# Automatic Speech Recognition Models: A Characteristic and Performance Review

**U. G. Patil[1], S. D. Shirbahadurkar[2], A. N. Paithane[3]**

[1, 3]Department of Electronics & Telecommunication, JSPM's Rajarshi Shahu College of Engineering, SPPU University, Pune, India.

[2]Department of Electronics & Telecommunication, Dr. D. Y. Patil college of Engineering, SPPU University, Pune, India.

[1]sakshi.paithane@gmail.com, [2]shirsd112@yahoo.in, [3]ajaypaithane@gmail.com

***Abstract*—This paper presents a review on few notable speech recognition models that are reported in the last decade. Firstly, the models are categorized into sparse models, learning models and domain – specific models. Subsequently, the characteristics of the models have been observed using speech constraints, algorithmic constraints and performance constraints. The performance of these models reported in the literature is investigated and the findings are summarized. Eventually, the research gaps revealed by the literature are discussed and the need for Hindi based speech recognition system is substantiated.**

***Keywords—speech; model; sparse; training; recognition; accuracy; Hindi***

## I. Introduction

Automatic Speech Recognition (ASR) is the primary component or subsystem of a human-computer interaction (HCI) system that requires human speech as decoded text input [1]. ASR is comprised of two major components, namely, acoustic model and language model [2]. The acoustic model models the pronunciation of a given word, whereas the language model predicts the likelihood of a given word sequence appearing in a language [2].

The components of an acoustic model can be the speech signal features and a pattern matching technique for a given word or phone [3], [65]. The term, 'phone' represents a basic unit of speech. A word may consist of one or more phones [2]. The most commonly used features of ASR is Mel-Frequency Cepstral Coefficients (MFCC) [4] and Perceptual Linear Predictive (PLP) Coefficients [5], whereas Hidden Markov Model (HMM) and neural network are the most commonly used pattern matching techniques [2] [6]. HMM understand the sequential nature of speech signal and model the output probability distribution as well as the state transition probability [7]. While recognizing the speech, various words are hypothesized against the acquired signal. HMM performs matching by determining the likelihood of a given word. The likelihood of a word is estimated based on the combination of likelihood of all the phones associated with the word [2]. Earlier, maximum likelihood (ML) estimation has been widely exploited to train HMM [8]. However, discriminative training has been found as more promising than ML in the later era [8] [9] [10] [11] [12].

On the other hand, the language model estimates the likelihood of a given word sequence appearing in the speech signal. N-gram language model is the most commonly used language model that predicts the probability of occurrence of a word in a sequence, when the history of word sequences is given. The probability calculation is based on a large text corpus given for training. Hypothetical word occurrences are handled by the scores obtained from acoustic and language models. Isolated word is recognized as the word, which has highest likelihood among the combined likelihoods of all the words [2]. However, Neural Network Language Models (NNLMs) have been proved to provide better performance than the conventional N-gram language models [13] [14] [15] [16] [17] [18]. Few hybrid models have also been used for improving the decoding process of ASR [19] [20].

## II. Noteworthy Research Contributions

### A. *Sparse Acoustic Models*

Weibin Zhang and Pascale Fung have given significant contributions on introducing sparse acoustic models. In 2013, they have proposed the sparse inverse covariance matrices for training acoustic models, despite the fact that the training data is insufficient. The traditional objective function for ML estimation has been improved by including L1 regularization term. The new objective function has sparsely satisfied the inverse covariance matrices and has trained the parameters of HMM using Expectation Maximization (EM) algorithm. The training process has been performed in a similar way as that of the ML estimation.

In 2014, they have focused on improving the discriminative training method by using acoustic models with sparse inverse covariance matrices, rather than using traditional full covariance matrices and diagonal covariance matrices. The sparse nature of the inverse covariance matrices has been accomplished by adding a lasso regularization term with the traditional objective function that attempts to extract maximum mutual information (MMI). The improved objective function has been subjected to maximization so that the training process can be performed. The experimental investigation on Wall Street Journal and Mandarin datasets has demonstrated that introducing sparse inverse covariance matrices has solved the over-fitting problem of discriminative training. They have also observed the ability of the methodology to regularize the model complexity and to improve the recognition accuracy. The observed results have proved that the acoustic models with sparse inverse covariance matrices are better than the conventional diagonal and full covariance models [1].

