International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Speech Based Query Searching Technique and It's Application In Library Management System

S.Shivaprasad, M. Sadanandam

ABSTRACT: Customary library inventories have turned into wasteful and badly designed in assisting library clients. Readers are spending too much time in looking through materials related to library by means of inventory copies. Pursuers need a smart and creative answer to sort out this issue using speech itself. Speech is the most effortless approach to speak with one another. Recently, Speech processing is mostly used in applications like devices used for security purposes, household appliances, cellular/mobile devices, Automatic Teller Machines (ATMs) ,computers and etc. The man machine interface is also created to associate people who are experiencing some kind of disabilities. Speech processing is nothing but the strategy to process and break down the speech signals. It has great advantages for disabled people and discovers their applications in our everyday lives. In this paper, we have connected different data mining models to distinguish the required course book which is available with using speech signals and furthermore findings in a book which is frequently issued to readers. It is very much useful to deaf peoples in society. In this we applied Mel-frequency cepstral Coefficients (MFCC) to extract the features from speech signal of isolated spoken words and applied GMM and DNN models for discovering whether required book is present or not and observed is DNN gives the great accuracy and We are also converting the given speech into text format and apply association rules to find the book that is frequently issued.

Keywords: GMM,DNN, MFCC, voice, Speech processing, HMM, Book searching.

I. INTRODUCTION

A library is a collection of all types of information and similar resources that are selected by experts and is given access to a defined community for reference. It not only provides physical access to the materials, but also the digital access, and also in either physical location or a virtual, or both. A library's collection usually include books, publications, journals, newspapers, films, maps, documents, micro form, CD's, cassettes, videotapes, DVD's, Blu-ray Discs, e-books, audio books, databases of different researches, and other formats. A Library Management System (LMS) is a product work to deal with the essential housekeeping elements of a library. Libraries mainly depend on the executive's frameworks to oversee resource gathering just as collaboration with their individuals.

Manuscript published on 30 September 2019
* Correspondence Author
S.Shivaprasad , Research scholar , Department of CSE, Kakatiya University, Warangal &Assistant professor, Vignan Foundation for Science and Technology, Guntur, A.P, India.
M.Sadanandam* ,Assistant professor, Department of CSE, KU college of Engineering, Warangal ,TS,India.

The executive's frameworks help libraries to monitor the books and their checkout procedures, just as individuals' memberships and profiles. Human to human interaction will be in several ways like facial expressions, eye contact, changing body gesture and mainly through speech. The speech is the important and easy mode of Communication among human beings and is also the effective and efficient form of transferring information.Speech-to-text conversion (STT) system is widely used in many application areas. In the educational field, STT or speech recognition system is the most effective on deaf or blind students. The recognition of speech without any noise is the most challenging and difficult part in speech processing. Speech Recognition is procedure of changing over discourse signal into a succession of words by methods for Algorithm executed as a product program.Now a days, we are utilizing the library to store the information on one server and each time readers enter their subtleties and check in database stockpiling whether the required book is accessible or not. These days, world readers are pulled in with most recent programming and they are not investing much energy in this. Also, this framework is having issues like(1)The data stored is prone to cyber hacks. Opting for a reliable online system eliminates the risk(2)Costly and Expensive(3)Complicated to operate(4)Online Systems require high-speed internet connectivity(5)Risk of computer virus (6)The automation feature is not available in offline/ open source systems thus, requires manual action to perform operations(7)Unlike online systems that utilize cloud computing, Open-source systems store data on computer hard drive. This increases the risk of data loss(8)Blind and deaf people can't able to access the present online book systems as they were unable to type the text or see the text .

Till now, Libraries using speech data base for speech processing research has not been made. In this paper, we not only designed a method to identify whether the required book is present or not, but have also developed the model to identified frequently issued book.

In Section 2 describes literature survey, Section 3 of this paper is proposed system, Section 4 different aspects of the experimental setup and the obtained results are described. The system accuracy is discussed in sub sections and conclusion is mentioned in Section5.

II. LITERATURE SURVEY

Roger W. Christian in 1975 was the first person who brought up the idea about "electronic library". F.W. Lancaster, in his book "Toward Paperless Information Systems (Lancaster,1978)", anticipated that electronic publication might replace paper publication after the year 2000, and those traditional libraries might shift into digital libraries. But the digital libraries have some disadvantages to overcome so that voice based searching can be implemented easily. Su Myat Mon, Hla Myo Tun, has designed a "Speech-To-Text(STT) System Using Hidden Markov Model(HMM)"[3], "implemented Speech to Text by using the MFCC" for extracting highlights and HMM as the recognizer.

