

Speaker Dependent and Independent Isolated Hindi Word Recognizer using Hidden Markov Model (HMM)

Ishan Bhardwaj
Student, NIT Raipur
69, S-2, GreenPark
Bareilly, U.P., India

Narendra D. Londhe
Asst, Professor, NIT Raipur
Electrical Department
NIT Raipur (C. G.)

ABSTRACT

Hindi is very complex language with large number of phonemes and being used with various accents in different regions in India. In this manuscript, speaker dependent and independent isolated Hindi word recognizers using the Hidden Markov Model (HMM) is implemented, under noisy environment. For this study, a set of 10 Hindi names has been chosen as a test set for which the training and testing is performed. The scheme instigated here implements the Mel Frequency Cepstral Coefficients (MFCC) in order to compute the acoustic features of the speech signal. Then, K-means algorithm is used for the codebook generation by performing clustering over the obtained feature space. Baum Welch algorithm is used for re-estimating the parameters, and finally for deciding the recognized Hindi word whose model likelihood is highest, Viterbi algorithm has been implemented; for the given HMM. This work resulted in successful recognition with 98.6% recognition rate for speaker dependent recognition, for total of 10 speakers (6 male, 4 female) and 97.5% for speaker independent isolated word recognizer for 10 speakers (male).

General Terms

Speech recognition, Isolated Hindi words.

Keywords

MFCC; HMM; Hindi; Isolated word; K-Means; Speech recognition.

1. INTRODUCTION

The paper is divided into two main sections first is about the speaker dependent isolated word recognizer and the second one is speaker independent isolated word recognizer; followed by conclusion. The reason behind choosing the Hidden Markov Model over other existing techniques; for example DTW, ANN etc., is that HMM has been proven to be one of the most powerful and dominating statistical approach over the last past years and have been applied in a no. of models like pattern [1, 2], DNA sequence analysis [3, 4], pathologies [5] or speech recognition [6,7]. The basic theory of HMM was published in a series of classic papers by Baum and his colleagues in the late 1960s and early 1970s which was then implemented for speech recognition applications by Baker at Carnegie Mellon University (CMU) and by Jelinek and his colleagues at IBM in the 1970s [8-20] and further explored by L. Rabiner, et al. [21-36] in 1980s and the early 1990s.

Theory of Hidden Markov Model and its applications are not new in the present time. However there is lot more to be done in the field of speech recognition despite of the work carried out till date, for languages like Hindi. First a brief introduction of HMM [28] and its parameters is presented here. A hidden

Markov model is basically a Markov chain where the output observation is a random variable generated according to an output probabilistic function associated with each state [40]. The HMM models a sequence of observations as a piecewise stationary process [37]. HMM works in two basic phases namely training, where the desired likelihoods are calculated in addition to model parameters for the input vectors and the testing phase, which comes up with result in form of recognized output. For the training phase, usually the Baum-Welch algorithm [38] is followed in order to re-estimate the parameters (π , A, B) for the HMM. Not only the Baum-Welch but the Forward- Backward Algorithm is also instigated in order to calculate the required alpha, beta parameters. The Viterbi algorithm [38] is also implemented for the estimation of most probable state sequences. Let us first introduce the parameters and their notations used throughout this paper.

N = number of states in the model

M = number of Distinct observation symbols.

T = length of observation sequence.

q_t = the state at time t .

$V = \{v_1, v_2, v_3, \dots, v_M\}$, the discrete set of possible observation symbols.

$\pi = \{\pi_i\}$, where $\pi_i = P(q_1 = S_i)$.

$A = \{a_{ij}\}$, where $a_{ij} = P(q_{t+1} = S_j | q_t = S_i) 1 \leq i, j \leq N$.

$B = \{b_{jk}\}$, where $b_{jk} = P(v_k \text{ at } t | q_t = S_j) 1 \leq j \leq N \text{ and } 1 \leq k \leq M$.

