ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Comparison of Warping Filter Banks using MATLAB Based Feature Extraction Techniques in ASR

S.Suganya[1], M.Selvaganapathy[2], N.Nishavithri[3]

Assistant Professor, Dept. of ECE., CK College of Engineering & Technology, Cuddalore, Tamil Nadu, India[1,2]

Assistant Professor, Dept. of ECE., Mailam Engineering College, Mailam, Tamil Nadu, India[3]

ABSTRACT: In recent years, Speech Recognition has the great development in the automation industry. This paper proposes an Automatic Speech Recognition (ASR) to facilitate an interaction between human and the electronic components. The main concern of this paper involves the suppression of various noises to achieve a robust speech recognition system. Discrete Hidden Markov Model is used to increase the speed of speech recognition. This paper explores the hardware realization of desired speech recognition system on the Field Programmable Gate array (FPGA). The accuracy has to be increased to get the clear and robust Speech Recognition. The speech features can be extracted through the cepstral coefficients by using warping filter banks. The cepstral coefficients are used to increase the robustness of Speech Recognition. To minimize the complexity of desired ASR system, the number of coefficients has to be minimized. The Speech- to-Text conversion is the main objective of this paper. This can be achieved by using Matlab software, Modelsim 6.4a for simulation and can be implemented in Altera DE2-70 board.

KEYWORDS: Feature Extraction, Discrete hidden Markov Model, Warping filter banks, speech-to-text.

I. INTRODUCTION

Speech Recognition has been the most dominant and convenient means of communication. Speech communication is not only a face-to-face interaction but also the individuals at any moment, via a wide variety of modern technological media. Speech Recognition plays an important role that a human to make an interaction with the electronic components for automation systems. Actually, Speech Recognition is also a kind of Pattern Recognition technique. Various applications of Speech Recognition includes voice controlled devices, speech-to-text, etc. Automatic Speech Recognition can be resolved into two phase viz., training phase and testing phase. In training phase, the speech feature vectors can be extracted and is trained in the codebook. In testing phase, the feature vectors can be obtained as in the testing phase and also comparing the testing features are matching with the codebook. If both the features are similar, this given speech can be recognized and that can be utilized for an authentication purpose.

Section II describes the survey from there we have got few ideas to implement this concept. Section III describes the proposed methodology of this paper. Section IV explores the steps has been followed in the feature extraction. Section V explains the algorithm by which the proposed strategy is designed. Section VI explores the hardware architecture which is designed to implement in FPGA. Section VII discussed the results which we obtained for the desired system. Section VIII explains the concepts which are concluded from this paper. Section IX explains the concepts from various papers that are referred.

II. LITERATURE SURVEY

Speech Recognition is an advance technique to be followed in the fields of Automation, Artificial Intelligence and so on. The improvement in the recognition accuracy, robustness may cause effects on the performance of ASR. MFCC and PLP are most widely used feature extraction techniques which are required to reconstruct the original signal [1]. The recent focus of researchers involved in the implementation of ASR into an embedded platform [2]. The speech-to-text conversion is a useful technique which is helpful for handicapped peoples [3]. This paper focusing on the speech-to-text conversion and can be implemented in Altera DE2-70 board. The robustness of speech recognition can be

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

improved by suppressing the noise from the speech signal. The conversion of speech-to-text depends on the variation in the frequency. As the number of cepstral coefficients increases, the desired system gets complex to achieve the desired goal. So the number of coefficients here is 12.

The speech signal has to be in wave format, then only the signal to be processed and feature extraction can be done. The speech signal is sampled at 16 Khz and the number of bits per sample as 32. The reasonable modeling of speech signal can be done according to the assumption that such a small segment of speech is sufficiently stationary [4].

III. PROPOSED ALGORITHM

The Proposed Algorithm consists of the feature extraction module and the codebook generation. The proposed architecture has been shown in Fig.1, which explodes the steps has been followed in this paper. The original signal has to be split up into number of frames according to the frame length. The low frequency signals are selected by blocking the low frequency in each and every frame. After frame blocking of each and every frames, Hamming windowing is applied to reduce the discontinuity of the signal. Determine the DCT coefficients for spacing to each and every windowed frame. Apply logarithmic values to get the DCT values as a single value [5].