In discourse database, fifty sound records are recorded. These are broken down to get highlight vectors. From that point onward, the test expressed word is tended by forward calculation of Hidden Markov Model (HMM). From the re-enactment results, it is unmistakably observed that the normal acknowledgment rate of 87.6% is accomplished by the quantity of states (N=5) is preferred exactness over some other states. However, on the off chance that the number of states is excessively enormous, there are no enough perceptions per state to prepare the model. This work, the presentation of the framework is progressively precise and solid by utilizing end point location calculation in prepossessing stage.

In[6],Na Li, Qiushi Li proposed Knowledge based Management Modes and Strategies for different University Libraries. The information based administration mode idea is to meet the improvement prerequisites of library inside the time. It is a blossoming the board mode to naturalize the most aggressive condition. The learning the executives approach will build up oneself - the board and innovative limit of college libraries. Despite the fact that the information the executives modes and techniques are investigated, an incorporated learning the board plot for college libraries is proposed.

Anita Gade,Yogesh Angal[7] proposed a framework which was tried with ongoing library the executives framework. It disentangled the procedure of structure and furthermore abbreviates the plan cycle in light of LabVIEW programming. The created framework diminishes the human obstruction endeavors with powerful administration of library. In this work, a programmed choice framework is executed to decide the control mode for consecutive undertakings. This framework is intended for control framework, for example, rescheduling, stockpiling and categorisation of inappropriately set books of library progressively applications. The created strategy have upgraded search speed and gives the correct arrangement.

In[5], Yeou-Jiunn Chen, Chung-Hsien Wu, and Gwo-Lang Yan proposed momentous confirmation Using Information for Mandarin Telephone Speech Keyword recognizable proof. they checked a few accomplishments in ceaseless Mandarin discourse keyword distinguishing proof and affirmation. In this framework, 59 setting independent sub syllables are utilized as the fundamental acknowledgment units. A two-organize methodology, with distinguishing proof pursued by adjustment, is grasped. For expression explanation, 12 against sub syllable HMM's, 175 setting subordinate discourse musical HMM's and 5 hostile to cadenced HMM's, are built. A keyword correction capacity joining phonetic-stage and prosodic-stage distinguishing proof is researched. Trial results demonstrate that prosodic data beats the standard framework without discourse musical data.

III. PROPOSED SYSTEM

Any recognition framework system must comprise two stages training and testing. In this, training stage is utilized to manufacture the framework. Meanwhile the testing stage is utilized to check the Frame work and its working during the preparation stage.

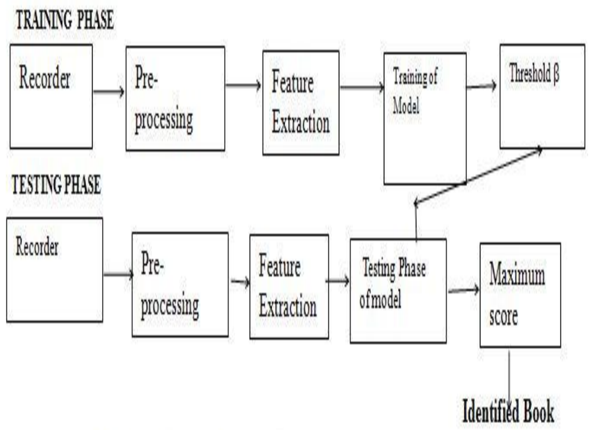


Fig.1 Block diagram of an identification system

Figure 1 gives a thought regarding the preparation and testing times of the present ID framework.

3.1 Development of Speech database

In identification of process of required book developed in 4 different modules. In this

firstly create a data base contains the all the book related information. For the current work, speech data was recorded using the Portable Recorder with 8 kHz sampling frequency. Recording was made like book name followed by author name. We create a data base contains 10 books with each book of 5 different authors. Example for C language we create a directory 'C' and record speech " C language by Balaguruswamy". Likewise we created for Data Structures, Compiler design, Algorithms etc. While recording itself applied prepossessing techniques to remove the noise by using average filtering techniques. The mean or average filtering is a basic sliding-window spatial channel that replaces the inside incentive in the window with the normal (mean) of all the pixel esteems in the window.

