6th International Conference on Multidisciplinary Research (ICMR-2019)
Osmania University Campus, Hyderabad (India)
30th-31st May 2019 www.conferenceworld.in
ISBN : 978-93-87793-89-7

ANALYSIS OF SPEECH RECOGNITION METHOD MFCC AND LDA CLASSIFER FOR HINDI WORDS

Priya Deokar[1], Sakshi Paithane[2]

[1]*ME Student, Electronics and Telecommunication Engineering Department, JSPM's RSCOE Pune (M.S.) India*

[2]*Assistant Professor, Electronics and Telecommunication Engineering Department, JSPM's RSCOE Pune (M.S.) India*

ABSTRACT

Speech recognition is that the task of taking Associate in Nursing vocalization of speech signal as input captured by electro-acoustic transducer and changing it into a text sequence as shut as attainable to what was described by the acoustic knowledge. In Asian nation there area unit twenty two govt languages and 419 alternative keep languages. Majority of population is unaware of English either in reading or writing however they'll cash in from resources of data Technology sector if they support native idiom. thus there's vast scope to develop such systems in Hindi language.

In this work, Indian Hindi language speech recognition is given mistreatment MFCC feature extraction and Linear Discriminant Analysis (LDA) classifier. The performance of algorithmic rule is compared with completely different classifiers. From the intensive experiment it's discovered that MFCC+LDA perform higher for low dataset than previous methodology.

Keywords: Hindi Language, Linear Discriminant Analysis, Mel Frequency Cepstral Coeffiicents, Nursing Vocalization, Speech Recognition

1. INTRODUCTION

Automatic Speech Recognition (ASR) is that the primary part or scheme of a human-computer interaction (HCI) system that needs human speech as decoded text input . ASR is comprised of 2 major parts, namely, acoustic model and language model . The acoustic model models the pronunciation of a given word, whereas the language model predicts the probability of a given word sequence showing in a very language. The parts of associate degree acoustic model is the speech signal options and a pattern matching technique for a given word or phone. The term, 'phone' represents a basic unit of speech. A word could include one or additional phones. the foremost unremarkably used options of ASR is Mel-Frequency Cepstral Coefficients (MFCC) and sensory activity Linear prognostic (PLP) Coefficients, whereas Hidden Andrei Markov Model (HMM) and neural network area unit the foremost unremarkably used pattern matching techniques. HMM perceive the consecutive nature of speech signal and model the output likelihood distribution further because the state transition likelihood. whereas recognizing the speech, numerous words area unit hypothesized against the no heritable signal. HMM performs matching by decisive

the probability of a given word. The probability of a word is calculable supported the mix of probability of all the phones related to the word. Earlier, most probability (ML) estimation has been wide exploited to coach HMM. However, discriminative coaching has been found as less dimmed than milliliter within the later era. On the opposite hand, the language model estimates the probability of a given word sequence showing within the speech signal. N-gram language model is that the most ordinarily used language model that predicts the likelihood of incidence of a word in a very sequence, once the history of word sequences is given. The likelihood calculation is predicated on an outsized text corpus given for coaching. hypothetic word occurrences area unit handled by the scores obtained from acoustic and language models. Isolated word is recognized because the word, that has highest probability among the combined likelihoods of all the words. However, Neural Network Language Models (NNLMs) are evidenced to produce higher performance than the standard N-gram language models. Few hybrid models have conjointly been used for up the cryptography method of ASR. This paper starts with the transient introduction on speech recognition system. Section II provides the literature survey of strategies for feature extraction and classification algorithms used for the speech recognition. Section III provides the outline of MFCC feature extraction formula. Section IV focuses on the linear discriminant analysis classifier. Section V elaborates the experimental results. Section VI provides the conclusion and future scope of the enforced system

2. LITERATURE REVIEW

Though there is not much attention gained by refereed publishers on speech recognition in Indian languages, we collect the other works and summarize to understand its current status. Referring the Wikipedia contents on Indian languages, we organize the methodologies based on the number of speakers of a language, as per the 2001 statistics.