A compact notation is used to indicate the complete parameter set of the model

$$\lambda = (A, B, \pi) \quad (1)$$

Now the question arises here is that how do one choose the type of model and the parameters. For isolated word speech recognition, with a distinct HMM designed for each word in the vocabulary, the left right model is found to be more appropriate than other models like the Ergodic model etc. Since the left right model is said to associate time with model states as distinct sounds of the word being modelled. This topology is, in fact, one of the most popular HMM structures used in state-of the art speech recognition systems. An Ergodic model has property that every state could be reached from any other state as shown in Fig. 1(a). Fig. 1(b) on the other hand shows two left-right models which are connected in a cross-coupled manner. The other models, A 4-state left-right or Bakis model [17,18] is also shown in Fig. 1(c). Whereas the Fig. 1(d) shows a modified Bakis model which is implemented in this paper. The fundamental property to be satisfied by all the left-right models is that state transition coefficients possess the following condition.

$$a_{ij} = 0, j < i \quad (2)$$

Furthermore, the initial state probabilities have the property

$$\pi = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (3)$$

These conditions result into a model with transitions from current state to another state only if its indices are higher and the state sequence is allowed to begin at state 1 only. It has been seen that the error is somewhat insensitive to the no. of states, N. The local minimum is said to be achieved at N=6 [39]. But here 4 states are chosen for faster computation without any discernible degradation in the recognition. M=256 code words are chosen for codebooks to use with discrete symbol models, hamming shaped filters with 24 number of filters in filter-bank. The other parameters (like π, A, B) are estimated with the help of above mentioned techniques and algorithms in addition to the initial values chosen based on some good guess because in the discrete HMM, if a probability is initialized as zero, it will remain zero even after infinite no. of iterations, in other words it will never converge. Thus, it is required to have sensible initial estimates. Empirical study has shown that, for discrete HMMs, uniform distribution can be used as the initial estimate [40]. Initial estimates are essential in the continuous-distribution case and can be obtained in a no. of ways including Segmental K-Means algorithm [35, 36] with clustering.

The recogniser discussed here is for isolated words recognition not for continuous speech. Block diagram of a typical HMM based isolated word recogniser [39] is shown in Fig. 2 Where first the input signal (speech signal) is fed to the feature analysis module. Then the observation sequences, obtained after the vector quantisation, are used for probability computation against all the models of the dictionary words and finally the word whose model likelihood found to be highest is said to be recognised. The detailed description of this recognition process is given in the section 2.1 specifically for the system implemented.

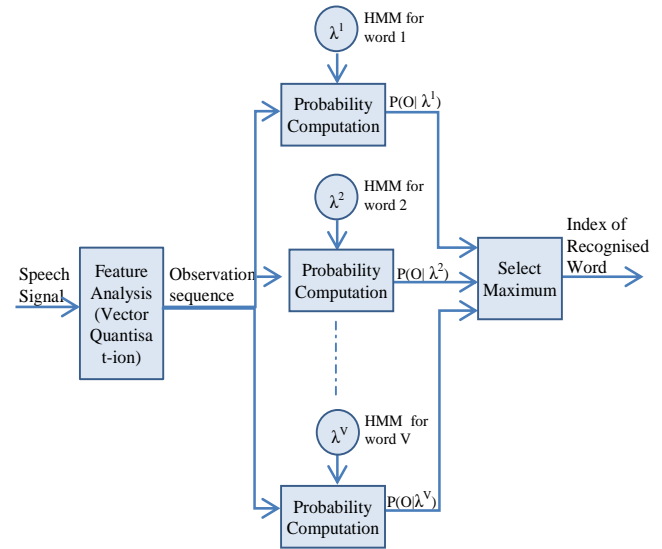


Fig. 2. Block Diagram of an isolated word HMM recognizer [31]

2. SPEAKER DEPENDENT ISOLATED HINDI WORD RECOGNIZER

Speaker dependent isolated word recogniser is one of the simplest kinds of recogniser which is based on the fact that the system will recognise the speech for a particular speaker if and only if he/she has already trained the system by first building the vocabulary of the recognition system for the words to be recognised. Otherwise the recognition rate may be null or unacceptably less. For building the isolated word speaker dependent recognition for Hindi speech here the HMM is implemented. A set of ten Hindi words is selected which is recorded by all the speakers individually. Each word is recorded at least 10 times by a speaker.

