Continuous Density Hidden Markov Model for Hindi Speech Recognition

Shweta Sinha[1], S S Agrawal[2], Aruna Jain[3]

***Abstract* – State of the art automatic speech recognition system uses Mel frequency cepstral coefficients as feature extractor along with Gaussian mixture model for acoustic modeling but there is no standard value to assign number of mixture component in speech recognition process.Current choice of mixture component is arbitrary with little justification. Also the standard set for European languages can not be used in Hindi speech recognition due to mismatch in database size of the languages.Parameter estimation with too many or few component may inappropriately estimate the mixture model. Therefore, number of mixture is important for initial estimation of expectation maximization process. In this research work, the authors estimate number of Gaussian mixture component for Hindi database based upon the size of vocabulary.Mel frequency cepstral feature and perceptual linear predictive feature along with its extended variations with delta-delta-delta feature have been used to evaluate this number based on optimal recognition score of the system . Comparitive analysis of recognition performance for both the feature extraction methods on medium size Hindi database is also presented in this paper.HLDA has been used as feature reduction technique and also its impact on the recognition score has been highlighted.**

Index Terms- GMM, Hindi speech Recognition, HLDA, MFCC, PLP

I. INTRODUCTION

Speech recognition is the task of converting an utterance of speech signal to the streams of symbols i.e words or phonemes that represent information contained in the utterance.The utterances are captured by a microphone or any other transducer, digitized and then processed to produce a text sequence as close as possible to what was actually represented by the acoustic signal.

Over the past few decades automatic speech recognition(ASR) has achieved huge success and its implementation domain has also expanded from simplest system of digit recognition to spontaneous dialogue systems.Such growth is mainly due to advancement in feature extraction techniques and powerful modeling techniques for representing real data such as speech signal.

Several feature extraction mechanism have been implemented to boost system performance.Selection of these mechanism is guided by their capability to suppress information irrelevant to correct classification.Currently most popular features are Mel frequency cepstral coefficients(MFCC) [1] and Perceptual linear prediction coefficients (PLP) [2]. Several works have been done using combination of different features together[3][4].

For acoustic modeling, methods based on statistical models, specially Hidden Markov Model (HMM) has become the dominant choice in speech recognition. It is a powerful statistical method for efficient evaluation of distribution of random vector sequences obtained during feature extraction. Several variants of HMM have been used in recognition process that belongs to either discrete or continuous density model. But, in speech recognition continuous density HMM(CDHMM) has got wide acceptance[5].

Although, remarkable progress have been achieved in the area of ASR using HMM, there are still large number of problems that need to be solved. Issues that are governed by language characteristics and available database resources for the language still needs more attention. Evaluation of optimal number of Gaussian mixtures of acoustic signal for CDHMM is one such problem which is computationally expensive and needs to be sorted out. Little work has been done for Hindi language in the area of ASR[9].This may be due to lack of standard database for this language.

In this paper, authors have taken into account both the important aspect of speech recognition process ;the feature extraction technique and the acoustic modeling technique.An iterative procedure is followed to compute optimal number of Gaussians to be used in CDHMM for Hindi database.The paper describes two extended feature extraction process along with heteroscedastic discriminant analysis (HLDA) for dimensionality reduction of feature vectors.Recognition performance have been compared based on these features and their extended version keeping the number of gaussians in the mixture to the optimal value obtained for existing database.

First author is with Birla Institute of technology,Mesra, Ranchi,India. e-mail: meshweta_7@rediffmail.com.

Second author is with KIIT Group of colleges,Gurgaon,India. e-mail: ss_agrawal@hotmail.com.

Third author is with Birla Institute of technology,Mesra, Ranchi,India. e-mail : arunajain@bitmesra.ac.in

DOI: 10.5176/2251-3043_3.2.264

The rest of the paper is organized as follows: section II reviews the related work and outlines the motivation and contribution of this work.section III describes ASR architecture,section IV illustrates continuous density hidden markov model for Hindi language.Section V describes experimental setup.Section VI dicusses results and findings,Conclusion is presented in the last section.