Hindi is spoken by 258-422 million Indians, because of its honour as national language. In contrast to other languages, valiant attempts have been made on adopting methodologies for Hindi speech recognition. HMM [3] [5] gains considerable attention from the researchers who are working in the Hindi speech recognition.

Tamil, Bengali and Marathi languages are spoken by 80-90 million people approximately. It is just 20% of the Hindi speakers. HMM [10] [11] gains considerable attention in Tamil speech recognition in addition with trigram language model, dynamic time wrapping [9] and decision tree model [15]. Decision tree model [15] and DTW [16] are found to be promising for Bengali and Marathi speech recognition, respectively.

Nearly, 50-75 million people speak Telugu, Kannada and Urdu languages. HTK [25] and ANN [24] have gained considerable attention from Telugu speech recognition attempts, whereas DWT [21] and SVM models [22] [23] are promising for Kannada speech recognition systems. Sphinx decoder [19] [20] has been used in the majority of the Urdu speech recognition system.

These languages are spoken by 30-50 million Indians, which is just 10% of the Hindi speakers. ANN models are found to be promising for Gujarati [26] and Malayalam [29] [30] language models. Malayalam speech recognition systems have also exploited HMM models [28] [1] and DWT models [29]. The HMM models were also found to be suitable for Odiya [17] [18] and Punjabi languages [7].

ISBN : 978-93-87793-89-7

They are the least spoken languages among the previously discussed Indian languages. Only 13 million people in India speak Assamese language and 1.5 million people speak Bodo. ANN models [13] [27] have gained considerable interest on these languages, yet LVQ [14] has also been found suitable for Assamese language

3. SYSTEM METHODOLOGY

After receiving speech signal from electro-acoustic transducer it's processed into 3 main steps. In beginning pre-processing is completed on the signal. In second step some helpful and distinctive feature vectors square measure extracted from the auditory communication of speech signal victimization MFCC. Finally these extracted feature vectors square measure matched against the options of information victimization pattern recognition rule. Fig one shows diagram of Isolated Word Recognition (IWR)

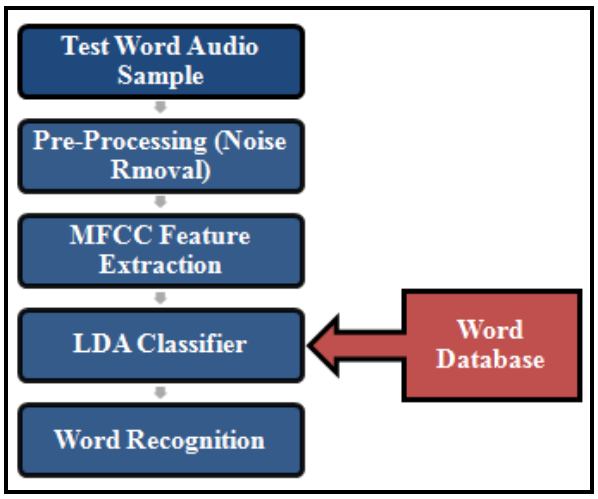

Fig.1 Block Diagram of IWR

Here the speech are given via electro-acoustic transducer connected to laptop that goes to be associate isolated word that should be recognized and displayed on laptop. To wipe out the impact of vocal organ and lip radiation signal pre-processing is employed. throughout utterance, lower frequencies square measure boosted whereas higher ones square measure suppressed, that causes loss of data in signal. to prevent this info loss and to preserve options of speech signal, high pass FIR filter is employed before feature extraction in speaker verification and speech recognition system. This method is named pre-emphasis. when pre-processing, the MFCC feature extraction stage extracts variety of predefined options from the processed speech signal. These extracted options should be able to discriminate between categories whereas being strong to any external conditions, like noise. Here the method of speech recognition is carried by LDA classifier. On PC, the text is outputted on screen. this can be nothing however the word given as input to electro-acoustic transducer

4. MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

MFCC is that the most evident and widespread feature extraction technique for speech recognition. It approximates the human sensory system response additional closely than the other system as a result of frequency bands area unit placed logarithmically here. {they area unit|they're} obtained from a Mel-frequency cepstrum wherever frequency bands are equally spaced on the Mel scale. Computation technique of MFCC is predicated on the short analysis and so from every frame MFCC vector is computed. Here we tend to used MFCC as feature extraction technique. In alternative words, MFCC is predicated on best-known variation of the human ear‘s crucial information measure with frequency. A diagram of the structure of AN MFCC processor is given in below Fig

