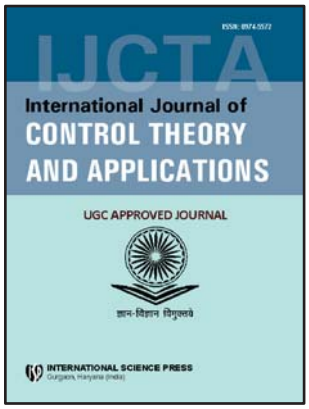

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 31 • 2017

Review of Speech Recognition on South Indian Dravidian Languages

Prathibha Sudhakaran[a] Ratnadeep Roy[b] and Archek Praveen Kumar[c]

[a]*Assistant professor, Xavier institute of engineering Mahim Mumbai, India*

E-mail: prathibha.s@xavierengg.com

Assistant professor, Amity University Rajasthan Jaipur, India

E-mail: rroy@jpr.amity.edu

Assistant professor, ECE Amity University Rajasthan Jaipur, India

E-mail: archekpraveen@gmail.com

Abstract: Speech is a one-dimensional quasi non-stationary time varying signal produced by a sequence of sounds. Speech signals are random in nature. Speech signals are easily corrupted by noise so recognition is an important role in speech processing. Speech is a general way of communication. Speech recognition systems are speaker dependent and speaker independent. Speech recognition is an important task for the interaction between human and machine. Many researches designed recognition system with challenging parameters. Speech recognition is classified in to 4 types speech database, preprocessing, feature extraction and feature classification. Speech database is created by recording the speech in silent environment. Preprocessing includes framing, de-noising, filtering etc., done by DWT etc. MFCC, LPC, RSTA etc., are some techniques used to extract the features. Pattern recognition, vector analysis and artificial networks (ANN) are some of the classification areas. This paper produces a comparative review of speech recognition for south Indian languages using various techniques with its recognition accuracy.

Keywords: *Speech recognition, South Indian language, Feature extraction, Feature classification.*

Nomenclature : **DWT:** Forward error correction.
AC: Arithmetic Coding
ANN: Artificial neural network.
ZCPA: Zero Crossing Peak amplitude
LPC: Linear Prediction Coding
MFCC: Mel Frequency Cepstral Coefficients
AC-MFCC: Arithmetic Coding Mel Frequency Cepstral Coefficients.

1. INTRODUCTION

Automatic Speech Processing is an important part of information and communication technology today. According to the development in technology most of the highly efficient companies use Robots and human machine interacting systems.

Biometric, security, mobile, healthcare, video games, weather forecasting, transcription etc., are the wide area applications where speech recognition is used. It has been an interesting and profound topic over decades.[1] Scientists and Researchers are formulating many algorithms and techniques for accurate error free systems. It is indeed a great challenge to make a computer to understand spoken language since a word cannot be repeatedly spoken by the same speaker with same slang, pitch and parameters. Vowels, semi-vowels, nasal consonants, unvoiced fricatives, voiced fricatives , voiced and unvoiced stops, diphthongs are the diverse sounds in speech. Every language has its own set of alphabets, vowels and consonant. Pattern recognition, Acoustic phonetic and Artificial intelligence are three main approached in speech recognition. Figure 1 shows different types of speech recognition systems.[2] Speech recognition systems are classified in to 3 types speech utterance, speaker model and vocabulary.