II. RELATED WORK

Speech recognition task has been performed using several feature extraction and acoustic modeling techniques by several researchers over the years. MJF Gales et al.[6] have investigated the use of Gaussian selection to increase the speed of a large vocabulary speech recognition system and shows that state likelihood estimation is highly computational and around 30-70% of the computation time is spent in calculating these probability.

F sha et al.[7] discusses problems of parameter estimation in CDHMM for ASR.GMM has been applied on TIMIT speech corpus and number has been optimized.They have further proposed a learning algorithm based on margin maximization.But size of Hindi speech corpus is not comparable to TIMIT database.

R K Aggarwal et al.[8] have used MFCC and its extended feature set in their work and have proposed some optimal value for their Hindi speech database.They have used maximum likelihood estimation technique,but for system to perform in unmatching train and test data this technique of estimation does not perforrm well. In [9] they have used MFCC and PLP in their research and have proposed several methods to combine them.But they have not compared the performance of these features independently.

Amrous Anissa Imen et al.[10] have used MFCC with delta and delta-delta features, while using full covariance matrix.This increases the computational cost, improving the recognition score merely by 1%.

C Kurian et al.[11] have worked for Malyalam speech and they have used PLP as feature vector .The recognition score reached is 83%.They did not include its delta and delta-delta feature to capture transition features of speech.

J T Chien et al.[12] have worked on Chinese speech corpus of around 32,000 words.They have used 32 Gaussian using different MFCC feature set and showed that recognition score increases by including feature derivatives.

E H Bourouba et al.[13] have used MFCC feature set with hybrid acoustic model to achieve 90% recognition score.They have showed that including derivatives of feature vector improves the recognition rate.

Z Hachkar et al.[14] have presented comparative evaluation of MFCC and PLP in noisy environment and they have shown that in noisy data PLP outperforms MFCC. They have used double derivatives of both the feature sets.

III. ASR ARCHITECTURE

ASR system consists of four basic modules comprising of feature extraction, acoustic model, language model and the recognizer for word and sentence level matching. A block diagram of an integrated approach to continuous speech recognition is shown in fig 1. The feature extraction module computes the acoustic feature vectors used to characterize the spectral properties of time varying speech signal.The acoustic match module evaluates the similarity between input feature vector sequence and a set of acoustic models created for each and every word in the task vocabulary. Language model is used by this module to determine the most likely sequence of words.Syntax and semantic rules are specified by word model and N-gram probabilities.The end decision is made by the recognizer by taking into account all likely word sequences and choosing the one which maximizes the matching score.

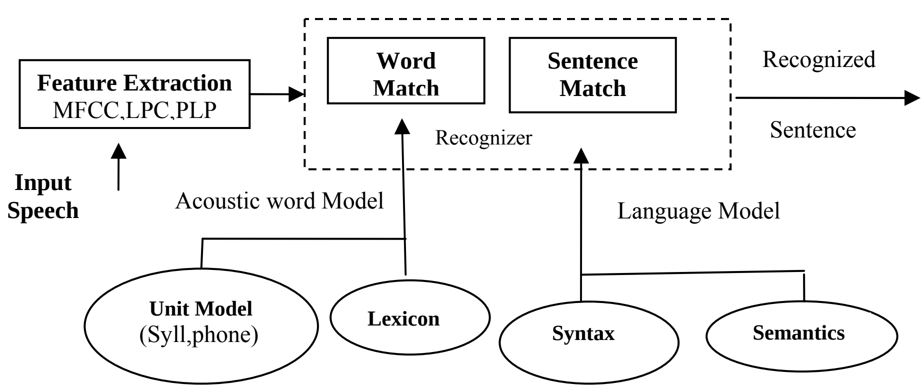

Fig 1. Block Diagram of continuous speech recognizer

A. *Speech Analysis and Feature Extraction*

The purpose of feature analysis is to parameterize the speech signal into a sequence of feature vector that contain relevant information about the sounds within the utterances[15]. A speech signal is formed by convolving the excitation signal from the lungs with the filter response of vocal tract.Research[1][2] shows that vocal tract component provides best discrimination between speech sounds.Most feature extraction methods use cepstral analysis to separate source and filter.Inclusion of attributes of psychophysical processes of human hearing into the analysis improves the recognition score.Mel frequecy cepstral coefficients (MFCC) and perceptual linear prediction(PLP) are the two widely used features based on human hearing perception of pitch.MFCC,PLP and their extended variations have been used in our experiment.