The window, or bit, is generally square however can be any shape.Fig.2. Window or mask of size 5 in 1D-average filter

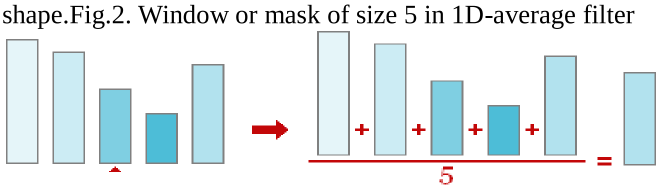
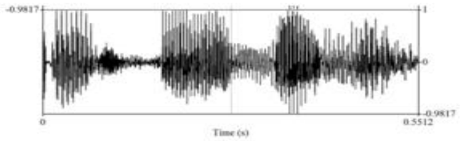


Fig.2. average filter working

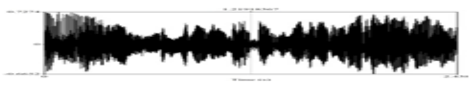
Simulation results of average filter

original signal



(a)

Results of average filter



(b)

Fig3. Working of Average filter (a)Original signal (b) Noise removal signal

2.1 Feature Extraction

The time area waveform of a context sign conveys all conceivable acoustical data from the perspective of phonemics, next to no can be decided based on the waveform. To have the option to locate significant data from approaching information, it is imperative to have a few techniques to decrease the data in each fragment of the sound sign into a nearly modest number of features. Feature extraction was characterized as the way toward rationing the significant data present in the linguistic signal while expelling the undesirable parts. Most of the spectral features of a speech signal are obtained by changing over the time area signal into the recurrence space. To get the mirage highlights, MFCC utilized as the element extraction strategy for both the preparation stage and evaluation period of the current framework.

Mel Frequency Cepstral Coefficient (MFCC) Highlight extraction is the most huge bit of the entire system. The feature extraction is to reduce the data size of the given signal preceding model portrayal or affirmation. The methods for Mel repeat Cepstral Coefficients (MFCCs) calculation are restraining, windowing, Discrete Fourier Transform (DFT), Mel recurrence isolating, numeric/logarithmic work and Discrete Cosine Transform (DCT).Fig.1 appears the square graph of MFCC process.

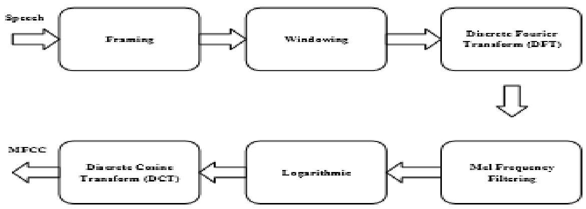


Fig.4 . Block diagram of MFCC

Framing is the initial step of the MFCC. It is the way toward hindering of the discourse tests acquired from the simple speech of the verbally expressed word, into the quantity of edge signal . Covering is expected to maintain a strategic distance from data loss. Windowing is done in request to lessen the interruptions toward the beginning and at the edge to be smooth of the earlier and least focuses in casing and windowing capacity are utilized. Discrete Fourier Transform series (DFT) is utilized as the Fast Fourier Transform (FFT) calculation. It changes over each edge of N tests from the time area into the recurrence space. The figuring is more exact in recurrence space as opposed to in time area. Mel recurrence sifting: The audio sign does not pursue the direct scale and there is a wide recurrence extension in FFT. It is non cognitive scale that recreates the manner in which human audibility works. It relates to the best goals at low frequencies and less at high. Logarithmic limit is the logarithmic change associated with the preeminent size of the coefficients got after Mel-scale change. The by and large degree action discards the stage information, making feature extraction less touchy to speaker subordinate assortments. Discrete cosine change (DCT) changes over the Mel-separated range again into the time region since the Mel Frequency Cepstral Coefficients are utilized as the periodic file in acknowledgment arrange.