ISBN : 978-93-87793-89-7

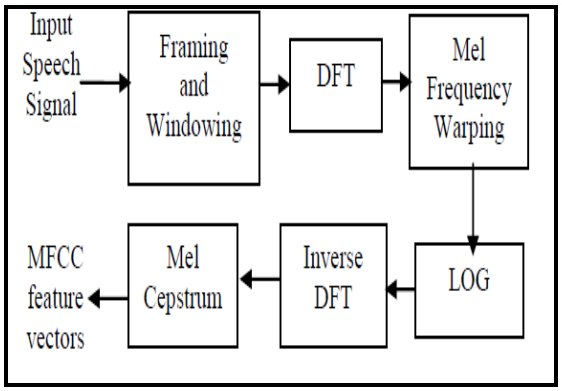

Fig.2 Block Diagram of MFCC

MFCC has 2 sorts of filters that area unit spaced linearly at low frequency below a thousand cps and power spacing on top of 1000Hz. A subjective pitch is gift on Mel Frequency Scale to capture necessary characteristic of phonetic in speech

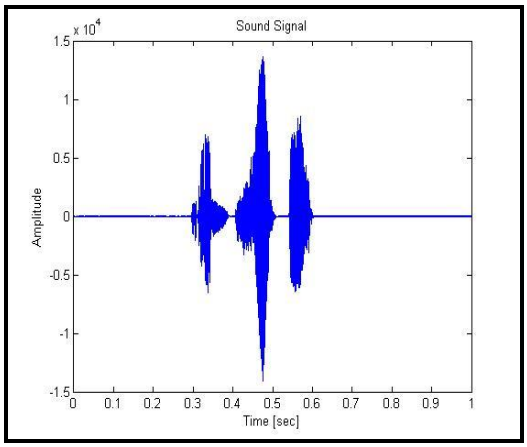

Fig.3 Original Sound signal

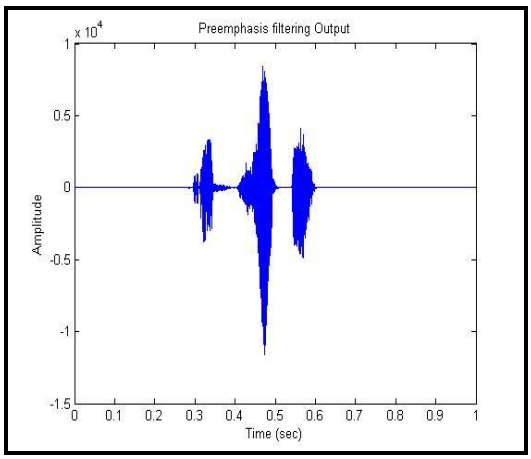

Fig.4 Pre-emphasis of signal

A. Framing

The first step is framing. The speech signal is go different ways into frames generally with the length of twenty to forty milliseconds[15]. The frame length is vital thanks to the exchange between time and frequency resolution. If it's too long it'll not be able to capture native spectral properties and if too short the frequency resolution would degrade. The frames overlap one another generally by twenty fifth to seventieth of their own length

B. Framing

After the signal is go different ways into frames every frame is increased by a window perform. additional acting window is employed for assortment shut frequency parts along ordinarily used windows throughout the frequency analysis of speech sounds area unit acting and Hanning window. The acting window is outlined as:

$$W(n) = 0.54 - 0.46(1 - \cos(\frac{2\pi n}{N-1})); 0 \le n \le N-1$$

Hamming window has highest side lobe attenuation and larger transition width of 8π/M where M is filter order. Hamming window is used to avoid Gibbs phenomenon. Every single frame is multiplied by hamming window to provide continuity of first and last point of the frame and to prevent abrupt changes at endpoint

