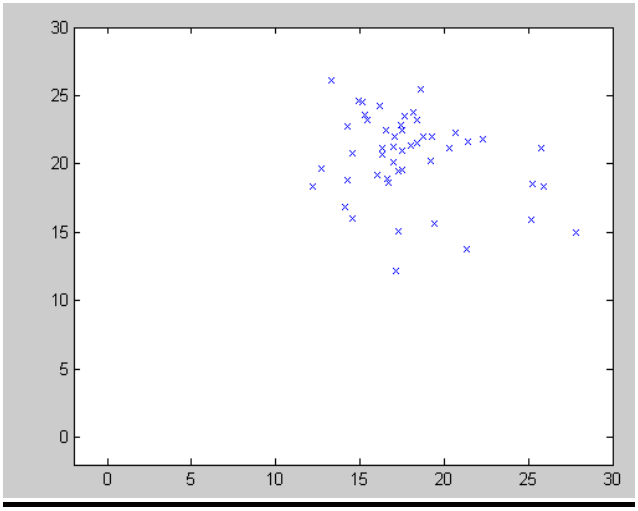

Fig. 8: Particle Searching for Best Position

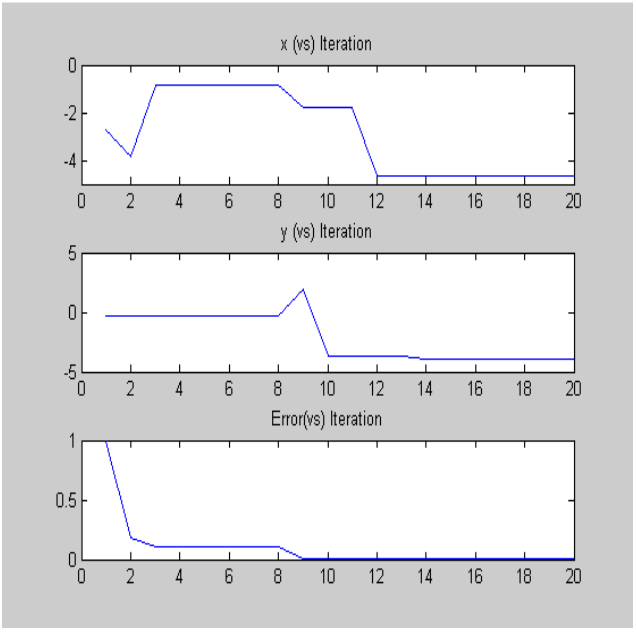

Fig. 9: Optimization for Weights using PSO.

ANN Weights	Optimized ANN weights by PSO
-0.3001	0.0655
-0.0935	0.0327
-0.7458	0.0147
-0.8370	0.0059
-0.4211	0.0021
0.3536	6.1872e-004
-0.0859	1.5468e-004
-0.0016	3.0936e-005
-0.8830	4.6404e-006
-0.8528	2.3202e-008

Fig. 10: Comparison for Weights of ANN & PSO.

V. Conclusion

We presented an efficient PSO-based optimization technique to enhance the performance of Artificial Neural Network for Speaker Recognition by means of optimizing ANN weights. Finding a more relevant fitness function or using a type of discriminative training may also improve the results and achieve superior performance.

Acknowledgment

Rita Yadav received her Degree from Institute of Electronics and Telecommunication Engineers (IETE), New Delhi, India in 2007. She is currently pursuing Master Degree in Electronics and Communication Engineering from Punjab Technical University, Jalandhar, India. She wants to thanks the college faculty members for their supervision and guidance.