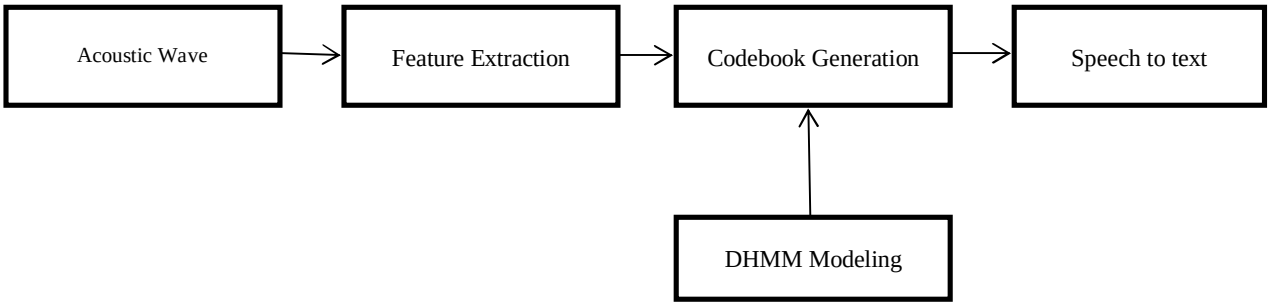

FIG.1. Proposed Algorithm

IV. FEATURE EXTRACTION TECHNIQUES

Fig.2 describes the block diagram of feature extraction techniques. The steps which are followed for the desired system has been explained below.

A. *Frame Blocking:*

Frame blocking is applied to divide the signals into matrix form with an appropriate time length for each frame. The speech signal taken in wave format and is sampled at 16,000 Hz and the number of frames could be assumed as 320 samples within a frame. Overlapping of frames would have the factor of separation of samples due to the effect of frame blocking [5].

Step 1: Hamming Windowing:

To reduce the discontinuities of signal at the end of each frames, hamming windowing is applied to each and every frames after frame blocking. The equation (1) representing the discrete time representation of signal,

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \quad \text{eq. (1)}$$

By introducing hamming windowing to each and every frames, windowing generates the least distortion [5].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

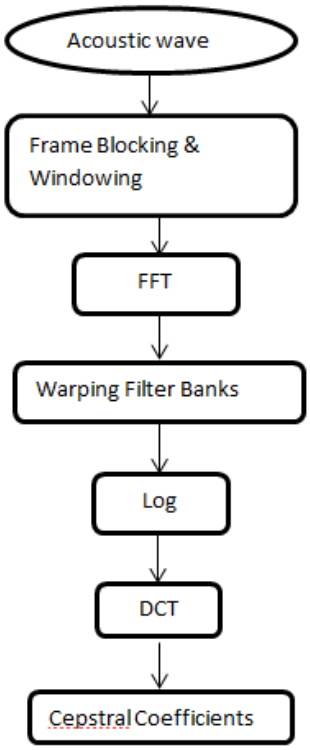

FIG.2. Feature ExtractionTechniques

B. *Fast Fourier Transform:*

Time domain signal can be converted into frequency domain by applying fast fourier transform to each and every windowed frames. The output of FFT can be complex numbers having both real and imaginary parts. Real time data has to be processed with the speech recognition system. The complex variables could be neglected by the FFT [5]. Equation (3.2) describes the spectral domain,

$$|X(n)| = \sqrt{\left[Re\left(x(n)\right)\right]^2 + \left[Im\left(x(n)\right)\right]^2}$$

eq. (2)

C. *Warping Filter Banks:*

The warping filter banks are used to fetch the exact data from the extracted features without any loss. The comparison of two filter banks were done with the help of Mel Frequency filter banks and Bark Frequency filter banks.