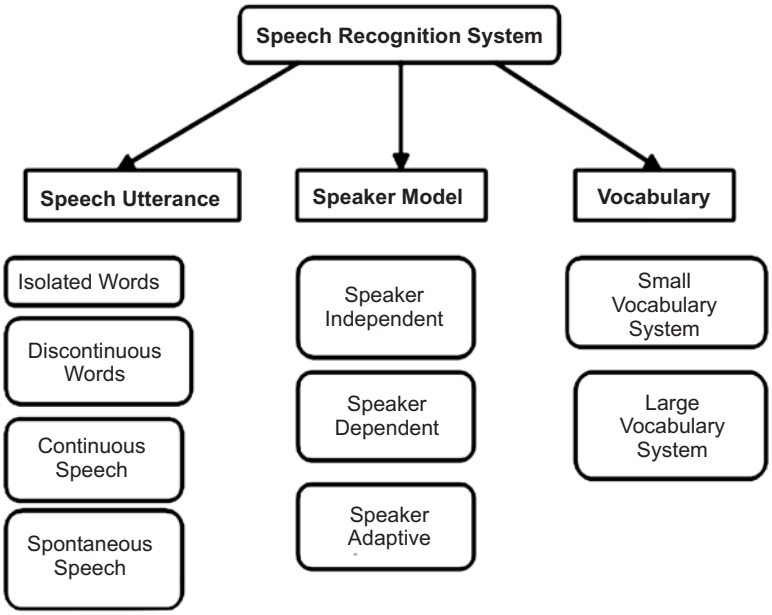

Figure 1: Speech Recognition Systems

There are various processes involved in formulating a speech processing system which includes the following steps as shown in figure 2. The detailed process includes, firstly selecting the language then speech data base is created by recording. The recorded speech is preprocessed for sampling, amplification and filtration. Pre-processed signal is framed and windowed. In pre-processing mainly there are two techniques end-point detection and wavelet denoising [25,32] Later smoothening, softening completes the front end of processing. After this entire feature extraction is done by various techniques as given in table 1. Lastly extracted features are classified by various techniques. [3]

2. LITERATURE SURVEY

2.1. Dravidian Languages – South India

Malayalam, Telugu, Tamil and Kannada are the four Dravidian linguistic classical languages of South India. The Malayalam Language has 52 alphabets of which 15 are vowels and 37 are consonants. Tamil has 12 vowels and 25 consonants. Kannada has 13 vowels and 34 consonants and Telugu language has 60 symbols where 16 are vowels vowel modifiers are 3 and 41 consonants.[9] Cini Kurian et.al, database consists of continuous set of digits in Malayalam Language. MFCC (Mel frequency Cepstrum Coefficient) is used for feature extraction and HMM (Hidden Markov Model) for recognition purpose. Database created involves voices of 21 male and

female in the age group of 20-40 years. Accuracy achieved in this paper is 98.5% for word recognition and 94.8% for sentence recognition. In this paper they were able to recognize any combination of digits pronounced properly without any pause. It showed satisfactory accuracy. Percentage accuracy is 95.7%. The author affirms that future work can be done using the above techniques on larger databases including large number of speakers of various age groups and having different accents. [4]. Authors Sonia Sunny et .al, showed The database consists of vowels of the Malayalam language. A composite design of Daubeches wavelet and ANN has yield a very good performance according to the author. High frequency resolution and low time resolution are some of the features of the wavelet employed. [23, 24] Features are extracted by using Discrete Wavelet Transforms (DWT). Three vowels show 100% accuracy and and an agreeable accuracy has been achieved with the other vowels as well using ANN (Artificial Neural Networks). ANN is excellent due to its features like Adaptive Learning, Robustness , Fault tolerance and Parallel organization .[27].An overall accuracy of 95% has been achieved by this technique[.5]. Wavelet transforms are used in various research areas which include image or signal due to their multi resolution and localization properties. [33]