2.2 Modelling Technique

In this proposed work two modelling techniques GMM and DNN have been used. Mixed models are examples for density model in which contains the density functions, like Gaussian and the segment functions are merged to bring a multi modal density. Gaussian lives to the exponential family. This family has a comparable range in both frequency as well as time area. GMM is characterized as a parametric Probability Density Function (PDF) which is spoken as a unit weighted entirety of a Gaussian segment density. Criteria of the Gaussian Mixture Model comprise of the mean and difference networks of the Gaussian segments and these loads demonstrating the commitment of each and every Gaussian to the estimation of PDF. The preparation information got after

Text book name	No.of Training Samples of book	Wave files length	No.of testing samples
C	5	2-7ms	3
Data structures and Algorithms	6	5-12ms	4
Compiler design	4	3-8ms	2
Strength of materials	3	2-9ms	2
Data Mining Techniques	5	3-12ms	3
Operating system	5	2-8ms	3

Retrieval Number C4779098319/2019©BEIESP
DOI: 10.35940/ijrte.C4779.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

component extraction are nourished to the preparation of GMM and DNN. Once if the training phase is done, then the weight, variance and mean are uniquely called as parameter β, are achieved for GMM. A separate GMM model is to be developed for each of the essential textbook. For the current system, ten GMM models are built one for each different text book. Not alike GMM only one DNN is developed to get β for all basic books. In the current identification system, the MFCCs of text books are used to train just GMM and DNN. This first β consists of information of all the textbooks being used in the system. For the current system, this β consists information of about 5 testing signals. a definite model for each constituent book is acquired by MAP (Maximum A-Posteriori) altering the trained models to training info of that book.

This gives altered criteria of the book. After MAP alteration, the altered criteria β is achieved for different textbook. For this DNN gives more accuracy in identification compare to GMM.

IV. EXPERIMENTAL SETUP AND OBTAINED RESULTS

We used different audio signals at the time of training and testing**.**A textbook is identified when a test affirmed gets greatest probability score for that specific book. For example in system, when a test asserted gets maximum possibility score for C, it entails that, most frequently searching and issued book is C. This is same for all remaining books also. In this, for identifying the book applied GMM and DNN with different testing samples and also tested the accuracy.

Every speech initially given to average filter to remove noise and unwanted data and sent to MFCC to get required features and apply the different models.

In identifying required book by Speech, GMM is 76.47% accuracy while using DNN it increases the accuracy to 82.35%. The data set used for training and testing given table1 and also the results described in table 2.and Performance of models will be in Fig2.

Tab.1 Data set description used for training and testing.

Text book name	No.of Training Samples of book	Wave files length	No.of testing samples
C	5	2-7ms	3
Data structures and Algorithms	6	5-12ms	4
Compiler design	4	3-8ms	2
Strength of materials	3	2-9ms	2
Data Mining Techniques	5	3-12ms	3
Operating system	5	2-8ms	3

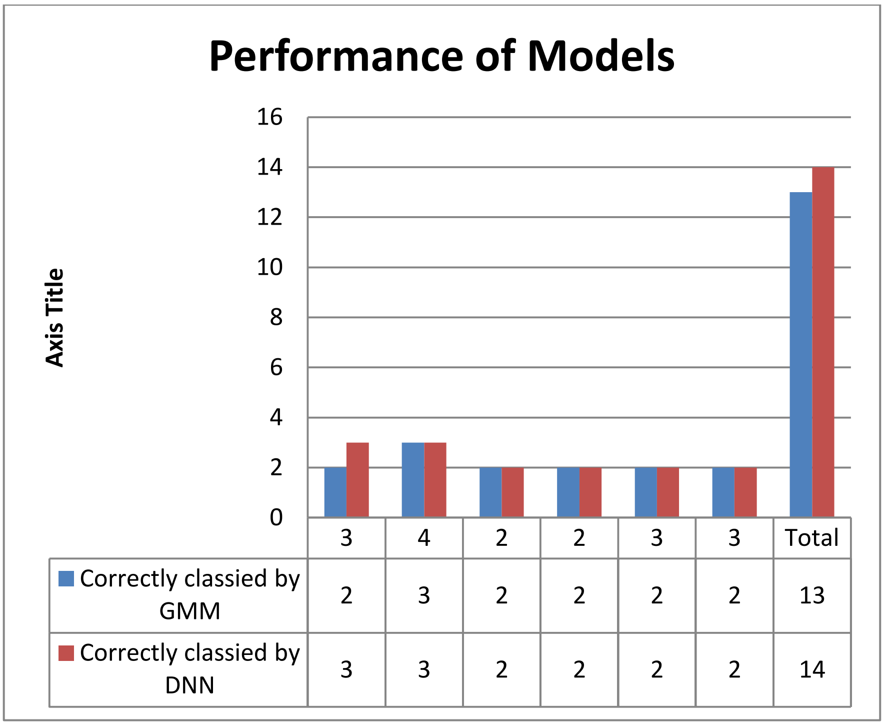


Fig.5. Performance of models

Tab 2.Measuring accuracy of the system

Text book name	Number testing samples	Total no. Of testing samples	Accuracy	
			GMM	DNN
C	3	17	76.47%	82.35 %
Data structures and Algorithms	4			
Compiler design	2			
Strength of materials	2			
Data Mining Techniques	3			
Operating system	3			