Step 1: Mel Frequency Filter Bank:

Based on the human perception, the Mel Frequency analysis is preferable. The human ear is very sensitive and it is proved that humans having high resolution to the low resolution frequency rather than a higher frequency. Speech signal does not be linear. To make a linear scale conversion for the frequency using Mel scale is used to warping a signal in frequency domain to the Mel scale. The conversion of speech signal from frequency domain to Mel scale can be done using the following eq. (3)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

$$Mel\,(f) = 2595\log\,(1 + \frac{f}{700})$$

eq. (3)

The Mel Filter bank spacing has to be applied to the FFT values to get the conversion for the frequency domain into the Mel scale. Triangular band pass filters are applied to as a filter bank spacing which has uniformly spaced on the linear frequency axis with the larger number of filters in the low frequency region and lesser number of filters in the high frequency region and is shown in Fig.3.

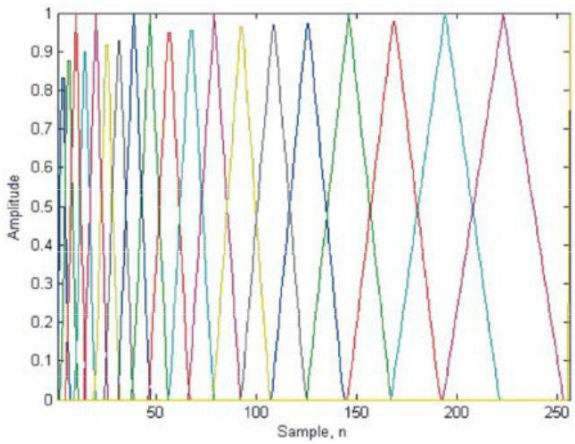

FIG.3. Mel Frequency Filter Bank

Step 2: Bark Frequency Filter Bank:

The bark scale is originally defined in zwicker(1961). A distance of 1 on the bark scale is known as a critical band. The implementation provided in this function is described in Traunmuller (1990). An approximate expression for the Bark scale frequency warping, due to the Schroedinger is used in this proposed method using eq.(4),

$$Bark\,(f) = 6log_{10}\,[\left(\frac{f}{600}\right) + \sqrt{\left(\frac{f}{600}\right)^2 + 1}\,]$$

eq. (4)

The Bark frequency filter ban spacing is applied to the FFT. It is uniform speed to collect more information with the input wave. It is shown in Fig.4.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

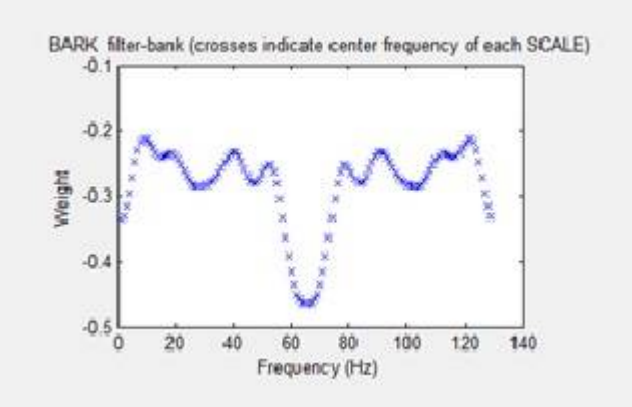

FIG.4. Bark Frequency Filter Bank

D. *Logarithm of Energies:*

To compute the log-energy, i.e., the logarithm of the sum of filtered components for each filter. Eq(5) expresses the computing logarithm, of weighted sum of spectral values in the filter-bank channel. At this stage, the number of rows equal to the number of columns to the number of filters in the filter bank.

$$S(m) = log_{10}[\sum_{n=0}^{N-1}(|x(n)|^2 . Hm(n)), 0 \leq m \leq M$$

eq. (5)

E. *Discrete Cosine Transform:*

The cepstral analysis includes the conversion of spatial domain to frequency domain by applying DCT to the Mel Scale values. DCT expresses a finite set of data points in terms of a sum of cosine functions. The conversion of DCT is similar to the DFT in the conversion process. DCT is more preferable since the value obtained will provide us the accuracy for increasing the robustness of ASR. Eq.(6) expresses the DCT,

$$C(n) = \sum_{m=0}^{M-1} S(m) . \cos(\pi n(m + \frac{1}{2})/M, 0 \leq n \leq N$$

eq. (6)