- MFCC – Mel cepstral feature extraction is used in almost every ASR system now a days.Its characteristics to model non linearity of human auditory system and capture phonetically important characteristics of speech has made it the most widely used.In this process speech samples are divided into overlapping frames of 20-30 ms (assuming stationarity of speech signal for this duration) at a frame rate of 10ms. Each frame is then windowed using Hamming window and its log magnitude is computed . To simulate the subjective spectrunm to human auditory system filter bank of triangular band pass filters spaced on mel scale is applied to obtain vector of log energies.The relation between linear frequency and mel frequency is given as:

$$mel(f) = 2595 * \log_{10}(1 + \frac{f}{700}) \qquad (1)$$

where f is the normal frquency measured in Hz and mel(f) is the frequecy output on mel scale.Discrete cosine transform(DCT) is applied to the output of the filterbank to obtain energy and cepstral coefficients.

- Extended Delta Features – Speech is a pseudo stationary signal and to capture the trajectories of the MFCC coefficients over time it has been observed that appending MFCC trajectories with original feature vectors increases ASR performance. Delta(Δ) and double delta(Δ-Δ) features using MFCC are the most commonly used feature in ASR research. Recently several work have been done including triple delta(Δ-Δ-Δ) feature and have shown that their inclusion enhances the system performance[8].

- PLP-Perceptual linear prediction technique is another method to compute speech features that simulate human auditory spectrum.This system uses three main concepts from psychophysics of hearing to derive an estimate of auditory spectrum. These are critical band spectrum resolution,equal loudness curve and intensity loudness power law[2].To obtain the auditory spectrum 17 band pass filter outputs are used. Their central frequencies are equally spaced on bark scale[16] given as :

$$\Omega(w) = 6\log\{(\frac{w}{600}) + [(w/600)^2 + 1]^{0.5}\} \qquad (2)$$

where w is the angular frequency in rad/sec and Ω is the Bark frequency. Inverse discrete fourier transform is applied to obtain cepstral coefficients.Extended delta features of PLP can also be computed and combined with the standard features.

B. *Feature Reduction*

Feature reduction is the mechanism that uses statistical methods to reduce the dimensions of the features,while maximizing the information that is preserved in the method feature space[17].These techniques are based on linear transformation schemes.Principal component analysis(PCA)[18], Linear discriminate analysis(LDA)[19] and Heteroscedastic linear discriminate analysis(HLDA) [20] are widely accepted feature reduction methods. HLDA concentrates on the larger difference of variance along one direction. It maximizes the likelihood of all the training data in the transformed space and each training sample contributes equally to the objective function[17].

C. *Acoustic Model*

In last three decades,methods based on the statistical models and specially on the Hidden Markov models(HMM) became one of the most powerful statistical methods for characterizing the spectral properties of speech signal.HMM is a very powerful statistical method of characterizing the observed data samples in the time series that can be discretely or continuously distributed.It is visualized as a finite state machine which uses a non deterministic process that generates output observation symbols in any given state[22].

During acoustic modeling process HMMs are created representing the basic phonetic units in the training data.Three interdependent problems need to be solved during this process.Most likelihood of the observation sequence has to be evaluated given the model φ, i.e. probability of model that generates the observation sequence has to be evaluated, by using Forward Algorithm. Most likely word sequence $\hat{W}$ given a set of feature vector has to be found by searching all possible state sequence arising from all possible word sequence,by using Viterbi algorithm and finally, the learning problem is solved using Forward-Backward algorithm where for each utterance the HMM that corresponds to the word sequence is found and composite HMM is created.