They have further introduced weighted lasso regularization to train the sparse banded acoustic models. They have attempted to reduce the computational speed by proposing the features that are able to minimize the band-width of sparse banded models. The sparse banded models have accomplished relatively improved results of about 9.5% and 15.1% over diagonal and full covariance models, respectively [34].

TABLE I: REVIEW OF THE SPEECH RECOGNITION MODELS BASED ON ITS CONSIDERATION OF SPEECH, ALGORITHM AND PERFORMANCE CONSTRAINTS

| **Speech Recognition Models** | | **MPE training of PMMs** | **Dynamic Features in the Linear and Linear-Log Hybrid Domains** | **Noise Adaptive Training** | **Sparse Inverse Covariance Matrices** | **Noise-Adaptive LDA** | **DT Sparse Inverse Covariance Matrices** | **Sparse Banded Acoustic Models** | **Temporally Varying Weight Regression** | **Derived Back-off Language Models** |
|---|---|---|---|---|---|---|---|---|---|---|
| Year [Citation] | | 2006 [8] | 2010 [26] | 2010 [35] | 2013 [24] | 2013 [25] | 2014 [1] | 2014 [34] | 2014 [7] | 2014 [19] |
| Speech constraints | Large vocabulary | ✓ | | | | | ✓ | | ✓ | ✓ |
| | Noisy environment | | | ✓ | ✓ | ✓ | | | ✓ | |
| | Reverberant environment | | ✓ | | | ✓ | | | | |
| | Speech resource requirement | High | High | High | Low | High | High | High | High | High |
| Performance constraints | WER | ✓ | | | ✓ | | ✓ | ✓ | ✓ | ✓ |
| | Word accuracy | | ✓ | ✓ | | ✓ | | | | |
| | CER/ P-metric | | | | ✓ | | | ✓ | | ✓ |
| ✓ - refers that the constraints are considered by the model | | | | | | | | | | |

## B. Adaptive Training Models

On the other end, researchers have also worked on reporting efficient training models and so the least error can be observed by solving the regular issues occurring while training the features extracted from the speech data. For instance, Khe Chai Sim and M.J.F. Gales [8] have studied the feasibility of using the precision matrix models for discriminative training in speech recognition systems. They have addressed the challenges ahead in constructing LVCSR systems as well as investigated the usage of minimum phone error criterion here. A generic framework has been used to approximate the precision matrices, in which numerous conventional models such as Semi-Tied Covariance (STC), Extended MLLT (EMLLT) and Subspace for Precision and Mean (SPAM) models have been included. Large vocabulary continuous telephone speech and broadcasting news in English are the tasks that have been used for experimental investigation.

In 2010, Ozlem Kalinli *et al* [35] have proposed a noise adaptive training (NAT) algorithm to handle the acoustic models under noisy environment. Rather than considering point estimates of the clean speech features, NAT has directly estimated the underlying "pseudo-clean" model parameters. The learnt pseudo-clean model parameters have been used with vector Taylor series (VTS) model adaptation for decoding noisy utterances at test time. Experiments have been conducted on Aurora 2 and Aurora 3 tasks, where substantial improvement of about 18.83% and 32.03%, respectively, have been accomplished by NAT over the traditional VTS model adaptation.

In 2014, Shilin Liu and KheChaiSim [7] have addressed the problem of using Standard HMM, which provides poor trajectory model for speech because of its assumption, "successive observations are independent to one another given the state sequence". Since, semi-parametric trajectory modeling techniques have proven its ability on handling large vocabulary continuous speech recognition tasks, they have focused on exploiting Temporally Varying Weight Regression (TVWR) to implicitly model HMM trajectory using time-varying Gaussian weights. In their work, they have provided a detailed formulation for Temporally Varying Weight Regression (TVWR) based on the probabilistic modelling framework. Parameters estimation has been accomplished based on ML and minimum phone error (MPE) criteria. Experimental results have asserted improved performance over the standard HMM systems under an environment containing 20k open vocabulary recognition task (NIST Nov'92 WSJ0) and 5k closed vocabulary noisy speech recognition task.