The cepstral coefficients are obtained by applying DCT to the Mel scale and Bark scale values. The coefficients which are obtained after the evaluation of DCT called Mel Frequency cepstral coefficients (MFCC) & Bark frequency cepstral coefficients (BFCC). When the number of coefficients increases the accuracy and also the increase in the complex of the designer complexity, there is a lag in the design of ASR system. The number of coefficients taken here is 12.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

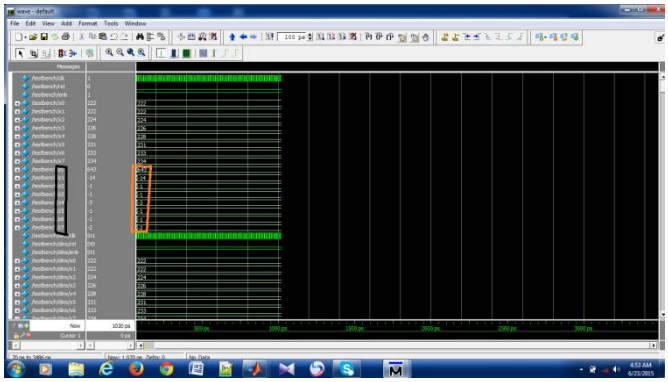

FIG.5. DCT Output

V. DISCRETE HIDDEN MARKOV MODEL

Discrete Hidden Markov Model is used to accelerate the speed of speech recognition. A Codebook is to be first generated with the values obtained from the feature vectors. Those feature vectors can be trained using DHMM in the codebook. From the training samples, the upper and lower bounds of each element has to be calculated to generate the codebook. The range of upper and lower bounds is divided into various sub-intervals from which the feature vectors are extracted. By randomizing the same number of vectors according to the number of classes, the initial codebook has to be formed. The codebook can be initialized with the values obtained from the feature vectors. DHMM id the only classifier based on the probability. This paper utilizes this technique as a comparator based on the probability basis. Since it is a time consuming process, it improves the efficiency of the desired system [7]..

VI. HARDWARE ARCHITECTURE

The desired system can be implemented in the Altera DE2-70 board. The desired system can be evaluated and can be implemented through the System on Chip architecture of FPGA. SoC architecture is used here due to its reconfigurablitiy. The SoC architecture has been used as similar in Fig.6. All the Algorithms of the desired methodology can be implemented through NIOS-II processor. The Altera DE2-70 development board in which the CYCLONE-IV processor is included in this experiment. The Push button is used here for noise suppression at various levels. The toggle switch can be used as an input for the FPGA board. This can be used as an authentication purpose. The microphone can be used as an output for checking the robustness of ASR. The LCD will be used to display the words that the user spoke which is to be converted into text. An audio controller is used to receive the speech signal. The I^2C protocol is used to control the register of the platform [7].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