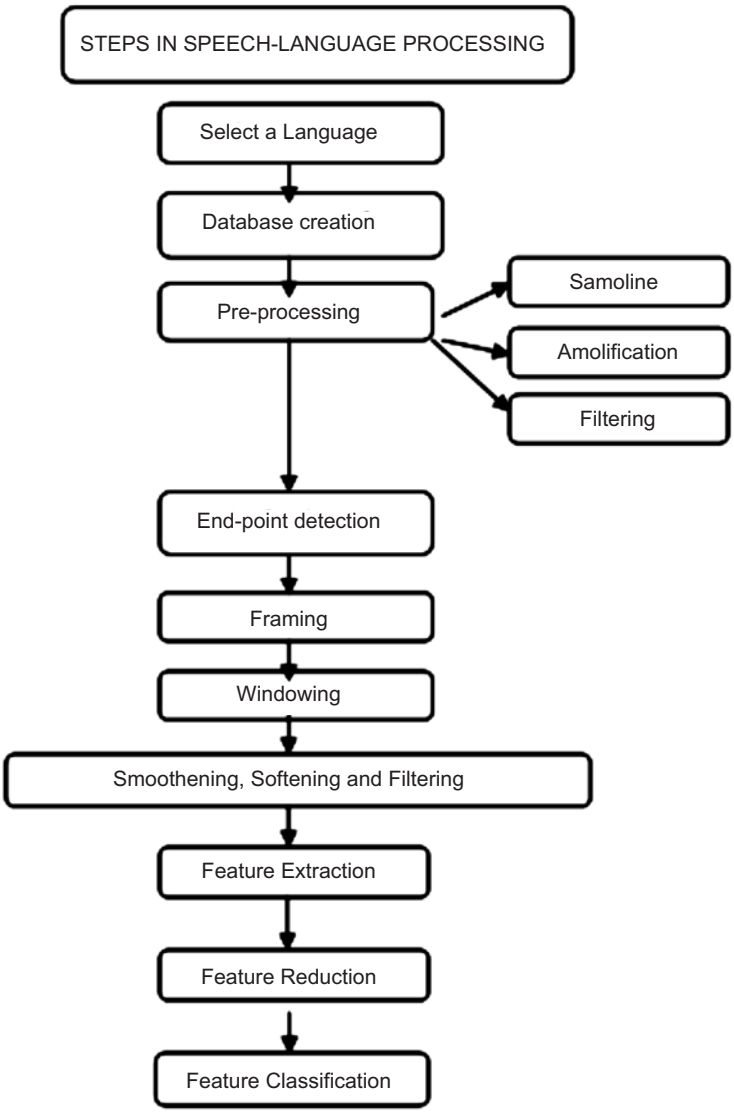

Figure 2: Steps in Speech – Language Processing

Researchers Anu V Anand et.al identified database is 30hours of speech data in Malayalam language based on Hidden Markov Model .Feature extraction has done using MFCC method (Mel-frequency Cepstral Coefficient).The accuracy of the system has been found to be 80%. In this paper a highly efficient machine has been proposed for semi-blind or completely blind persons. They have created a user friendly system for the visually impaired. It is an integrated product of TTS and Open Office Writer version 2.4. They have implemented only for basic commands for making of a document. Further research can be done by including all the commands of Open Office Writer. [6]

Extreme work is presented by Sreejith C et.al. The database is created by 100 speakers and six words in Malayalam. The words are stored in a database and later identified. This involves feature extraction using K-clustering and MFCC .A real time speaker independent recognition system has been proposed in this paper. Quantization distance is computed between features of each work and individual words in the training and testing period respectively. The author points out that future realizable task lies in creating a large continuous vocabulaory system. [7]. Another interesting research has been done by Karpagavalli S et.al ,This paper throws light on the fact that research in speech can benefit the illiterate /semi-literate in rural areas. The performance of any ASR system is highly affected by even small margins of noise. The factors such as resonance , robustness to noise and transducer characteristics are active issues still to be resolved . Many feature extraction techniques have been discussed such as PCA, LDA, ICA, LPC , MFCC, CMS. Many auditory –based feature extraction methods such as zero crossing peak amplitude (ZCPA) , average localized synchrony detection (ALSD), perceptual minimum variance distortionless response (PMVDR) , power – normalized cepstral coefficients (PNCC) , invariant integration features (IIF) , sparse auditory reproducing kernel (SPARK) have been examined effectively.

Researchers Archek Praveen Kumar et.al proved some efficient work for recognizing Telugu speech. In this paper compression and feature extraction has been performed on Telugu Language using AC-MFCC (Arithmetic Coding Mel-Frequency Cepstrum Coefficients). The results have been compared in tabular form with ADPCM, LD-CELP, CS-AELP, CELP, LPC techniques. The extracted parameters include LSP, Pitch prediction filter, code base indexes. MATLAB software has been used for the analysis. 15 male and 15 female speakers each were uttering 10 words at a sampling rate of 8 kHz in the IPA format. Recognition Accuracy obtained is 88.64%. Researcher Archek Praveen Kumar et.al, again recognized Telugu speech by other techniques by improvising accuracy. In this paper feature extraction using MODGDF and MFCC techniques and Naïve Bayes classifiers have been implemented for 100 speakers each uttering 10 words in Telugu Language. Pitch prediction filter, line spectrum, code base indexes, gain, synchronization, forward error correction have been extracted. Naïve Bayes are simple to implement and interpret , have independent attributes and are fast and efficient in the training procedure. [30] Accuracy of 93.76% has been obtained. Sampling rate of 16 kHz is used and two frame sizes 144bits/frame and 80bits/frame have been used. [9][10]. Scientists Dr.V.Radha et.al, designed In this paper , initial pre-processing is done by using four types pre-emphasis , median , average and Butterworth band stop filters and windowing has been performed.[11] LPCC is used for feature extraction and feed-forward neural networks for classification of Tamil spoken words. MSE and PSNR are used as performance measures. Author Hanitha Gnanathesigar use Corpus, which is a collection of spoken or wriiten text which can be understood by machines, thereby increasing the authenticity of the research conducted. It was conducted using CMU's Sphinx Train acoustic trainer model. Accuracy was measures using Pocket Sphinx. Trained Corpus achieved accuracy of 99.1% compared to test corpus of 53.9%. The sampling rate is 16 kHz. The training of acoustic models with Sphinxtrain has been explained in detail for the recognition of Tamil Speech using semi-continuous models.[12]