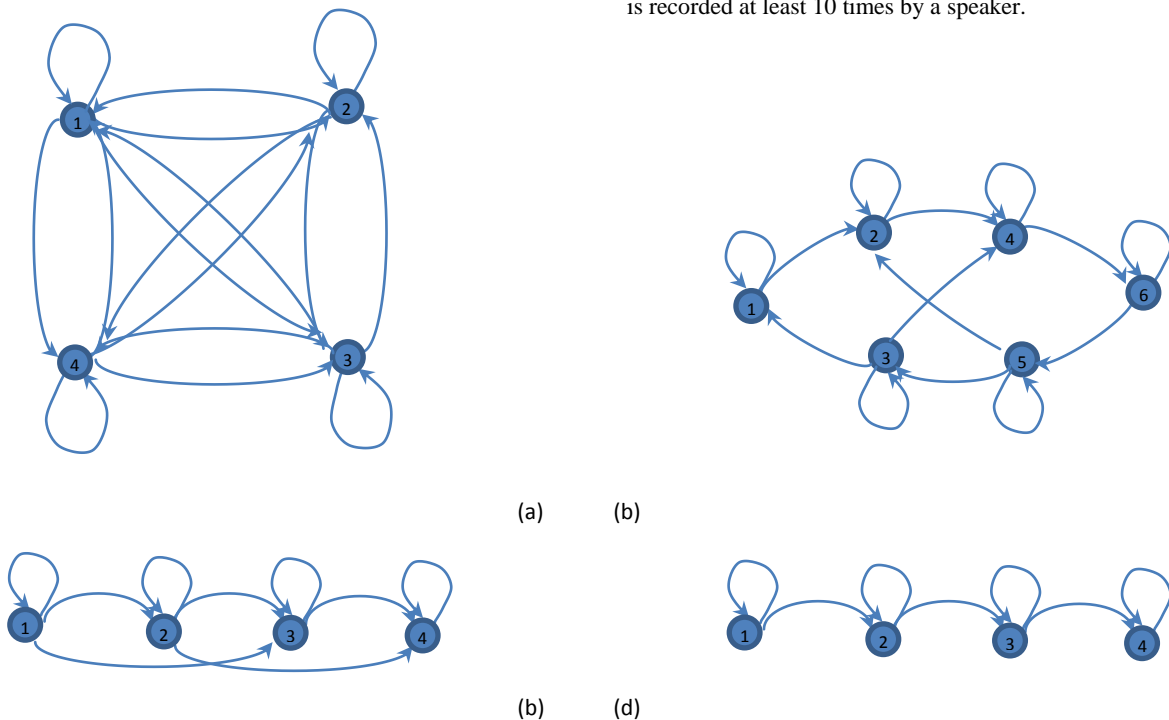


Fig. 1. Illustration of types of models (a) a 4-state Ergodic Model; (b) a 6-state parallel path left-right model; (c) a 4-state left-right model; (d) a 4-state left-right model implemented in this paper

After that the training is done and then the recognition (testing) is carried out by the same speaker. This process is repeated by all of the speakers. The working of this speaker independent recognition is as follows.

2.1 System Overview

For recording a regular microphone (Logitech PN 981-000214 Headset) is used with the following specification : Supra-aural, Over-the-head design; Rotating boom microphone; 100 Hz - 16000 Hz Microphone frequency response; -38 dB/mW Microphone sensitivity at 1 kHz. The environment for the recording was not noise-free. The same conditions and hardware were used at the time of recognition by all the 10 speakers.