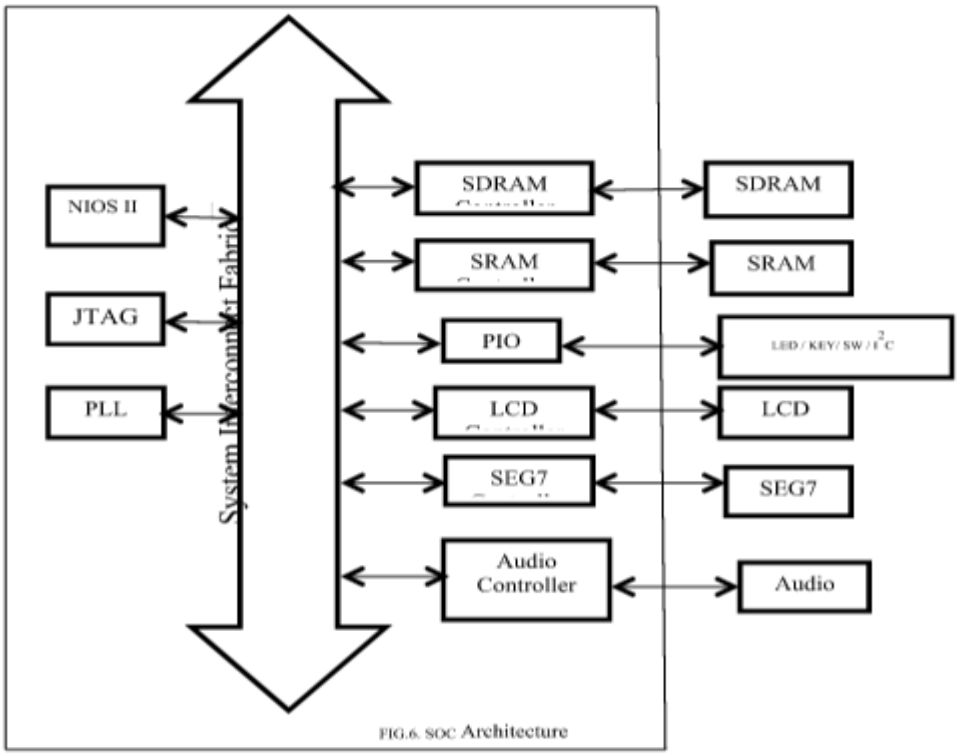

Fig.6. SoC Architecture

VII. RESULTS & DISCUSSIONS

The variousnoise levels were added to the input speech signal has been given to extract the features to train the data. Fig.7 shows that the information collected using MFCC & BFCC for Car Noise such as 0dB, 5 dB &10 dB. It describes that the BFCC retrieves more information than MFCC.So, the loss of information is very less in Car noise. Fig.8 shows that the information collected using MFCC & BFCC for Babble noise for the various noise levels such as 0dB, 5 dB &10 dB.It describes that the BFCC retrieves more information than MFCC.So, the loss of information is very less in Babble noise.

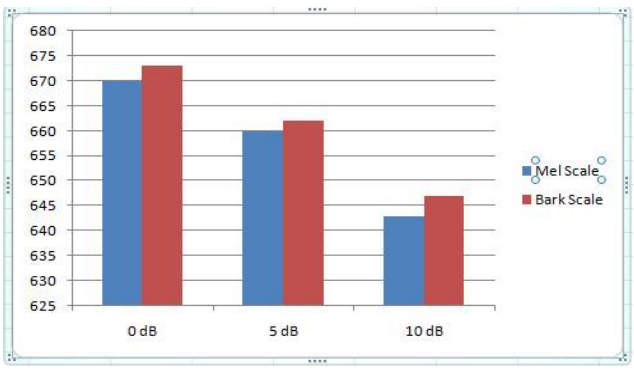

Fig.7. Car Noise

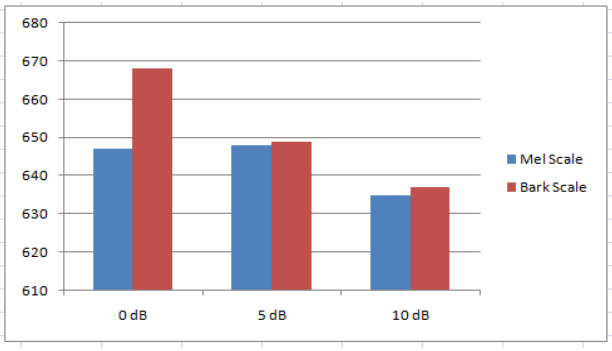

Fig 8. Babble Noise

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

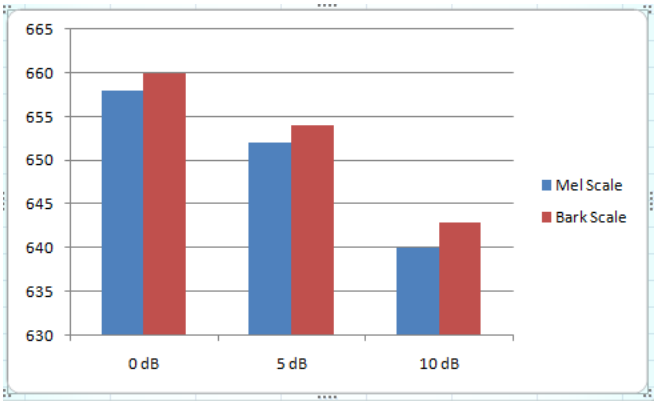

Fig 9. Airport Noise

Fig. 9 shows that the information retrieved using MFCC & BFCC for Airport noise for the various noise levels such as 0dB, 5 dB &10 dB.It describes that the BFCC retrieves more information than MFCC.So, the loss of information is very less in Airport noise. On Comparing with these different noises, BFCC is very suitable with Babble noise tha Car & Airport noise.