Researcher P.Ishwarya et.al, gives us a comparative analysis between LPCC and MFCC conducted on 10 isolated Tamil words. 10 words have been repeated 10 times by 3 speakers so in all 300 utterances. Accuracy of 97% was achieved using MFCC. A detailed mathematical explanation with regard to both the methods has been clearly stated. Classification has been performed using PNN (Probabilistic Neural Networks). Error ranging from 3% to 29% has been observed. Scientist[13] Dr.E.Chandra et.al discussed Continuous Speech Recognition for Tamil Language. It uses MFCC for feature extraction. The classification is done using a combination of EWTLBO (Enhanced weighted Teaching – Learning Based Optimization). The weighted is introduced to increase the convergence rate. MFCC system acquired 100% accuracy and the testing time produced 95.26% accuracy. The execution of the proposed system has been measured using FAR (Fakse Acceptance Rate) and FRR (False Rejection Rate). This method can be further applied to continuous speech. [14]. Again authors C.sivaranjan et.al, has proposed a system for continuous speech recognition in Tamil language. Segmentation performed using Viterbi Algorithm. MFCC has been used for feature extraction and HMM for classification purpose. Accuracy of 95% achieved for speaker identification and 98% for speech recognition. Speech data was recorded for 6 minutes for 20 speakers. [15] Additional to this authors M.A.Anusuya et.al, worked on new silence removal algorithm is being discussed in Kannada Language. PRATT software has been used to acquire the speech signal. Error recognition rate has been improvised and decreased from 2.59 to 1.56 by VQ1 clustering algorithm and from 2.5 to 1.45 by VQ2 algorithm. Vector quantization has been used for the purpose of clustering.SVM have features like ease of training, capacity for large attributes,high accuracy and flexibility .[28,29] Accuracy has been tabulated for both speaker independent and speaker dependent systems. MFCC has been used for the feature extraction process. [16]. Researchers Prashanth Kannasaguli et.al, have worked on an Automatic Phoneme Recognition System which uses GMM (Gaussian Mixture Modeling). Feature extraction technique used is MFCC for 15 Kannada phonemes were recorded 500 times during training and 200 times in the testing phase. Phonemes error rate (PER) is implemented for performance analysis of models (5% to 30%). They have included 7500 phonemes and 3000 phonemes in the training and testing database respectively . Classification has been performed using HMM. The GMM uses MAP (Maximum a Posteriori). [17] Phoneme is a basic unit of speech in which speech phonemes are obtained and then processed. [20] Continuation to the work Sarika Hegde et.al, researched that Kananda is an alpha syllabify language where each alphabet has a syllable like structure. It contains 13vowels and 34 consonants. In this paper they have considered 5 vowels and 10 consonants. For feature extraction MFCC and LPC have been implemented and a combination of two classifiers SVM (Support Vector Machine) and HMM (Hidden Markov Model) has been used to improve the efficiency of existing systems. [17]Lastly Sharada C Sajjan et.al, proved in their paper about creating a database of 943 different words in Kannada language and 1753 internal Triphones. Kannada has 46 phonemes. GMM and HMM techniques have been observed to have increased the overall recognition accuracy. Viterbi Algorithm is used to decode the test data. HMM is a statistical model in the Markov process with not known parameters. [31]Comparison of sentences, words and tied state Triphone systems for single Gaussian HMM is implemented. [18]. The main purpose of any efficient speech recognition system is to extract the speech sounds and match it to the input signal. [2,21] Development of an ASR system is a complex and tedious task as it is based on various challenges with regard to channel ,speaker, style of speaking which varies from person to person , regional differences in pronunciation , background noise , speed of speech , pitch and phonetic identity. [22,35]The pattern classifier and the feature vector set play an important role in the recognition accuracy of an ASR system. [26] Speech enhancement is another important factor that has attracted many researchers as removing noise from a signal is always a main concern for developing a robust speech recognition system. It is great challenge as the properties of the original signal needs to be retained. [34]