*C. Domain based Models*

The domain based models are the feature models that extract features from the given speech signal represent it in a different domain such as logarithmic domain, linear domain, etc. Few of the notable contributions include the feature models proposed by Osamu Ichikawa *et al* [26] in 2010 and Dorothea Kolossa *et al* [25] in 2013.

In [26], the problem of using logarithmic delta to represent the speech signals under reverberant environment has been explored. It has introduced schemes to determine the delta and the delta – delta features to diminish the reverberation effect on speech recognition system. Further, dynamic features have been introduced to adopt with their schemes. The experimental investigation has been carried out in Corpus and Environments for Noisy Speech RECognition-4 (CENSREC-4) database under reverberant environment. The results have demonstrated dominant error reduction, when the dynamic features are used rather than the conventional features.

Dorothea Kolossa *et al* [25] have introduced a strategy for ASR under natural (noisy) environment. According to their strategy, "Reducing the dimensionality of the speech feature for optimal discriminance under observation uncertainty can yield significantly improved recognition performance". They have also stated that the attempt can be achieved using Fisher's criterion of discriminant analysis.

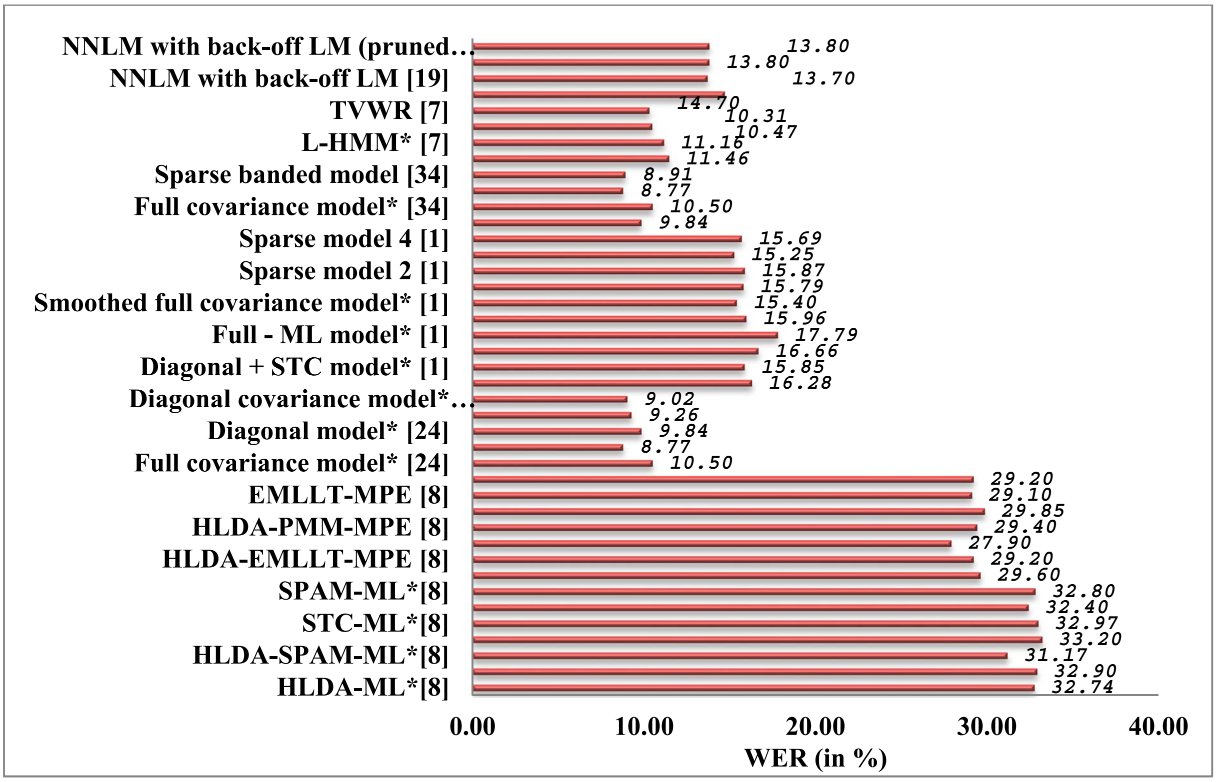


**Fig. 1:** WER VERSUS SPEECH RECOGNITION MODELS. WER IS AVERAGED OVER MULTIPLE EXPERIMENTAL SCENARIOS. *[REF] INDICATES THE STATE-OF-THE-ART RESULTS PUBLISHED IN THE PARTICULAR 'REF'

## III. ACOUSTIC MODELS ON PRACTICAL CONSTRAINTS

The models that are considered in our review have numerous peculiar characteristics and the ability to handle the practical constraints The constraints are categorized into three, namely, speech constraints, algorithmic constraints and performance constraints. The speech constraints are the practical challenges reside on handling a speech signal. For instance, noise and reverberant environment are common, when acquiring speech signal. Moreover, speech signal associated with a language has a large vocabulary and the models can have the multilingual support.

The algorithmic constraints are the challenges or limitations reside on the algorithms used in the models. Such constraints include computing efficiency, impact of initialization, overfitting or regularization problem and the impact of constants. The performance constraints are the metrics considered to explore the performance of the models. Renowned metrics are word error rate (WER) and word recognition accuracy. Apart from them, character error rate (CER) and perplexity (called further in this paper as P-metric) have also been considered by the researchers. The models are reviewed in the perspective of these constraints and the report is summarized in Table I.