3-state HMM with each state having associated output probability distribution can be modelled for any of the discrete and continuous space problem.Speech is a continuous signal and the feature vectors derived in feature extraction phase is considered to be a sequence of vectors.These vectors can be converted into discrete symbols using vector quantization codebooks and other methods,but this may give rise to quantization error[21].Hence it would be advantageous to use HMMs with continuous observation densities to model continuous signal representation.

IV. CONTINUOUS DENSITY HIDDEN MARKOV MODEL FOR HINDI SPEECH

A. Database for Hindi Language

Hindi is the most common official language of India which is spoken in almost nine states of India.Though standard Hindi, the official language of India is based on Khariboli dialect, variations can be observed in the spoken Hindi in different parts of India.To make any speech recognition system ubiquitous, it is important that the system should be independent of speaker and language characteristics such as accents, speaking styles, disfluencies, syntax and grammar.

Hindi language uses Devnagri script. To process our data using standard tools (HTK,CMUSphinx) available for speech recognition we need to transliterate it using Roman characters. For Example: “अयोध्या नगरी एक दर्शनीय स्थल है” has to be written as “ayodhyaa nagrii ek drshaniie sthal hai”. The 46 phonemes of Hindi have to be represented with the help of 26 English alphabets.Many transliteration standards exists, but they are mainly based on the combination of both lower and upper case letters along with special characters like period and hyphen.Tools like CMUSphinx requires simple transliteration without any special characters in the database for proper functioning.ASR-11 developed by TIFR Mumbai defines some transliteration standards meeting to the constraints of these tools.In our experiment we have used this standard with few modifications.

B. Phonetic Transcription

One of the major task in speech recognition system building is to decide the phonetic structure of the recorded signal.In the beginning itself it is essential to identify how speech and non-speech units shall be represented.Whole word models are the simplest of all linguistic unit that can yeild good results in small databases and are useful in closed vocabulary system.Other options are syllable and phone based representation of speech signal.Phone based modeling of a system is complex but solves issues related to flexibility and feasibility of whole word models.These models based on single phoneme are called context independent model(CI).Phonemes may have several acoustic realization that may depend on context in which it occurs. This co-articulation effect needs to be captured for proper training of the system.To model these context dependent (CD) subword units senones are created [23].These senones are the combination of phonemes with its preceeding and following phonemes.Hidden Markov Model is used to represent these senones.Deciding the number of senones for the training database is also a major performance issue.Example:

Word	Transliteration	Phone model	Tri-phone model
नगरी	nagarii	n a g r ii	(sil-n-a, n-a-g, a-g-r, g- r-ii, r-ii-sil)

C. Continuous Density Hidden Markov Model

In using continuous observation density, some restriction must be placed on the form of model probability density function(pdf) to ensure that the parameters of the pdf can be reestimated in a consistent way[23]. In making a choice for continuous output probability density function, the first candidate is multivariate Gaussian mixture density function.

A Gaussian mixture model(GMM) is a parametric probability density function represented as a weighted sum of Gaussian component density[24].Weighted sum of M component Gaussian densities is given by the equation:

$$p(x|\gamma) = \sum_{i=1}^{M} w_i \, g\left(x \middle| \mu_i, \sum i\right) \qquad (3)$$

where, γ is a D-dimensional continuous feature vector, w_i, i=1...M are the mixture weights and $g(x|\mu_i,\Sigma i)$,i=1......M are the component Gaussian density.Each component density is a D-variate Gaussian function of the form,

$$g(x|\mu_i,\Sigma i) = \frac{1}{(2\pi)^{D/2}|\Sigma i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)}\right\} \qquad (4)$$

with mean μ_i, co-variance Σi and constraint on mixture weigh$\sum_{i=1}^{M} w_i$ = 1, where covariance can be full or diagonal. To estimate GMM parameters maximum a posteriori parameter estimation(MAP) is used. The choice of this estimation technique is based upon the fact that when cardinality of vector set is small and the training and the test condition differ, adaptation to changing environment is required. Since the data collected is from speakers of different regions along with the fact that training utterance is less as compared to the test utterance MAP is assumed to be suitable for this purpose. Value of M, ie number of mixture component per state is an important parameter for performance of ASR. In general, for European languages where large databases are available it is set as high as 64, but due to limit on the size of Hindi language database this value cannot give optimal result.