VIII. CONCLUSION AND FUTURE WORK

From the results appeared for various levels of various noises. The BFCC method retrieves more information than MFCC. The BFCC method of feature extraction technique is better for Babble noise than Car & Airport noise. The codebook can be initialized with the help of feature vectors are obtained through the training phase by collecting the data from the DCT output. The speech vectors can be randomized and that can be evaluated through the suppression of environmental noises. The speech signal can be processed and that can be compared with the various hidden states using DHMM. When the features vectors are similar and the pattern can be recognized using Matlab 13.1a software. The simulation results can be obtained with the help of Modelsim6.4a and this can be synthesized through the help of QUARTUS II software and which is helped to implement in the hardware realization.

REFERENCES

1. Yuan Mang,'Speech Recognition on DSP : An Algorithm on Optimization & Performance Analysis', The Chinese University of Hong Kong, pp. 1-18, 2004.
2. D. Huggins Daines, M. Kumar, A. Chan, A. Black, M.Ravishankar and A.Rudnicky, 'PocketSphinx: A free real-time continuous speech recognition system for hand-held devices', Proceedings of theICASSP,2006.
3. Rumia Sulthana and Rajesh Palit,'A Survey on Bengali Speech-To-Text Recognition Techniques', 9th International Forum on Strategic Technology, Cox's Bazar, 2014.
4. Muda Lindasalwa , Mumtaj Begum and I.Elamvazhuthi,Voice Recognition Algorithm using MFCC & DTW Techniques',*Journal of*computing, ISSN 2151-9617, 2(3):138-143.
5. AlGabri Malek,Chunlin LI, Z. Yang, Naji Hasan.A.H and X.Zhang,' Improved the Energy of Ad hoc On- Demand Distance Vector Routing Protocol', International Conference on Future Computer Supported Education, Published by Elsevier, IERI, pp. 355-361,2012.
6. D.Shama and A.kush,'GPS Enabled EEnergy Efficient Routing for Manet', International Journal of Computer Networks (IJCN),Vol.3, Issue 3, pp. 159-166,2011.
7. Shilpa jain and Sourabh jain,'Energy Efficient Maximum Lifetime Ad-Hoc Routing (EEMLAR)', international Journal of Computer Networks and Wireless Communications, Vol.2, Issue 4, pp. 450-455, 2012.
8. Vadivel, R and V. Murali Bhaskaran,'Energy Efficient with Secured Reliable Routing Protocol (EESRRP) for Mobile Ad-Hoc Networks', Procedia Technology 4,pp. 703- 707,2012.

BIOGRAPHY

Ms. S.Suganyais currently working as an Assistant Professor in CK College of Engineering & Technology(A Constituent of Anna University, Chennai). She pursued her Masters in Applied Electronics from Saveetha Engineering

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

college, Thandalam (A Constituent of Anna University, Chennai) in 2015. Her research interests in VLSI design, Image Processing, Signal Processing and Embedded systems.

Mr. M. Selvaganapathy is currently working as an assistant professor in CK College of Engineering & Technology, Cuddalore (A Constituent of Anna University, Chennai). He pursued his Masters of Engineering in Mailam Engineering College, Mailam and currently pursuing his Ph.D. Degree in Annamalai University, Chidambaram. His research areas are Digital Image Processing, Neural Networks, Embedded Systems and Wireless Communication / Networks.

Mrs. N. Nishavithri is currently working as an Assistant Professor in Mailam Engineering College, Mailam (A Constituent of Anna University, Chennai). She pursued her Masters of Engineering in Mailam Engineering College, Mailam. Her research interests are Embedded Systems and Wireless Communication / Networks.