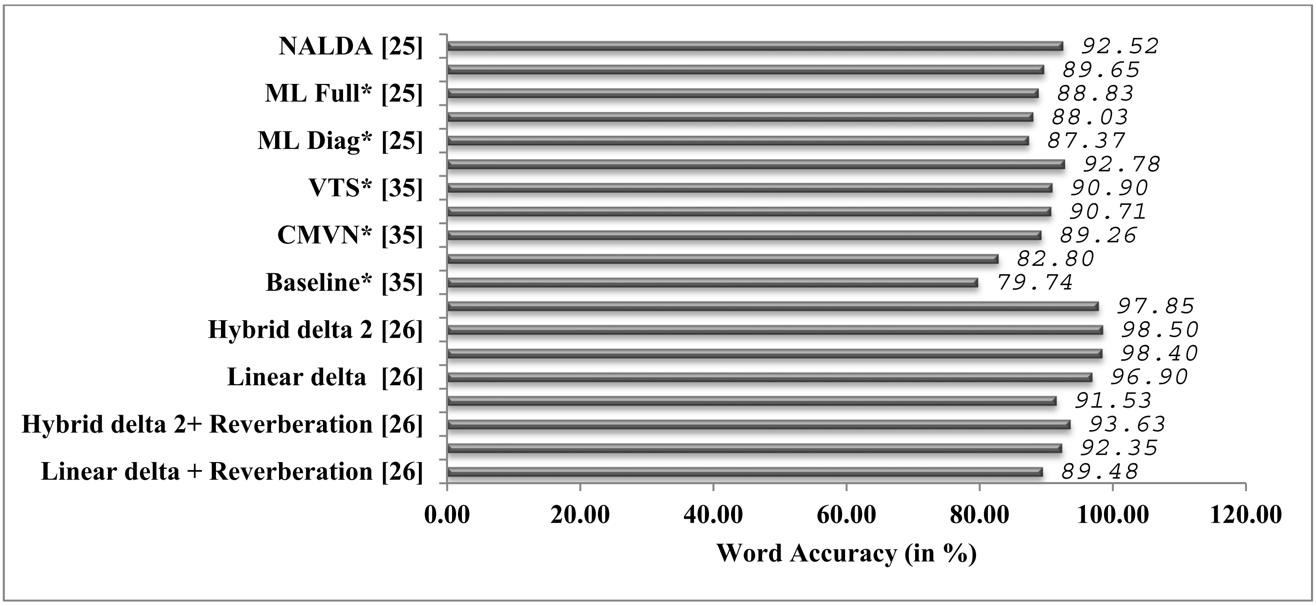


**Fig. 2:** WORD ACCURACY VERSUS SPEECH RECOGNITION MODELS. WER IS AVERAGED OVER MULTIPLE EXPERIMENTAL SCENARIOS. *[REF] INDICATES THE STATE-OF-THE-ART RESULTS PUBLISHED IN THE PARTICULAR 'REF'

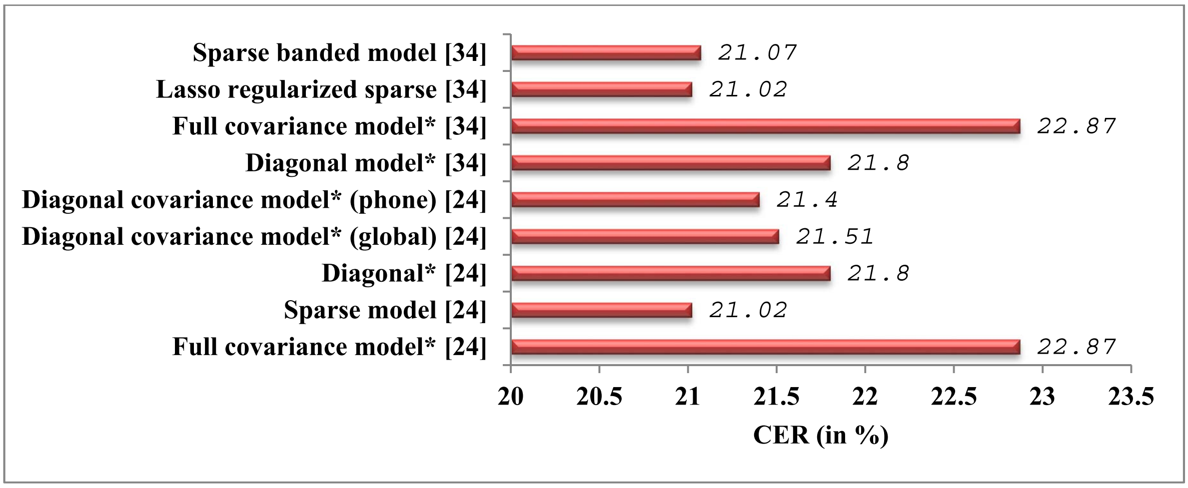


**Fig. 3:** CER VERSUS SPEECH RECOGNITION MODELS. WER IS AVERAGED OVER MULTIPLE EXPERIMENTAL SCENARIOS. *[REF] INDICATES THE STATE-OF-THE-ART RESULTS PUBLISHED IN THE PARTICULAR 'REF'

## IV. PERFORMANCE REVIEW

### A. *Word Error Rate (WER)*

Since the recognized word sequence and the reference word sequence express varying length, it is practically challenging to determine the performance. WER is promising on handling such problem using Levenshtein distance. The practical definition for WER can be mathematically given as

$$WER = \frac{S + 0.5D + 0.5I}{N} \qquad (1)$$

where, S, D, I and N refer to the number of substitutions, deletions, insertions and reference words, respectively.

### B. *Word Accuracy*

The word recognition accuracy or word accuracy is a derived form of WER. In other words, WER is a minimization function, whereas the word accuracy is a maximization function derived from WER. According to Fig. 2, the models [26] have accomplished more than 95% accuracy under no reverberant environment. Under reverberant environment, the word accuracy has reduced to 90% (approximated). The baseline model given in [35] has accomplished the least word accuracy of about 79%. The other models [25] [35] have reported 85-92 % accuracy (approximated).