References

[1] D.O. Shaughnessy, "Speaker Recognition", ASSP Magazine, IEEE Signal Processing Magazine, Vol. 3, No. 4, Part. 1, pp. 4-17, October 1986.
[2] J. P. Campbell, JR, "Speaker Recognition: A Tutorial" , Proceedings of the IEEE, Vol. 85, No. 9, September 1997.
[3] A.E. Rosenberg, "Automatic speaker verification: A review," Proc IEEE, vol. 64(4), pp. 475-87, Apr. 1976.
[4] H. Gish, and M.Schmidt, "Text-indepent speaker identification," IEEE Signal Process. Mag., vol. 18, pp.18-32, Oct. 2002.
[5] Sirko Molau, Michael Pitz, Ralf Schlü ter, and Hermann Ney, " Computing Mel-Frequency Cepstral Coefficients on the Power Spectrum" Lehrstuhl für Informatik VI, Computer Science Department, RWTH Aachen – University of Technology, 52056 Aachen, Germany.
[6] T. Kohonen, "Self-organization and Associative Memory" Springer-Verlag, Berlin- New York, 1988a.
[7] B.D. Ripley, "Neural Networks and Related Methods for Classifications" Journal of Royal Statistics Society, B56,pp. 409-456,1994.
[8] (PSO Tutorial), Xiaohui Hu. Available: http://www.swarmintelligence.org/tutorials.php
[9] J. Kennedy and R. Eberhart, "Particle Swarm Optimization", Proceedings of IEEE International Conference on Neural Networks (ICNN'95), Vol. IV, pp.1942-1948, Perth, Australia, 1995.
[10] Campell J.P. and Jr., "Speaker recognition: a tutorial" Proceeding of the IEEE, Vol 85, pp. 1437-1462, 1997.
[11] Y.-Yan, M. Fanty, and R. Cole, "Speech Recognition Using Neural Networks with Forward-backward Probability Generated Targets", Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Munich, April 1997.
[12] Shukla, Anupam, Tiwari, Ritu, "A novel approach of speaker authentication by fusion of speech and image features using Artificial Neural Networks", Int. J. of Information and Communication Technology 2008-Vol.1,No.2 pp . 159 – 170.
[13] R. P. Lippmann, "Review of Neural Networks for Speech Recognition," Neural Computation, Vol. 1, No. 1, pp. 1-38, 1989.
[14] Ehab F., M. F. Badran , Hany Selim "Speaker Recognition Using Artificial Neural Networks Based on Vowel phonemes" Electrical Engineeling Department, Assiut University.
[15] Zahorian, S. A., "Reusable Binary-Paired Partitioned Neural Networks for Text-Independent Speaker Identification, Proc. ICASSP-99, pp. II: 849- 852, 1999.
[16] Zebulum, R.S. Vellasco, M. Perelmuter, G. Pacheco, 'A comparison of different spectral analysis models for speech recognition using neural networks" Departamento de Engenharia Eletrica, PUC, Rio de Janeiro, Brazil; IEEE 1996.
[17] A. Abraham, H. Guo, and H. Liu, "Swarm intelligence: Foundations, perspectives and applications, swarm intelligent systems," in Studies in Computational Intelligence. Berlin, Germany: Springer Verlag, pp. 3-25, 2006.
[18] www23.Homepage.Villanova.edu/Varadarajan…./PSO_meander-line .ppt

Rita Yadav received her Degree from Institute of Electronics and Telecommunication Engineers (IETE), New Delhi, India in 2007. She is currently pursuing Master Degree in Electronics and Communication Engineering from Punjab Technical University, Jalandhar, India. She worked as Junior Engineer (R &D) in Recorders and Medicare System Pvt. Ltd., Chandigarh, India in 2006. She worked in C-DAC, Mohali, India as a part of PCB Designing Lab. in 2006-2010. She is presently working as Junior Technical officer (JTO) in Ministry of Defence, New Delhi, India. She wants to thanks the college faculty members for their supervision and guidance.

Danvir Mandal received his Bachelors Degree in Electronics and Communication Engineering from Punjab Technical University, Jalandhar, India in 2001, and Masters Degree in Electronics and Communication Engineering from Punjab Technical University, Jalandhar, India in 2006. He is currently pursuing Doctorate degree from NITTTR, Chandigarh, India. He was a lecturer with Department of Electronics & Communication Engineering, Institute of Engineering & Technology, Bhaddal, Punjab, India in 2006. Presently, he is an Assistant Professor with Department of Electronics & Communication Engineering, Institute of Engineering & Technology, Bhaddal, Punjab, India. His research interests include digital signal processing, image processing, antenna design and analysis, FDTD methods.