V. EXPERIMENTAL SETUP

To develop an ASR for any language it is necessary that the database should be rich enough to capture all the phonemes.Due to unavailability of Hindi speech data, authors have created their own database for this experiment. Phrases with 1000 unique words have been created. Recording of this database was done in a sound proof studio by 55 speakers, out of which 35 are male and 20 are females. 40% of the recorded utterance was used for training and 60% was used as testing token.Speech signal is digitized at a sampling rate of 16KHz.

Speech features of the recorded utterances were extracted by analysing the signal every 10 ms with a frame width of 25ms. MFCC and PLP feature sets were extracted from the recorded utterance. Two experiments have been performed ; first, to find the optimal number of gaussian component for different vocabulary size for both the MFCC and LPC features.Second experiment is conducted to compare the recognition performance of our system using different feature sets of the two feature extraction techniques along with feature reduction approach.Number of Gaussians for this experiment is set to the optimal value obtained from the first experiment.

The recognition system used in our experiment is an HMM based speaker independent speech recognizer.Context dependent triphone models are created.The models are left-to-right with no skip state transition.3-state HMMs are used for each model.Mixture of multivariate Gaussian distribution with diagonal covariance matrix is used for each state to approximate the probability density function.

VI. RESULTS and DISCUSSIONS

A. *Vocabulary Size vs. Number of Gaussians*

The first experiment was performed to find the optimal number of Gaussian density component for medium sized Hindi speech database. 39 MFCC(MFCC,log Energy,Delta and Delta-Delta) and 39 PLP feature vectors were extracted.3-state HMM CD model was used in this experiment.Varying number of words in the vocabulary and number of Gaussian density component recognition score was obtained .Bi-gram model of text collected from Hindi news papers were used as language model. Table1 shows the recognition performance in different size vocabulary based on Gaussian component Formula used for calculating the recognition score is as:

$$\%\ Recognition\ score = 100 * \frac{Successfully\ Recognized\ Word}{Total\ Words\ in\ Test\ Set}$$

Further analysis of result obtained shows that number of gaussians have significant impact on the recognition score, also as the vocabulary size change optimal value of Gaussian also changes. For our medium sized corpus 8 gaussian component gives the optimal result with both the MFCC and PLP features.

TABLE 1. IMPACT OF NUMBER OF GAUSSIAN COMPONENT ON RECOGNITION SCORE

Number of Gaussian Mixture Compo-nent	**% Recognition for MFCC and PLP Based on Vocabulary Size in Words**							
	135		**250**		**400**		**1000**	
	MFCC	**PLP**	**MFCC**	**PLP**	**MFCC**	**PLP**	**MFCC**	**PLP**
2	86	83	84	83	87	81	78	76
4	98	92	93	91	90	88	84	82
8	98	96	96	95	92	91	89	87
16	89	89	89	89	84	85	82	83

B. *Recognition Performance vs. Speech Features*

This experiment aims to measure suitability of feature extraction techniques. These experiments were performed in phases where the authors have used their complete database of 650 words and have applied PLP and MFCC as feature extraction methods.Similar to the previous experiment speech signal was sampled at 16KHz and then processed at 10ms frame rate with a Hamming window of 25ms to extract MFCC (12 coefficients,1 energy coefficient) and PLP (12 coefficients,1 energy coefficient) features.To measure the transition features in the next phase we included first(13-delta) and second(13-delta-delta) order derivatives also and in the third phase, third(13-delta-delta-delta) order derivatives of these features were also included in the feature sets.While experimenting with third order derivatives, HLDA was applied as feature reduction technique to obtain 39 features from 52 features.CDHMM with 8 Gaussian was used during acoustic modeling. The results obtained from our experiments using MFCC and PLP features are outlined in Table 2.