### C. Character Error Rate (CER)

The CER exhibits the error rate in terms of character rather than word. Though CER is not well established like WER, it is considered in [24] and [34] for performance study. As per Fig. 3, the sparse model [24] and regularized sparse model [34] have accomplished least CER of about 21%, whereas full covariance models have achieved 23% (approximated).

## V. SPEECH RECOGNITION METHODOLOGIES FOR INDIAN LANGUAGES

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 staus. Speech Recognition on Hindi Language

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 [38] [40] gains considerable attntion from the researchers who are working in the Hindi speech recognition.

### A. Speech Recognition on Tamil, Bengali and Marathi Languages

Tamil, Bengali and Marathi languages are spoken by 80-90 million people approximately. It is just 20% of the Hindi speakers. HMM [45] [46] gains considerable attention in Tamil speech recognition in addition with trigram language model, dynamic time wrapping [44] and decision tree model [50]. Decision tree model [50] and DTW [51] are found to be promising for Bengali and Marathi speech recognition, respectively.

### B. Speech Recognition on Telugu, Kannada and Urdu Languages

Nearly, 50-75 million people speak Telugu, Kannada and Urdu langugaes. HTK [60] and ANN [59] have gained considerable attention from Telugu speech recognition attempts, whereas DWT [56] and SVM models [57] [58] are promising for Kannada speech recognition systems. Speech Recognition on Gujarati, Malayalam, Odiya and Punjabi Languages

### C. Speech Recognition on Assamese and Bodo Language

They are the least spoken languages among the previusly discussed Indian languages. Only 13 million people in India speak Assamese language and 1.5 million people speak Bodo. ANN models [48] have gained considerable interest on these languages, yet LVQ [49] has also been found suitable for Assamese language.

## VI. RESEARCH GAPS

### A. Global Challenges

Speech recognition system finds numerous applications in healthcare sectors, military, automobiles, commercial sectors, etc. This necessitates the robustness of the speech recognition system to be adopted for the applications. Recently, the research works on large-vocabulary continuous speech recognition (LVCSR) have been extended to spontaneous speech such as telephone conversations, ,lectures and meetings [21]. Though various advancements and improvements have been included in ASR, there are many grand challenges ahead, even in recent days [22] [23].

For instance, the performance of modern speech recognition systems is mainly decided by the volume of training data. However, it is not possible to collect sufficient data practically. In fact, collecting and transcribing such a huge amount of speech data for acoustic model training is complex as well as costly [24]. A deadlock always persists between the training of acoustic model and the speech database, because phonetically aligned speech database is the primary requirement for training an acoustic model. However, an acoustic model is mandatory for automatic alignment of speech database. Manually aligning a large speech database is infeasible because of its high time consumption and imprecision [24].

### B. Hindi Speech Recognition System: A Mandatory Prerequisite

In India, Hindi is the official language and the most widespread language, next to English. According to 2001 census report, 41% of Indians speak Hindi out of other Indian languages. The percentage is relatively higher than that of the other languages, because the second highly spoken language, Tamil and Bengali grab 8.1% [36]. Hence, the need for speech recognition systems in commercial as well as governmental applications is high.

## VII. CONCLUSION AND FUTURE WORK

This paper reviewed few significant contributions on the speech recognition models reported in the last decade. The review summary had revealed the three categories of the contributions, namely sparse models, adaptive learning models and domain based models. The models have been investigated under the renowned speech, algorithm and performance constraints that reveal the strengths and weaknesses of the models. Further, we have explored the performance of these models in terms of WER, word accuracy and CER.

The review has revealed important research gaps to be considered in the future research. The gaps have been presented under global and language specific challenges.

# References


[1] Weibin Zhang ; Fung, P., "Discriminatively Trained Sparse Inverse Covariance Matrices for Speech Recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 22, Issue 5, pp 873 - 882, 2014

[2] Kumar, M. ; Rajput, N. ; Verma, A., "A large-vocabulary continuous speech recognition system for Hindi", IBM Journal of Research and Development, Volume 48, Issue 5.6, pp 703 – 715, 2004

[3] Yongwon Jeong, "Adaptation of Hidden Markov Models Using Model-as-Matrix Representation", IEEE Transactions on Audio, Speech, and Language Processing, Volume 20, Issue 8, pp 2352 – 2364, 2012

[4] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustic, Speech and Signal Processing, Volume 28, Issue 4, pp. 357–366, 1980.