ISBN : 978-93-87793-89-7

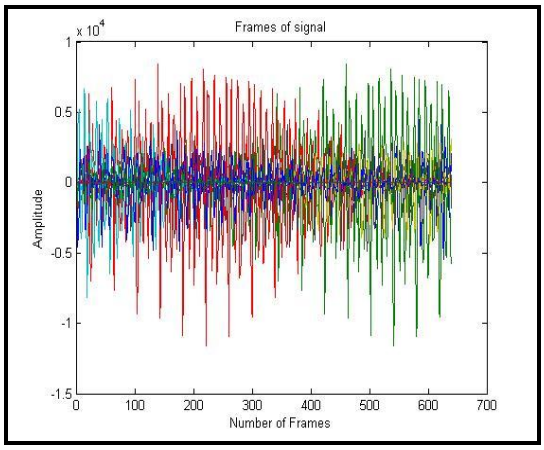

Fig.5 Framing of signal

A. Quick Fourier Transforming-

The third step is to use the separate Fourier remodel on every frame. Once windowing is finished, DFT is applied on every frame to convert samples of each frame to frequency domain from time domains

The quickest thanks to calculate the DFT is to use FFT that is AN algorithmic rule that may speed up DFT calculations by hundred-folds. we have a tendency to take log of spectrum so as to represent amplitude of spectral lines in decibel. Log of FFT shows speedy variations correspond to the basic frequency and slowly variable envelope corresponds to vocal tract parameters.

B. Mel- Frequency Warping

The Mel scale is predicated on however the human hearing perceives frequencies. it had been outlined by setting a thousand Mels capable a thousand Hz as a reference. Then listeners were asked to regulate the physical pitch till they perceived it as two- fold ten-fold and [*fr1] and people frequencies were then labelled as 2000 Mel, ten thousand Mel and five hundred Mel severally. The ensuing scale was known as the Mel scale and is just about linear below frequencies of a thousand Hz and exponent higher than. The Mel frequency are often approximated by the subsequent equation

$$m = 2595 \log_{10}(1 + \frac{f}{700})$$

Where f is the actual frequency and m is the Mel frequency

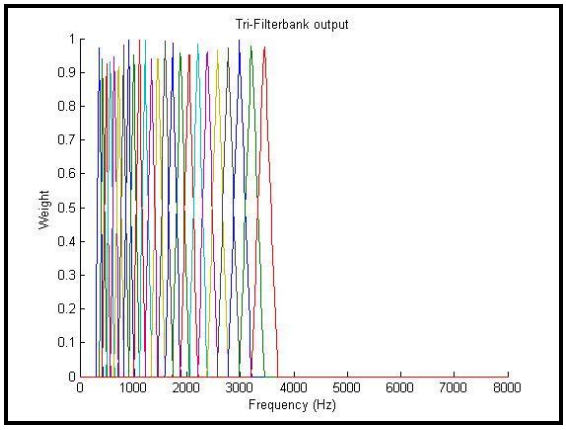

Fig.6 Triangular filter bank response

C. Log compression and discrete cosine transforming

Step 5 is to use compression by employing a log on the filter outputs Y (i) then to use the distinct trigonometric function rework that yields the MFCCs c[n] in keeping with the subsequent formula-

$$DCT = \sum_{i=1}^{M} \log(Y(i)) * \cos((\Pi n / M) * (i - 0.5))$$

The use of distinct trigonometric function reworking is completed since it's the property of high de-correlation, that is, it de-correlates all the mfcc coefficients. within the final step, the log Mel spectrum needs to be reborn back to time. The result's referred to as the Mel frequency cepstrum coefficients (MFCCs). The cepstral illustration of the sound spectrum provides an honest illustration of the native spectral properties of the signal for the given frame analysis