First a vocabulary of $V=10$ is selected which have to be recognized. Here the recordings are done by the talkers from various states of India in order to consider the variations in their accent, which includes Jammu and Kashmir, Rajasthan, Gujarat, Uttar Pradesh, Madhya Pradesh, Bihar, Jharkhand, Orissa, Chhattisgarh and Assam; who have inured their native languages. Then for each word in vocabulary $K=10$ occurrences were recorded and each occurrence here epitomize to an observation sequence. These observations are nothing but some appropriate representation of the temporal or spectral characteristic of the words. The feature analysis here implemented is the MFCC, a representation defined as the real cepstrum of a windowed short-time signal derived from the FFT of that signal. MFCC is preferred here because study has showed the MFCC representation to be beneficial for speech recognition [42]. For each word a distinct HMM is built and the model parameters are thus required to be estimated. The good initial estimates of the parameters of $b_j(O_t)$ densities are vital for appropriate and fast convergence of the re-estimation formulas. After initializing the HMM, on the basis of current λ the set of training observation sequences are segmented into states. This segmentation is achieved by implementing the Viterbi algorithm, and then backtracking along the optimal path. As a result of this segmentation of each of the training sequence, for each of the N states, is a maximum likelihood estimate of the set of observations that occur according the current model within each state S_i . The updated initial parameters are calculated with the help of K-means algorithm by means of Euclidian distortion measure. Baum welch algorithm is used for re-estimating the parameters [43].

In recognition phase any utterance is recognised by first obtaining the feature vectors from the speech signals (using MFCC). Then the measurement of Observation sequence is done $O = \{O_1, O_2, O_3, \dots, O_T\}$; followed by calculation of model likelihoods, by implementing Viterbi algorithm, for all the models in order to have it for all the words in vocabulary. Then using the function $v^* = [P(O|\lambda_v)]$, $1 \leq v \leq V$ to find out the word whose model likelihood is highest. This word is said to be recognised. The flow diagram of the system for a test signal is illustrated in the figure 3.

2.2 System Interface

The interface of the system was developed using Matlab®. It is designed to be simple and user-friendly. The Menu for recording is shown in Fig. 4. The Fig. 4(a) shows the initial recording interface whereas the Fig. 4(b) shows the recording menu for additional words when one or more than one words have been already recorded by the speaker.

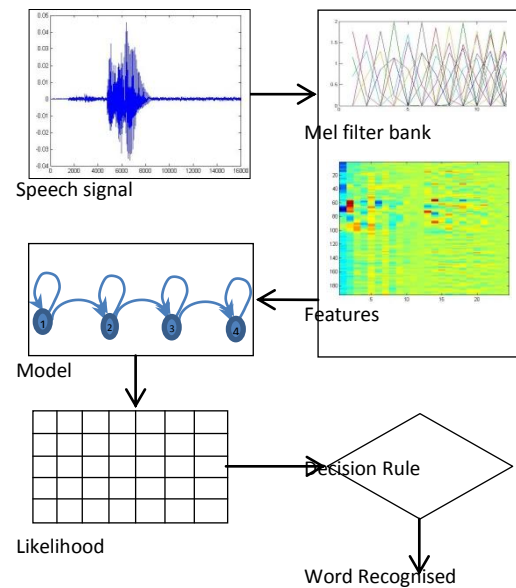


Fig.3. Block Diagram of Recognition system implemented

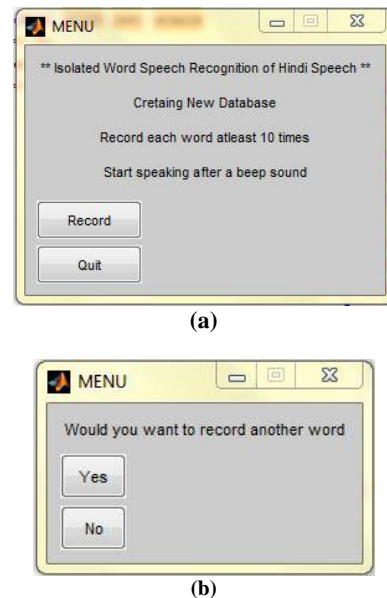


Fig. 4. User Interface (a) initial recording menu; (b) menu for recording additional words;

2.3 Results

Results achieved for speaker dependent isolated hindi word recognizer are quite satisfactory. As the recording environment at the time of building the vocabulary as well as the time of testing was not noise free and the successful recognition here obtained is 98.6%. The speech signals, used for training, were manually trimmed to remove silence from start and end of the signal. Table 1 is showing the results obtained by this speaker dependent isolated Hindi word recognizer for 10 speakers (4 Females, 6 Males) where each speaker has tested every word in the dictionary 10 times. The table entry shows out of 10 the number of correctly recognized utterances.