[5] H. Hermansky, "Perceptual Linear Predictive (PLP) analysis of speech," Journal of the Acoustic Society of America, Volume 87, Issue 4, pp. 1738– 1752, 1990.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, Volume 77, Issue 2, pp. 257–286, 1989.

[7] Shilin Liu ; Khe Chai Sim "Temporally Varying Weight Regression: A Semi-Parametric Trajectory Model for Automatic Speech Recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 22, Issue 1, pp 151 – 160, 2014

[8] K.C.SimandM. J.Gales, "Minimum phone error training of precision matrix models," IEEE Transactions on Audio, Speech, and Language Processing, Volume 14, Issue 3, pp. 882–889, 2006.

[9] G. Evermann, H. Y. Chan, M. J. F. Gales, T. Hain, X. Liu, D. Mrva, L. Wang, and P. C. Woodland, "Development of the 2003 CU-HTK conversational telephone speech transcription system," in Proc. ICASSP, 2004.

[10] D. Y. Kim, G. Evermann, T. Hain, D. Mrva, S. E. Tranter, L. Wang, and P. C. Woodland, "Recent advances in broadcast news transcription," in Proc. ASRU, pp. 105–110, 2003

[11] J. McDonough and A. Waibel, "Maximum mutual information speaker adapted training with semi-tied covariance matrices," in Proc. ICASSP, 2003

[12] V. Goel, S. Axelrod, R. Gopinath, P. Olsen, and K. Visweswariah, "Discriminative estimation of Subspace Precision and Mean (SPAM) models," in EUROSPEECH, 2003.

[13] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," J. Mach. Learn. Res., Volume 3, pp. 1137–1155, 2003

[14] [H. Schwenk and J.-L. Gauvain, "Training neural network language models on very large corpora," in Proc. HLT-EMNLP, pp. 201–208, 2005

[15] H. Schwenk, "Continuous space language models," Comput. Speech Lang., vol. 21, no. 3, pp. 492–518, Jul. 2007

[16] Hai-Son Le ; Oparin, I. ; Allauzen, A. ; Gauvain, J. ; Yvon, F., "Structured Output Layer Neural Network Language Models for Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, Volume 21, Issue 1, pp 197 – 206, 2013

[17] T.Mikolov,M.Karafiat, L. Burget, J. Cernocky, and S. Khudanpur, "Recurrent neural network based language model," in Proc. Inter-speech, pp. 1045–1048, 2010.

[18] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network language model," in Proc. ICASSP, pp. 5528–5531, 2011.

[19] Arisoy, E. ; Chen, S.F. ; Ramabhadran, B. ; Sethy, A., "Converting Neural Network Language Models into Back-off Language Models for Efficient Decoding in Automatic Speech Recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 22, Issue 1, pp 184 – 192, 2014

[20] E. Arisoy, S. F. Chen, B. Ramabhadran, and A. Sethy, "Converting neural network language models into back-off language models for efficient decoding in automatic speech recognition," in Proc. ICASSP, pp. 8242–8246, 2013.

[21] Akita, Y.; Kawahara, T., "Statistical Transformation of Language and Pronunciation Models for Spontaneous Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, Volume 18, Issue 6, pp 1539 - 1549, 2010

[22] J. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, and D. O'Shaughnessy, "Research Developments and directions in speech recognition and understanding, Part 1," IEEE Signal Processing Mag., vol. 26, no. 3, pp. 75–80, 2009.