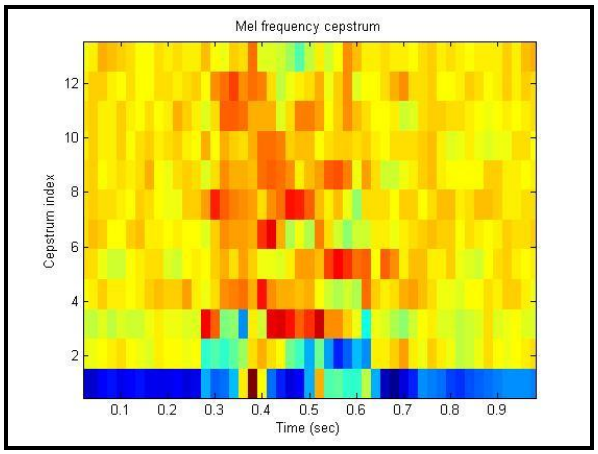

Fig.7 Mel Cepstrum Coefficients

5. LINEAR DISCRIMINANT ANALYSIS CLASSIFIER

LDA is additionally closely associated with principal part analysis (PCA) and correlational analysis therein they each explore for linear combos of variables that best justify the information. LDA expressly makes an attempt to model the distinction between the categories of information. PCA on the opposite hand doesn't take into consideration any distinction at school, and correlational analysis builds the feature combos supported variations instead of similarities. Discriminant Analysis is additionally completely different from correlational analysis therein it's not an mutuality technique: a distinction between freelance variables and dependent variables (also referred to as criterion variables) should be created.

LDA works once the measurements created on freelance variables for every observation area unit continuous quantities. once coping with categorical freelance variables, the equivalent technique is Discriminant correspondence analysis.[5][6]

Discriminant analysis is employed once teams area unit proverbial a priori (unlike in cluster analysis). every case should have a score on one or additional quantitative predictor measures, and a score on a gaggle live.[7] In easy terms, Discriminant perform analysis is classification - the act of distributing things into teams, categories or classes of identical sort.

In the case wherever there area unit quite 2 categories, the analysis employed in the derivation of the Fisher discriminant is extended to search out a mathematical space that seems to contain all of the category variability.[14]This generalization is thanks to C. R. Rao. Suppose that every of C categories contains a mean u and therefore the same variance . Then the scatter between category variability is also outlined by the sample variance of the category means that

$$\Sigma_b = \frac{1}{C}\sum_{i=1}^{C}(\mu_i - \mu)(\mu_i - \mu)^T$$

Where u is the mean of the class means

$$S = \frac{\vec{w}^T \Sigma_b \vec{w}}{\vec{w}^T \Sigma \vec{w}}$$

This means that when is an eigen vector of the separation will be equal to the corresponding eigen value

If $\Sigma^{-1}\Sigma_b$ is diagonalizable the variability between features will be contained in the subspace spanned by the eigenvectors corresponding to the $C-1$ largest eigenvalues (since Sigma is of rank $C-1$ at most)

These eigenvectors square measure primarily utilized in feature reduction, as in PCA. The eigenvectors like the smaller eigenvalues can tend to be terribly sensitive to the precise alternative of coaching knowledge, and it's usually necessary to use regularisation as delineate within the next section

If classification is needed, rather than dimension reduction, there square measure variety of different

techniques accessible. for example, the categories could also be divided, and a typical Fisher discriminant or LDA accustomed classify every partition. a typical example of this can be "one against the rest" wherever the points from one category square measure place in one cluster, and everything else within the alternative, and so LDA applied. this may end in C classifiers, whose results square measure combined. Another common methodology is pairwise classification, wherever a replacement classifier is made for every combine of categories giving $C(C-1)/2$ classifiers in total), with the individual classifiers combined to produce a final classification

6. EXPERIMENTAL RESULTS

For working of this system firstly we develop Hindi database. The database is obtained by recording Hindi isolated words and the recording is done for both male and female speakers at 16 KHz. The time given for recording speech samples is two seconds, because it was found that two seconds are enough for recording isolated words. If the time given for recording was more than two seconds that would result in having so much silence time in the recorded speech sample or the word's utterance. Recorded speech files were in .wave file. There are few isolated words as 'आपना', 'जल', ' घेन्हू', ' दाल', 'हिंदुस्तान', 'सपना' spoken by different speakers. Hence there are total 44 samples in database. This database is used for training and for testing purpose. Along with these 6 isolated words we have used 4 continuous sentences. Hence total there are 60 samples of isolated words and continuous sentence also the collected speech samples are then going to pass through the features extraction, features training and features testing stages

For the implementation the system used has following specification.