3. SPEAKERINDEPENDENTISOLATE-D HINDI WORD RECOGNIZER

The same basic blocks and techniques are used as in speaker dependent recognizer discussed above. The only difference lies in the training phase where instead of recording utterances for individual speakers, a dictionary is made from recording by 10 speakers, vocabulary of V=10 is selected which have to be recognized. Now for each word in vocabulary there are exactly two occurrences per speaker and thus a training set of total K=20 occurrences was used. All the talkers who have participated in recognition (10 males) are different from speakers (10 males) who have contributed in building the dictionary for training phase. Thus the recognizer here works in a speaker independent manner. The idea is inspired from the theory and results given by Rabiner et al. [43]. Table 2 is

showing the results obtained by this speaker independent isolated Hindi word recognizer. The speaker independent isolated word recognizers [44] as well as continuous speech recognizers [46-48] both are the area of interest for the researchers and significant work has been done incorporating various techniques and introducing modification [49-54] to the existing system.

3.1 Results

The performance of this isolated Hindi word recogniser has not been degraded too much with the testing in a speaker independent fashion as compared to the speaker dependent isolated Hindi word recogniser. Even here no manual trimming of training set signals has been done. The recognition results obtained here, 97.5%, are thus quite motivating.

Table 1. Recognition Results for speaker dependent isolated Hindi word recognition

Word to be recognised	Recognition Results									
	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9	Speaker 10
Ishan	10/10	10/10	10/10	10/9	10/10	10/10	10/10	10/10	10/10	10/10
Ashwini	10/10	10/9	10/10	10/10	10/10	10/9	10/10	10/10	10/10	10/10
Darshan	10/9	10/10	10/10	10/8	10/10	10/10	10/10	10/10	10/10	10/10
Shastry	10/10	10/10	10/10	10/10	10/9	10/10	10/10	10/10	10/10	10/10
Devendra	10/8	10/10	10/9	10/10	10/10	10/9	10/10	10/10	10/10	10/10
Tejash	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Sudeshna	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/9	10/10
Shilpi	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Mitali	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Ashok	10/10	10/10	10/10	10/10	10/9	10/10	10/10	10/10	10/10	10/9
Recognition Rate (%)	97	99	99	97	98	98	100	100	99	99

Table 2. Recognition Results for speaker independent isolated Hindi word recognition

Word to be recognised	Recognition Results									
	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Speaker 6	Speaker 7	Speaker 8	Speaker 9	Speaker 10
Ishan	10/10	10/9	10/10	10/10	10/10	10/10	10/10	10/10	10/9	10/10
Ashwini	10/9	10/10	10/9	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Darshan	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/9	10/10	10/10
Shastry	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Devendra	10/10	10/7	10/8	10/10	10/10	10/9	10/10	10/9	10/10	10/9
Tejash	10/10	10/9	10/10	10/10	10/9	10/8	10/10	10/10	10/10	10/10
Sudeshna	10/10	10/10	10/10	10/9	10/10	10/10	10/10	10/10	10/10	10/10
Shilpi	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Mitali	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Ashok	10/9	10/8	10/10	10/10	10/10	10/9	10/10	10/10	10/9	10/8
Recognition Rate (%)	98	93	97	99	99	96	100	98	98	97

4. CONCLUSION

The isolated word recognition for Hindi speech is performed in both speaker dependent and independent manner. The existing models developed for English language can't be used directly for Hindi because out of 64 phones in Hindi language only 39 are present in English. It is hard to achieve 100% recognition in Hindi language with models designed for English language. This happens due to their phonetic space differences raised because of their different acoustic characteristics. Despite Hindi being a phonetic language, for which better rules can be formed to generate the base-forms from spelling, with the way people from different part of India use this language these rules are not followed at times.