[23] J. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, and D. O'Shaughnessy, "Updated MINDS report on speech recognition and understanding, Part 2," IEEE Signal Processing Mag., vol. 26, no. 4, pp. 78–85, 2009

[24] Weibin Zhang ; Fung, P., "Sparse Inverse Covariance Matrices for Low Resource Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, Volume 21, Issue 3, pp 659 – 668, 2013

[25] Kolossa, D. ; Zeiler, S. ; Saeidi, R. ; Fernandez Astudillo, R., "Noise-Adaptive LDA: A New Approach for Speech Recognition Under Observation Uncertainty", IEEE Signal Processing Letters, Volume 20, Issue 11, pp 1018 - 1021, 2013

[26] Ichikawa, O.; Fukuda, T.; Nishimura, M., "Dynamic Features in the Linear-Logarithmic Hybrid Domain for Automatic Speech Recognition in a Reverberant Environment", IEEE Journal of Selected Topics in Signal Processing, Volume 4, Issue 5, pp 816 - 823, 2010

[27] Rennie, S.J.; Hershey, J.R.; Olsen, P.A., "Single-Channel Multitalker Speech Recognition", IEEE Signal Processing Magazine, Volume 27, Issue 6, pp 66 – 80, 2010

[28] R. Rose, E. Hofstetter, and D. Reynolds, "Integrated models of signal and background with application to speaker identification in noise," IEEE Trans. Speech Audio Process., Volume 2, Issue 2, pp. 245–257, Apr. 1994.

[29] Hasan, T. ; Hansen, J.H.L., "Maximum Likelihood Acoustic Factor Analysis Models for Robust Speaker Verification in Noise", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 22, Issue 2, pp 381 - 391, 2014

[30] Hasan, T.; Hansen, J.H.L., "Acoustic Factor Analysis for Robust Speaker Verification", IEEE Transactions on Audio, Speech, and Language Processing, Volume 21, Issue 4, pp 842 – 853, 2013

[31] X. Fan and J. H. L. Hansen, "Speaker identification within whispered speech audio streams," IEEE Trans. Speech Audio Process., Volume 19, Issue 5, pp. 1408–1421, Jul. 2011.

[32] C. Zhang and J. H. L. Hansen, "Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing," IEEE Trans. Speech Audio Process., Volume 19,no.4, pp. 883–894, 2011

[33] Tanja S and Alex W, "Language-independent and language-adaptive acoustic modeling for speech recognition ", Speech Communication, Volume 35, Issue 1-2, pp 31-51, 2001

[34] Weibin Zhang ; Fung, P., "Efficient Sparse Banded Acoustic Models for Speech Recognition ", IEEE Signal Processing Letters, Volume 21, Issue 3, pp 280 – 283, 2014

[35] Kalinli, O. ; Seltzer, M.L. ; Droppo, J. ; Acero, A., "Noise Adaptive Training for Robust Automatic Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, Volume 18, Issue 8, pp 1889 – 1901, 2010

[36] 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.

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

[38] 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.

[39] 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.

[40] 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.

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

[42] 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.

[43] 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.

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

[45] 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.

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

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

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

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

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

[57] Sarika Hegde, Achary K.K., Surendra Shetty " 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.

[58] Sarika Hegde ,K K Achary,Surendra Shetty- Analysis of Isolated Word Recognition for Kannada Language using Pattern Recognition Approach-International Journal of Information Processing , 7(1), 73 - 80. -2013.

[59] 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.

[60] Vijai Bhaskar 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.

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

[62] 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) .

[63] 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.

[64] 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.

[65] Paithane, A.N.; Bormane, D.S., "Analysis of nonlinear and non-stationary signal to extract the features using Hilbert Huang transform," in Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on , vol., no., pp.1-4, 18-20 Dec. 2014

[66] Paithane, A.N.; Bormane, D.S., "Electrocardiogram signal analysis using empirical mode decomposition and Hilbert spectrum," in Pervasive Computing (ICPC), 2015 International Conference on , vol., no., pp.1-4, 8-10 Jan. 2015.

[67] Paithane, A. N., D. S. Bormane, and Sneha Dinde. "Human Emotion Recognition using Electrocardiogram Signals." International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), Volume 2, Issue 2,February 2014.