Table 1
Signal Preprocessing Parameters

S. No	*Authors*	*Year*	*Language*	*Extraction Technique*	*Classification Technique*	*Category of Tokens*	*Accuracy*
1.	Cini Kurian and Kanan Balakrishnan	2009	Malayalam	MFCC, HMM	ANN	Continuous speech	word recognition-98.5%
2.	SoniaSunny, David P.	2010	Malayalam	Daubechies DWT	ANN	Vowels in Malayalam	95%
3.	Anu.V. Anand, P. Shobana Devi, Jose Stephen, Bhadran VK	2012	Malayalam	MFCC	ANN	Spoken words	80%
4.	Sreejith C, Reghuraj PC	2012	Malayalam	MFCC, K-Clustering	ANN	isolated spoken words	88%
5.	Archek Praveen Kumar, Neeraj Kumar	2016	Telegu	AC-MFCC	ANN	10 words are spoken by 15 male voices and 15 female voices (Sampling rate is 8kHz)	88.64%
6.	Archek Praveen Kumar, Ratnadeep Roy ,Sanyog Rawat.	2016	Telegu	MODGDF and MFCC	Naïve Bayes	100 male voices and 100 female voice – 10 words are spoken (Sampling rate is 16kHz)	93.76%
7.	Dr.V. Radha, Vimala.C, M.Krishnaveni	2011	Tamil	LPCC	Feed-forward Neural Networks	Isolated words	NA
8.	Hanitha Gnanathesigar	2012	Tamil	(CMU)'s Carnegie Mellon University's SphinxTrain Acoustic Model Trainer	Pocket Sphinx.	Speech corpus of 37 Tamil phones	Trained corpus-99.1% Test corpus-53.9%
9.	P.Aishwarya and V Radha	2013	Tamil	LPCC and MFCC	PNN (Probabilistic Neural network)	10 words + 10 repetetions by 3 persons so in all 300 utterances	MFCC – 97% LPCC-82.3%

S. No	*Authors*	*Year*	*Language*	*Extraction Technique*	*Classification Technique*	*Category of Tokens*	*Accuracy*
10.	Dr.E.Chandra , S.Sujiya	2014	Tamil	MFCC, LPCC, LPC	EWTLBO and HMM.	Audio clip	96.26%
11.	C.Sivaranjani, B. Bharathi	2016	Tamil	MFCC	HMM	Isolated words from 20 persons	NA
12.	M.A.Anusuya and S.K.Katti	2012	Kannada	LPC,MFCC.	SVM	100 signals -10 words repeated 10 times. Speaker independent	NA
13.	Prashanth Kannadaguli , Vidya Bhat	2014	Kannada	MFCC	GMM	Training database- 755 phonemes Testing database- 3000 phonemes	NA
14.	Sarika Hegde · K. K. Achary · Surendra Shetty	2014	Kannada	LPC and MFCC	SVM , HMM	5 vowels and 10 consonants	MFCC + SVM = 65.8% MFCC + HMM = 66.33%
15	Sharada C. Sajjan, Vijaya C	2016	Kannada	MFCC	GMM, HMM	943 different words, 1753 word internal triphones	NA