Table 1. Implementation System Details

System	Specification
Environment	Windows Operating System
System Details	Personal Laptop with 64 bit, Intel core i3 processor (2.53 GHz speed) and 3 GB RAM
Software	MATLAB R2013b (8.1.0.604 (R2013a))
Toolboxes	Image Processing Toolbox, Digital Signal processing Toolbox
Sampling frequency	16000 Hz
Recording Time	2 sec

The cross validation accuracy for the dataset is compared with MFCC+KNN and MFCC+SVM as shown in table 2. From the experimental results is it observed that MFCC+LDA performs better than the other two algorithms by giving 92.00 % accuracy

Table 2. % accuracy for different methods

Method	MFCC+KNN	MFCC+SVM	MFCC+LDA
% Accuracy	85.00 %	88. 50 %	92. 50 %

7. CONCLUSION

In this work we have done application MFCC method extraction algorithm for hindi word feature extraction. For the recognition of Hindi word LDA classifier is selected. The performance of the algorithm is

ISBN : 978-93-87793-89-7

measured on the basis of cross validation accuracy and compared with the several previous methods. For the experiment Hindi database is created which is having 100 different Hindi words and we have observed that results with MFCC+LDA methods has proven best results

References

[1] Anuj Mohamed, K. N. Ramachandran Nair- Continuous Malayalam speech recognition using Hidden Markov Models. Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010, Tamilnadu, India; 01/2010

[2] N. Rajput M. Kumar and A. Verma. A large-vocabulary continuous speech recognition system for Hindi. IBM Journal for Research and Development 2004

[3] Kumar, K. and Aggarwal, R. K., "Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research, ISSN (Online): 2229-6166, Volume 2 Issue 2 May 2011

[4] Aggarwal, R.K. and Dave, M., -Using Gaussian Mixtures for Hindi Speech Recognition System , International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No. 4, December 2011

[5] Mishra, A. N., Biswas, A., Chandra, M., Sharan, S. N., "Robust Hindi connected digits recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 2, June, 2011

[6] Sivaraman. G.; Samudravijaya, K., Hindi Speech Recognition and Online Speaker Adaptation, International Conference on Technology Systems and Management: ICTSM-2011, IJCA

[7] Kumar, R., Comparison of HMM and DTW for Isolated Word Recognition System for Punjabi Language, International Journal of Soft Computing 5(3):88-92, 2010

[8] R. Thangarajan, A.M. Natarajan, M. Selvam, "Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language", Wseas Transactions on Signal Processing, Issue 3, Volume 4, March 2008

[9] V.S.Dharun, M.Karnan, 2012, Voice and Speech Recognition for Tamil Words and Numerals, IJMER, Vol. 2, 5, pp 3406-3414

[10] C.Vimala, M.Krishnaveni , 2012, Continuous Speech Recognition system for Tamil language using monophone-based Hidden Markov Model , Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology CCSEIT '12 , pp 227-231

[11] Ms.Vimala C., V. Radha, Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM, Procedia Engineering, Volume 30, 2012, Pages 1097-1102

[12] S. Karpagavalliet. al , 2012 , Isolated Tamil Digits Speech Recognition using Vector Quantization, IJERT, Vol.1, 4, pp 1-9

[13] M. P. Sarma and K. K. Sarma, "Speech Recognition of Assamese Numerals using combinations of LPC-features and heterogenous ANNs", Proceedings of International Conference on Advances in Information and Communication Technologies (ICT 2010), Kochi, Kerala, India (V. V. Das and R. Vijaykumar (Eds.): ICT2010, CCIS 101, pp.8-11,2010, Springer-verlag, 2010

[14] Sarma, M. P.; Sarma, K. K.,Assamese Numeral Speech Recognition using Multiple Features and

Cooperative LVQ – Architectures, International Journal of Electrical and Electronics 5:1, 2011

[15] Banerjee, G. Garg, P. Mitra, and A. Basu, "Application of triphone clustering in acoustic modeling for continuous speech recognition in Bengali," in International Conference on Pattern Recognition, ICPR.,Dec. 2008, pp. 1–4