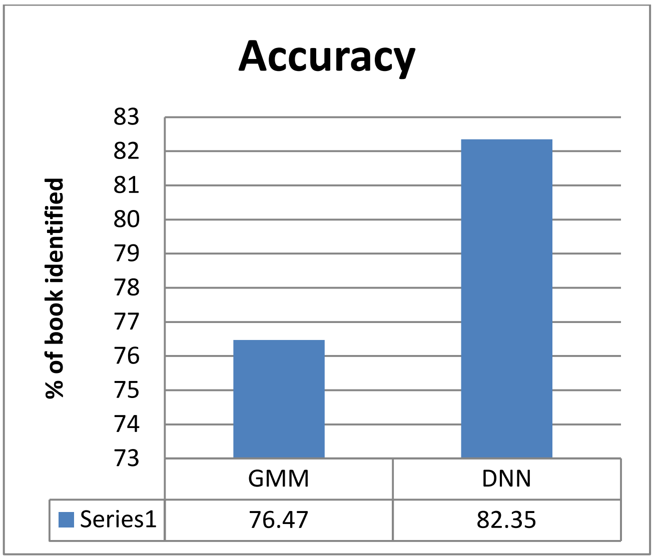


Fig.6. Accuracy percentage of Models

2.2.1 Identifying frequent issued book

In order to find out frequent issued book , we have to convert given speech in to text(STT) . For this one we use MFCC and HMM model. Firstly testing sample utterance is recorded and compared with data base speech corpus and if it is correctly identified by GMM or DNN, then we take that sample as input to HMM model and we convert that speech sample to text. This data is stored in excel data sheet and the count of a book that indicates number of times book searched.

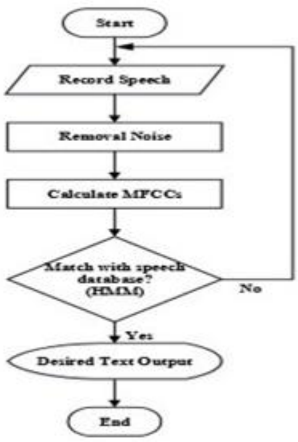


Fig. 7. Flow chart diagram of speech to text conversion

To transfer input audio to text output, mainly four steps are developed by using PYTHON. The steps are audio database, pre-processing, feature derivation (extraction) and identification. Firstly, audio file which is recognized by model taken as input and the congestion of surroundings, the unnecessary noise may affect the identification rate worse. The end point identification strategy can overwhelm this issue. In the wake of pre-processing stage is done, the discourse tests are separated to highlights or coefficients by using the Mel Frequency Cepstral Coefficient (MFCC). At long last, these MFCC coefficients utilize the contribution of Hidden Markov Model (HMM) recognizer to group the ideal verbally expressed word. The ideal content yield can be created by HMM technique is going to store in document. From that exceed expectations document we are ascertaining the most frequent issued book.

Fig.Fig.8. Details of number of times search for a book

V. CONCLUSION

In this paper implemented Vocal based book searching system. It will helps visually challenged people to access their required books searched by their voice aptly. It's been observed that almost 40% complete deaf population all over the world is in India itself. This system overcomes complications faced by blind or partially blind people as well as unlettered people. Therefore, a voice based books (library) can predict a reader's necessity in advance, rely upon the mining results.

We apply GMM and DNN model to search for required book and we find that DNN provides great accuracy compare to GMM. And also we find the frequent issued book by applying HMM model and frequent rules. In future the data base size may also increase and many mining algorithms we can apply.