TABLE 2. RECOGNITION SECORE USING DIFFERENT FEATURE SET AND HLDA

Feature Set	% Recognition	Feature set	% Recognition
PLP (12 coefficients+ 1 Energy Feature)	81	MFCC (12 coefficients + 1 Energy Feature)	78
PLP+Δ+Δ-Δ (39 Features)	87	MFCC+Δ+Δ-Δ (39 Features)	89
PLP+Δ+Δ-Δ+Δ-Δ-Δ+HLDA (52 feature reduced to 39 feature)	91	MFCC+Δ+Δ-Δ+Δ-Δ-Δ +HLDA (52 feature reduced to 39 feature)	93

Results thus obtained show that MFCC out performs PLP. Also, using delta-delta-delta features recognition score can further be enhanced by 3-4% .The experiment highlights the efficiency of HLDA for feature reduction, which is able to captures all the dominant feature vectors while discarding others to give enhanced recognition score.

VII. CONCLUSION

Recognition of human speech by machine is a very challenging task. Despite research being carried out in this area for last so many decades no accurate system has yet been developed. There are still many open problems that need fast and precise solution in speech recognition process.A novel approach to develop a speech recognition system for Hindi using MFCC,PLP and their extensions have been discussed. Context independent units were taken as the smallest unit for recognition and CDHMM was applied. Experimental results show that MFCC are more robust as compared to PLP. The performance of system is tested by inclusion of third order derivative of speech features. Results show further improvement in the recognition rate by 3-4%. To cater with the increased amount of data processing, HLDA is applied as feature reduction mechanism. To estimate the optimal number of gaussian component experiments have been performed on different size vocabulary and results show that for a medium sized vocabulary of 500 to 1000 words 8 gaussian component gives optimal result. During the execution of these experiments it was observed thar there are still some open problem, especially for Hindi speech that need further attention , such as feature combination, inclusion of articulatory features in the feature set , number of tied state etc. Development of robust,adaptive and efficient system for Hindi speech recognition is the research frontier in ASR development.

REFERENCES

[1] Joseph Picone, Signal Modeling Techniques in Speech Recognition, Proc IEEE June1993

[2] H Hermansky, Perceptual linear predictive(PLP) analysis of speech, J. Acoustic Society of America,Vol 87,No 4, 1990

[3] András Zolnay, Daniil Kocharov, Ralf Schlüter, Hermann Ney, Using Multiple Acoustic Feature Sets for Speech Recognition,Speech Communication, April 2007.

[4] Zolnay, R. Schl¨uter, and H. Ney, Robust speech recognition using a voiced-unvoiced feature, in Proc. Int. Conf. on Spoken Language Processing, Denver, CO, vol 2,Sept. 2002.

[5] Alberto Abad, Thomas Pellegrini,Isabel Trancoso and Jo˜ao Neto, Context dependent modeling approaches for hybrid speech recognizers, INTERSPEECH 2010, 26-30 September 2010, Makuhari, Chiba, Japan

[6] MJF Gales,K M Knill,S J Young, State based gaussian selection in large vocabulary continuous speech recognition using HMMs, IEEE transactions on Speech and Audio Processing,vol 7,No.2,1999.

[7] F Sha,L K Saul, Large margin hidden markov model for automatic speech recognition,Proc. of Advances in neural information processing system, 2007

[8] R. K. Aggarwal and M. Dave, Using Gaussian Mixtures for Hindi Speech Recognition System, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 4, No. 4, December, 2011

[9] R.K. Aggarwal · M. Dave, Performance evaluation of sequentially combined heterogeneous feature streams for Hindi speech recognition system,Telecommunication Systems (2011):1-10

[10] Amrous Anissa Imen, Mohamed Debyeche,A Amrouche, Robust arabic speech recognition in noisy environment using prosodic features and formants, International journal of Speech Technology,14(4),2011.