[16] Gawali, Bharti W., Gaikwad, S., Yannawar, P., Mehrotra Suresh C., ―Marathi Isolated Word Recognition System using MFCC and DTW Features (2010), Int. Conf. on Advances in Computer Science 2010, DOI: 02.ACS.2010. 01.73

[17] Mohanty, S.; Swain, B. K., ―Continuous Oriya Digit Recognition using Bakis Model of HMM, International Journal of Computer Information Systems, Vol. 2, No. 1, 2011

[18] Mohanty, S.; Swain, B. K., ―Markov Model Based Oriya Isolated Speech Recognizer-An Emerging Solution for Visually Impaired Students in School and Public Examination, Special Issue of IJCCT Vol. 2 Issue 2, 3, 4; International Conference On Communication Technology-2010

[19] Raza, A., Hussain, S., Sarfraz, H., Ullah, I., and Sarfraz, Z., An ASR System for Spontaneous Urdu Speech, Proceedings of OCOCOSDA' 09 and IEEE Xplore, 2009

[20] Sarfraz, H.; Hussain, S.; Bokhari, R.; Raza, A. A.; Ullah, I.; Sarfraz, Z.; Pervez, S.; Mustafa, A.; Javed, I.; Parveen, R., ―Large Vocabulary Continuous Speech Recognition for Urdu, International Conference on Frontiers of Information Technology, Islamabad, 2010

[21] M. A. Anusuya, S. K. Katti" Kannada speech Recognition using Discrete wavelet Transform-PCA" international conference on computer applications --24-27 - December 2010 - Pondicherry – Indi.

[22] SarikaHegde, Achary K.K., SurendraShetty " Isolated Word Recognition for Kannada Language Using Support Vector Machine" Wireless Networks and Computational Intelligence Communications in Computer and Information Science Volume 292, 2012, pp 262-269

[23] SarikaHegde ,K KAchary,SurendraShetty-Analysis of Isolated Word Recognition for Kannada Language using Pattern Recognition Approach-International Journal of Information Processing , 7(1), 73 - 80. -2013

[24] Sai Prasad P. S. V. S, .Girija P. N Speech Recognition of Isolated Telugu Vowels Using Neural Networks. Proceedings of the 1st Indian International Conference on Artificial Intelligence, IICAI 2003, Hyderabad, India, 28-34 December 18-20, 2003

[25] VijaiBhaskar P , Rama Mohan Rao S , Gopi A - "HTK Based Telugu Speech Recognition" International Journal of Advanced Research in Computer Science and Software Engineering – Volume 2, Issue 12, December 2012

[26] Patel Pravin, HarikrishnaJethva -" Neural Network Based Gujarati Language Speech Recognition" International Journal of Computer Science and Management Research Vol 2 Issue 5 May 2013

[27] M.K.Deka , C.K.Nath , S.K.Sarma , P.H. Talukdar - "An Approach to Noise Robust Speech Recognition using LPC-Cepstral Coefficient and MLP based Artificial Neural Network with respect to Assamese and Bodo Language" International Symposium on Devices

ISBN : 978-93-87793-89-7

MEMS, Intelligent Systems & Communication (ISDMISC) 2011

[28] Syama R, Suma Mary Idikkula (2008) "HMM Based Speech Recognition System for Malayalam", ICAI'08 – The 2008 International Conference on Artificial Intelligence, Monte Carlo Resort, Las Vegas, Nevada, USA (July 14-17, 2008)

[29] Krishnan, V.R.V. Jayakumar A, Anto P B (2008) , "Speech Recognition of isolated Malayalam Words Using Wavlet features and Artificial Neural Network". DELTA2008. 4th IEEE International Symposium on Electronic Design, Test and Applications, 2008.Volume,Issue, 23-25 Jan. 2008 Page(s):240 – 243

[30] A.R. Sukumar, A.F. Shah, and P.B. Anto, "Isolated question words recognition from speech queries by using Artificial Neural Networks", in proc. of IEEE 2nd International conference on Computing, Communication and Networking Technologies(ICCCNT), Karur, India, 2010, pp. 1-4