3. RESEARCH COMPARASION

These are the main papers which are reviewed and most relevant to the south Indian language speech recognition. All these papers worked on different suitable techniques for the relevant language. Some papers used ZCR (Zero crossing rate), DWT (Discrete wavelet transformation), DCT (Discrete cosine Transformation) etc., for preprocessing. Some papers use different techniques for feature extraction. MFCC, LPC, LPCC, RSTA, AC-MFCC, MODGDF, ZCPA etc., are used according to the database created. Feature classification can be done in three areas called as pattern recognition, vector analysis and artificial neural network.[2] The detailed description of Authors name, published year, chosen language, extraction techniques used, classification techniques used, category of tokens and recognition accuracy is shown in table 2. Recognition accuracy is the major factor to be considered for a perfect recognition system. Every researcher need to work on various parameters like accuracy, efficiency, speed, bit rate, word error rate, gain, code base indices, forward error correction, pitch, synchronization, stability. All the recognition systems deal with these parameters.

4. CONCLUSION AND DISCUSSION

Native language speech recognition is a complex frame work and has a huge vocabulary requirement. We are currently working to develop a novel model of Speech recognition in Malayalam language with the highest accuracy and least error. An attempt has been made through this paper to review the various methods used in developing tools for speech recognition in south Indian languages applied to words, sentences, continuous speech with various sampling rates.

REFERENCES

[1] Sonia Sunny,David P,K Paulose Jacob, " Performance of different Classifiers in Speech Recognition", International Journal of Research in Engineering and Technology ISSN: 2319-1163.

[2] Lawrence R. Rabiner, and Ronald W. Schafer, "Introduction to Digital Speech Processing," Foundations and Trends in Signal Processing, vol. 1, nos. 1–2, pp. 1-194, Jan. 2007

[3] Sonia Sunny (2013), "A hybrid architecture for recognising speech signals in Malayalam",http://hdl.handle.net/10603/25496

[4] Cini Kurian, Kannan Balakrishnan, Speech Recognition of Malayalam Numbers,IEEE,2009.

[5] Sonia Sunny,David P,K Paulose Jacob, "Recognition of Speech signals "An experimental comparison of Linear Predictive Coding and Discrete Wavelet Transforms''. International Journal of Engineering Science and Technology (IJEST).

[6] Anu V Anand, P.Shobana Devi, Jose Stephen, Bhadran V K, "Malayalam Speech Recognition System and Its Application for visually impaired people",IEEE 2012

[7] Sreejith C , Reghuraj P C ," Isolated Spoken Word Identification in Malayalam using Mel-frequency Cepstral Coefficients and K-means clustering", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064. Volume 1 Issue 3, December 2012.

[8] Karpagavalli S and Chandra E, A Review on Automatic Speech Recognition Architecture and Approaches, International Journal of Signal Processing, Image Processing and Pattern Recognition , Vol.9, No.4, (2016), pp.393- 404, http://dx.doi.org/10.14257/ijsip.2016.9.4.34.

[9] Archek Praveen Kumar, Neeraj Kumar, Chereku Sandesh Kumar, Ashwani Yadav, Abhay Sharma,"Speech Recognition using Arithmetic Coding and MFCC for Teugu Language",978-9-3805-4421-2/16/$31.00©2016IEEE.

[10] Archek Praveen Kumar ,Ratnadeep Roy, Sanyog Rawat,Achyut Sharma, Amit Chaurasia , "Telugu Speech Features Extraction by MODGDF and MFCC using Naïve Bayes Classifier", IJCTA,9(21),2016,pp.97-104.

[11] Dr.V.Radha1, Vimala.C2, M.Krishnaveni,"Isolated word recognition system for tamil spoken language using back propagation neural network based on LPCC features", Computer Science & Engineering: An International Journal (CSEIJ), Vol.1, No.4, October 2011.

[12] Hanitha Gnanthesigar,"Tamil Speech recognition using semi continuous models", International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2012 1 ISSN 2250-3153.

[13] P.Iswarya and V.Radha,"Comparative Ananlysis of Feature Extraction Techniques in Tamil Speech Recognition Systems",Proceedings of International conference on "Emerging research in Computing ,Information ,Communication & Applications" ERCICA 2013 ISBN:9789351071020.

[14] Dr.E.Chandra, S.Sujiya,"Tamil Speech Recognition using Hybrid technique of EWTLBO and HMM", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6664-6669,ISSN:0975-9646.

[15] C.Sivaranjani, B. Bharathi,"Syllable based continuous speech recognition for Tamil Language", Bharathi et al., International Journal of Advanced Engineering Technology E-ISSN 0976-3945.