[11] C.Kurian, K.Balakrishnan, Malayalam isolated digit recognition using HMM and cepstral coefficients, International journal of advanced information technology, 1(5),2011.

[12] J T Chien,C H Chueh, Joint acoustic and language modeling for speech recognition, Speech Communication,Vol. 52,2010.

[13] E H Bourouba,M Bedda, R Djemili,Isolated word recognition system based on hybrid approach DTW/GHMM,Informatica, vol. 30,2006.

[14] Z Hachkar, B Mounir, A Farchi, J Eiabbadi, Comparision of MFCC and PLP parameterization in pattern recognition of Arabic alphabet speech, Canadian journal of AI,Machine Learning and Pattern Recognition,vol 2(3),2011.

[15] Chin-Hui Lee, Biing-Hwang Juang, A Survey on Automatic Speech Recognition with an Illustrative Example on Continuous Speech Recognition of Mandarin, Computational Linguistics and Chinese Language Processing vol.1, no.1, August 1996.

[16] Schroeder, MR, Recognition of complex acoustic signals, life sciences research report 5 , edited by T H Bullock.

[17] Stefan Geirhofer,Feature Reduction with Linear Discriminant Analysis and its Performance on Phoneme Recognition University of Illinois at Urbana-Champaign, Department of Electrical and Computer Engineering, May 8, 2004

[18] Amaro Lima, Heiga Zen, Yoshihiko Nankaku, Chiyomi Miyajima, Keiichi Tokuda, Tadashi Kitamura, On the Use of

Kernel PCA for Feature Extraction in Speech Recognition, EUROSPEECH 2003
[19] Xiao-Bing Li, Jin-Yu Li, Ren-Hua Wang, Dimensionality Reduction Using Mce-Optimized Lda Transformation, Microsoft Research, ICASSP 2004
[20] N. Singh-Miller, M. Collins, and T. J. Hazen, Dimensionality reduction for speech recognition using neighborhood components analysis, in Proceedings of Interspeech-07, 2007
[21] B.H. Juang, Lawrence R. Rabiner, Automatic Speech Recognition – A Brief History of the Technology Development, Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2005
[22] Xuedong Huang, Alex Acero,Hsiao-Wuen, Spoken Language Processing A Guide to Theory, Algorithm and System Development, Prentice-hall Inc. 2001
[23] CMU-Sphinx Speaker Independent Speech Recognition System http://www.speech.cs.cmu.edu/sphinx/
[24] Reynolds, D.A., A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification. PhD thesis, Georgia Institute of Technology (1992)

Shweta Sinha has completed her M.Tech from GGSIP University Delhi, India. She has done MCA,B.SC in Physics. She has 10 years of teaching experience. She has also worked on live project based on speech technology. She has published 9 papers in reputed journals and National and International Conferences. Currently, she is pursuing her full time Ph.D research work in Dept. of IT , BIT Mesra, Ranchi.

Prof. S. S. Agrawal is presently working as Director General with KIIT group of colleges, Gurgaon. He has done his Ph.D in 1970 from AMU,Uttar Pradesh,India. He has also worked as visiting scientist in various research labs outside the country. Around 150 publications are to his credit. He is actively associated with development of annotated data for various languages. He is also engaged with many international speech groups and Projects.

Dr. Aruna Jain has received her Ph.D.from BIT Mesra, Ranchi, Jharkhand, India in the year 2009. She has done M.Tech in Computer Science and M.Sc. in Physics. She has published around 30 papers in reputed Journals and National and International Conferences. She has acted as resource person in various National and International conferences and editorial board member in reputed Journals. Her fields of Research are Computer Networks & Security,Data Mining, Soft Computing, and Web Engineering. She has more than 20 years teaching experience. Currently she is working as Associate Professor in Department of IT, BIT Mesra, Ranchi, Jharkhand, India and guiding Ph.D. research scholars.