The results obtained here in both the cases are quite satisfactory and could be further improved by using better noise canceling hardware and techniques. Moreover, the recognition of similar words could be improved by using better codebook generation. The speaker independence is achieved successfully which could be further enhanced by increasing the training data set by more number of speakers as well as instigating specialized paradigm for speaker independence. The implementation of large vocabulary speech recognition as well as recognition of continuous speech is left as a part of future work.

5. REFERENCES

- [1] Ferrer, M. A., Camino, J. L., Travieso, C. M., Morales, C., 1999, Signature Classification by Hidden Markov Model, 33rd Annual IEEE International Carnahan Conference on Security Technology, (IEEE ICCST'99), Comisaría General de Policía Científica, Ministerio del Interior, IEEE Spain Section, COIT, SSR-UPM, SeguridadSeguridadEspaña S.A, Madrid, Spain, Oct. 1999, 481-484.
- [2] Sánchez, J. A., Travieso, C. M., Alonso, I. G., Ferrer, M. A., 2001, Handwritten Recognizer By Its Envelope and Strokes Layout Using HMM's, 35rd Annual 2001 IEEE International Carnahan Conference on Security Technology, (IEEE ICCST'01), London, UK, 267-271.
- [3] Yin, M.M., Wang, J.T.L., 1999, Application of Hidden Markov Models to Gene Prediction in DNA, International Conference on Information Intelligence and Systems, 1999. Proceedings., 40 – 47.
- [4] Cohen, A., 1998, Hidden Markov Models in Biomedical Signal Processing, Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conf. of the IEEE, 3, 1145 – 1150.
- [5] Alonso, J. B., Carmona, C., León, J. de, Ferrer, M. A., 2002, Combining Neural Networks and Hidden Markov Models for Automatic Detection of Pathologies, 16th Biennial International Eurasip Conference Biosignal 2002, Brno, Check Republic.
- [6] Bahl, L.R., Brown, P.F., de Souza, P.V., and Mercer, R.L., 1986, Maximum Mutual Information Estimation of HMM Parameters For Speech Recognition., In Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal processing., Tokyo, Japan.
- [7] Renals, S., Morgan, N., Bourlard, H., Cohen, M. & Franco, H. 1994, Connectionist Probability Estimators in HMM Speech Recognition, IEEE Transactions on Speech and Audio Processing 2(1), 161-174.
- [8] Buam, L. E. and Petrie, T., 1966, Statistical Inference for Probabilistic Functions of Finite State Markov Chains, Ann. Math. Stat., 37, 1554-1563.
- [9] Buam, L. E. and Egon, J.A., 1967, An Inequality with Applications to Statistical Estimation for Probabilistic Functions of a Markov Process and to a Model for Ecology, Bull. Amer. Meteorol. Soc., 73, 360-363.
- [10] Buam, L. E. and Sell, G. R., 1968, Growth Functions of Transformation on Manifold, Pac. J. Math., 27(2), 211-227.
- [11] Jelinek, F., 1969, A Fast Sequential Decoding Algorithm Using a Stack, IBM J. Res. Develop., 13, 675-685.
- [12] Buam, L. E., et al., 1970, A Maximisation Technique Occurring in the Stochastic Analysis of Probabilistic Function of Markov Chains, Ann. Math. Stat., 41(1), 164-171.
- [13] Buam, L. E., 1972, An Inequality and Associated Maximisation Technique in Stochastic Estimation of Probabilistic Function of Markov Process, Inequalities, 3, 1-8.
- [14] Bahl, L. R. and 1975, Decoding for Channels With Insertions Deletions Substitution with Applications to Speech Recognition, IEEE Trans. Informat. Theory, IT-21, 404-411.
- [15] Jelinek, F., et al., 1975, Design of a Linguistic Stochastic Decoder For The Recognition of Continuous Speech," IEEE Trans. Informat. Theory, 64, 250-256.
- [16] Baker, J. K., 1975, The Dragon System-An Overview, IEEE Trans. Acoust. Speech Signal Processing, vol. ASSP-23, 1, 24-29.
- [17] Bakis, R., 1976, Continuous Speech Word Recognition Centisecond Acoustic States, in Proc. ASA Meeting. (WDC).
- [18] Jelinek, F., 1976, Continuous Speech Recognition by Stochastic Methods, Proc. IEEE, 64, 532-536
- [19] Jelinek, F., 1982, Continuous Speech Recognition : Stochastic Methods, in Handbook of Stastics, II, P.R. Krishnaiah, Ed. Amsterdam, The Netherlands: North-Holland.
- [20] Bahl, L. R. and Jelinek, F., 1983, A Maximum Likelihood Approach to Continuous Speech Recognition." IEEE Trans. Pattern Anal. Machine Intell., PAMI-5, 179-190.
- [21] Rabiner, L. R. and Wilpon, J. G., 1979, Speaker Independent, Isolated Word Recognition for a Moderate Size (54 word) Vocabulary, IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP--27., 6, 583-587
- [22] Juang, B. H. and Rabiner, L. R., 1985, A Probabilistic Distance Measure for Hidden Markov Models, AT&T Technical Journal, 64(2), 391-408.
- [23] Rabiner, L. R., Levinson, S. E. and Sondhi, M. M., 1983, On The Application of Vector Quantization and Hidden Markov Models to Speaker Independent Isolated Word Recognition, Bell Syst. Tech., 62 (4), 1075-1105.
- [24] Levinson, S. E., et al., 1983, Speaker Independent Isolated Digit Recognition Using Hidden Markov Models, Conference Record 1983 International