REFERENCES

1. Matthew Nicholas Stuttle "A Gaussian Mixture Model Spectral Representation for Speech Recognition" Dissertation , July-03
2. Andrew L. Maas *, Peng Qi, Ziang Xie, Awni Y. Hannun, Christopher T. Lengerich, Daniel Jurafsky, Andrew Y. Ng "Building DNN acoustic models for large vocabulary speech recognition" Computer Speech and Language 41 (2017) 195–213
3. Su Myat Mon, Hla Myo Tun" Speech-To-Text Conversion (STT) System Using Hidden Markov Model (HMM)" International Journal of Scientific & Technology Research Volume 4, Issue 06, June 2015.
4. M. Sadanandam, Dr. V. Kamakshi Prasad, Dr. V. Janaki "AUTOMATIC LANGUAGE RECOGNITION SYSTEM USING NEW FEATURES AND THEIR WEIGHTAGE", Recent Science: International Journal of Advanced Computing, Vol.35 Issue.7 380 © RSIJAC 2012.
5. Yeou-Jiunn Chen, Chung-Hsien Wu, and Gwo-Lang Yan **"**Utterance verification using prosodic information for Mandarin telephone speech keyword spotting" ICASSP, IEEE Internationa;l Conference on Acoustics,Speech and Signal Processing 2:697-700 · March 1999.
6. Na Li and Qiushi Li "Knowledge Management Modes and Strategies for University Libraries" 201O International Conference on Future Information Technology and ManagementEngineering.
7. Anita Gade , Yogesh Angal "Development of Library Management Robotic System" 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI) Zeal Education Society, Pune, India, Feb 24-26, 2017.
8. M. Sadanandam, Dr. V. Kamakshi Prasad,Dr. Janaki, A. Nagesh "Text Independent Language Recognition System Using DHMM with new Features", IEEE –Int. Conf. Signal Processing,2012, Beijing, China.
9. M. Sadanandam and Dr. V. Kamakshi Prasad, "Robust features for GMM Based Language Identification System", International Journal of Speech Technology: Volume 17, Issue 2 (2014), Page 99-105, Springer
10. S. Nakagawa, Y. Ueda, and T. Seino, ''Speaker-independent, text-independent language identification by HMM,'' in Proc. Int. Conf. Spoken Lang. Process., Banff, AB, Canada, 1992, pp. 1011–1014.
11. J. Navratil, ''Spoken language recognition A step toward multilinguality in speech processing,'' IEEE Trans. Speech Audio Process., vol. 9, no. 6, pp. 678–685, Sep. 2001.
12. IITG Multivariability SpeakerRecognition Database A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. Journal Royal Statist.Soc. Ser. B. (methodological), 39:1 38, 1977.
13. D.B. Paul, 'Speech Recognition Using Hidden Markov Models', The Lincoln Laboratory Journal, Volume 3, Number 1 (l990).
14. Mathew Magimai Doss,'Using Auxiliary Sources Of Knowledge For Automatic Speech Recognition', Computer Science and Engineering ,2005.
15. Ying Chen and Zhenmin Tang, "Speaker Recognition of Noisy Short Utterance Based on Speech Frame Quality Discrimination and Three-stage Classification Model", International Journal of Control and Automation, SERSC Australia, ISSN: 2005-4297 (Print); 2207-6387 (Online), vol. 8, no. 3, March 2015, pp. 135-146.

AUTHORS PROFILE

Mr. S Shivaprasad is a research scholar in department of CSE, Kakatiya University, Warangal, TS, India,
Presently working as assistant professor in department of CSE,VFSTR (deemed to be)University,Guntur. He is IBM Certified Specialist Rational Team Convert V3 and published 11 papers in peer reviewed IEEE and Scopus journals and conferences. His research focuses on the Speech Recognition and Processing, Image processing and Pattern recognition. He is a fellow of the IEEE,IJCTT.

Mr. M. Sadanandam received the PhD degree from JNTUH, Hyderabad in 2013. He is currently working at the Department of Computer Science &Engineering, Kakatiya University, Warangal. His research focuses on the Speech Recognition and Processing, Image processing and Pattern recognition. He has published more than **25** papers in peer reviewed IEEE and ACM journals and conferences. He has served as HOD of Computer Science and Engineering from Nov-2013 to Sept 2015, *BOS chairman for CSE &IT during 2016*-18, Joint Director for KU Engineering Hostels and Placement officer for University College of Engineering, Kothagudem. He is a fellow of the IEEE, Fellow of Institute of Engineers (FIE), ISTE(Indian Society for Technical Education) , Member CSI, IAENG.