[16] M.A.Anusuya , S.K.Katti, "Speaker Independent Kannada Speech Recognition using Vector Quantization", Proceedings published by International Journal of Computer Applications® (IJCA)ISSN: 0975 – 8887. 7-8 April, 2012.

[17] Prashanth Kannadaguli, Vidya Bhat,"Phoneme modelling for speech recognition kannada using gaussian mixture model ",ISRASE First International Conference on Recent Advances in Science & Engineering -2014 (ISRASE-2014)2014-15

[18] Sarika Hegde · K. K. Achary · Surendra Shetty," Statistical analysis of Feature and classification of alphasyllabary sounds in Kannada Language", International Journal of Spoeech Technology, DOI 10.1007/s10772-014-9250-8.

[19] Sharada C. Sajjan, Vijaya C,"Continuous Speech Recognition of Kannada Language using Triphone modelling". IEEE WiSPNET 2016 conference.

[20] Namrata Dave, " Feature Extraction Methods LPC ,PLP and MFCC in Speech Recognition ",International Journal for Advance Research in Engineering and Technology ,www.ijaret.org, Volume1,Issue VI,July 2013.

[21] Samudravijaya K, "Speech and Speaker recognition : a tutorial , " in Proc. International Workshop on Technology Development in Indian Languages , Kolkata , Jan 2003.

[22] Markus Forsberg .(2003).Why is Speech Recognition Difficult ? [Online].Available:www.speech.kth.se/~rolf/gslt_papers/MarkusForsberg.pdf.

[23] Robi Polikar , " The story of Wavelets ," in Proc.IMACS/IEEE CSCC,1999,pp.5481-5486.

[24] George Tzanetakis, Georg Essl, and Perry Cook, " Audio Analysis using Discrete Wavelet Transform", in Proc.WSES International Conference , Acoustics and Music : Theory and Applications , 2001, pp.318-323.

[25] Kapil Sharma,H.P.Sinha and R.K.Aggarwal,"Comparative Study of Speech Recognition System Using Various Feature Extraction Techniques ," International Journal of Information Technology and Knowledge Management, vol.3,no.2pp.695-698,Jul-Dec.2010.

[26] Malaly Kumar, R.K.Aggarwal, Gaurav Leekha , and Yogesh Kumar "Ensemble Feature Extraction Modules for Improved Hindi Speech Recognition System , " IJCSI International Journal of Computer Science Issues, vol.9,issue 3,no.1,pp.175-181, May 2012.

[27] S.N.Sivanandam, S.Sumathi and S.N.Deepa , Introduction to Neural Networks using Matlab 6.0 , New delhi , India : Tata McGraw-Hill,2006.

[28] V.N.Vapnik,Statistical Learning Theory ,New York, USA: J.Wiley, 1998.

[29] N.Christianini, and J.Shawe-Taylor , An Introduction to Support Vector Machines , Cambridge , UK: Cambridge University Press,2000.

[30] Laszlo Toth , Andras Kocsor and Janos Csirik , "On Naïve Bayes in Speech Recognition ," International Journal of Applied Mathematics and Computer Science , vol.15,no.2,pp.287-294,Jun.2005.

[31] L.R.Rabiner and B.H.Juang , fundamentals of Speech Recognition , New Jersey , USA : Eagle –wood Cliffs Publisher, 1993.

[32] D.L.Donoho,"Denoising by soft thresholding ", IEEE Transactions on Information Theory , Vol.48,PP.927-940,1995.

[33] Mohammed Bahoura, Jean Rouat.Wavelet –based denoising by customized thresholding ,IEEE International Conference on Acoustics , Speech and Signal Processing ; 2004; 925-928.

[34] Byung-Jun Yoon , P.P.Vaidyanathan. Wavelet –based denoising by customized thresholding, IEEE International Conference on Acoustics , Speech and Signal Processing ; 2004; 925-928.

[35] Shaik Shafee, Prof.B.Anuradha,"Speaker Identification and Spoken word recognition in Noisy Enviroment using Different techniques", International Journal on Recent and Innovation trends in Computing and Communications, vol:4,Issue:6,ISSN:2321-8169,pp.590-595.