- Conference on Acoustics, Speech, and Signal Processing, Paper 22.8, 1049-1052.
- [25] Rabiner L. R., and Levinson, S. E., 1985, A Speaker Independent, Syntax Directed Connected Word Recognition System Based on Hidden Markov Models and Level Building, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33, No. 3, 561-573.
- [26] Rabiner, L. R., Juang, B. H., Levinson, S. E. and Sondhi, M. M., 1985, Some Properties of Continuous Hidden Markov Model Representations, *AT&T Tech. J.*, 64(2),391-408
- [27] Juang, B. H. and Rabiner, L. R., Mixture Autoregressive Hidden Markov Models for Speaker Independent Isolated Word Recognition, *Conference Record 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2.1, 41-44, 1986.
- [28] Rabiner L. R. and Juang, B. H., 1986, An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, 3 (1), 4-16.
- [29] Ephraim, Y. et al., 1987, A Minimum Discrimination Information Approach for Hidden Markov Modeling, *Conference Record 1987 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Paper 1.8.1, pp. 25-28.
- [30] Rabiner, L.R., Wipon, J.G., 1987, Application of Hidden Markov Models to Automatic Speech Endpoint Detection, *Computer Speech and Language*, 2 (3/4),701-714.
- [31] Rabiner, L. R., Wilpon, J. G., and Soong, F. K., 1988, High Performance Connected Digit Recognition, Using Hidden Markov Models, *Conference Record 1988 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Paper S3.6, 119-122.
- [32] Rabiner, L. R., Wilpon, J. G. and Soong, F. K., 1989, High Performance Connected Digit Recognition Using Hidden Markov Models, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 8, 1214-1225.
- [33] Rabiner, L. R., Lee, C. H., Juang, B. H. and Wilpon, J. G., 1989, HMM Clustering for Connected Word Recognition, *Conference Record 1989 IEEE International Conference on Acoustics, Speech, and Signal Processing*, S 8.5, 405-408.
- [34] Ljolje, A., Ephraim, Y. and Rabiner, L. R., 1990, Estimation of Hidden Markov Model Parameters By Minimizing Empirical Error Rate, *Conference Record 1990 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Paper S13.8, 709-712
- [35] Wilpon J. G. and Rabiner, L. R., 1985, A Modified K-means Clustering Algorithm for Use in Speaker Independent Isolated Word Recognition, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33 (3), 587-594.
- [36] Juang, B.H., Rabiner, L.R., 1990, The segmental K-means Algorithm for Estimating Parameters of Hidden Markov Models" *Acoustics, Speech and Signal Processing*, *IEEE Transactions*, 38, 9, 1639 – 1641.
- [37] David, S. et al, gpdsHMM: A Hidden Markov Model Toolbox in The Matlab Environment, Dpto. De Señales y Comunicaciones, Universidad de Las Palmas de Gran Canaria SPAIN
- [38] Viterbi, A. J., 1967, Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm, *IEEE Trans. on Information Theory*, 13(2), pp.260-269
- [39] Rabiner, L. R., 1989, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE*, 77 (2), 257–286.
- [40] Xuedong, H., et al., *Spoken Language Processing : A Guide to Theory, Algorithm, and System Development*, Printice-Hall NJ.
- [41] Rabiner, L. R., Levinson, S. E., May 1981, Isolated and Connected Word Recognition - Theory and Selected Applications, and *IEEE Trans. on Communications*, COM-29, (5), 621-659.
- [42] Davis S. and Mermelstein P., 1980, Comparison of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentences, *IEEE Trans. on Acoustics, Speech and Signal Processing*, 28(4), 357-366.
- [43] Rabiner, L., Juang, B. H., Yegnanarayana B., *Fundamentals of Speech Recognition*, Pearson Publication.
- [44] Swaranjali: Isolated Word Recognition for Hindi Language using VQ and HMM, 2000, *International conference on multimedia processing and systems (ICMPS)*.
- [45] Kumar, M., Rajput, N., 2005, Verma, A., Hybrid Baseform Builder for Phonetic Languages, *Proceedings of International Conference on Intelligent Sensing and Information Processing IEEE conference publications*, 382 – 386.
- [46] Kumar, M., 2004, A Large-Vocabulary Continuous Speech Recognition System For Hindi, *IBM Journal of Research and Development*, 48 (5.6), 703 – 715.
- [47] Agarwal, A., Jain, A., Prakash, N., Agrawal, S.S., 2010, Word Boundary Detection In Continuous Speech Based On Suprasegmental Features For Hindi Language, *2nd International Conference on Signal Processing Systems (ICSPPS)*, 2, V2-591 - V2-594.
- [48] Rajput, N. Adapting Phonetic Decision Trees Between Languages For Continuous Speech Recognition, 2010, *Sixth International Conference on Spoken Language Processing*, Beijing, China, 3, 850-852.
- [49] Sekhar C. C., Yegnanarayana B., 2000, A constraint satisfaction model for recognition of stop consonant-vowel (SCV) utterances, *IEEE Transactions on Speech and Audio Processing*, 10 (7), 472 – 480.
- [50] Poonam Bansal et al., 2008, Optimum HMM Combined with Vector Quantisation for Hindi Speech Word Recognition, *IETE journal of research*, 54 (4).
- [51] Lichi Yuan, 2008, An Improved HMM Speech Recognition Model, *ICALIP 2008. International*

Conference on Audio, Language and Image Processing, IEEE Conference Publications, 1311 – 1315.

- [52] Ranjan, S., 2010, A Discrete Wavelet Transform Based Approach to Hindi Speech Recognition”, International Conference on Signal Acquisition and Processing, IEEE Conference Publications.
- [53] Mehta,K., Anand,R.S., 2010, Robust Front-End and Back-End Processing for Feature Extraction for Hindi

Speech Recognition” IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 1 – 4.

- [54] Gupta, V.K., 2011, Speech Enhancement Using MMSE Estimation and Spectral Subtraction Methods International Conference on Devices and Communications (ICDeCom), IEEE Conference,1– 5.