Bibliography

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., & Penn, G. (2012). Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4277–4280.

Abdelaziz, A. H., & Kolossa, D. (2016). General hybrid framework for uncertainty-decoding-based automatic speech recognition systems. Speech Communication, 79, 1–13.

Aggarwal, R. K., & Dave, M. (2011). Acoustic modeling problem for automatic speech recognition system: advances and refinements (Part II). International Journal of Speech Technology, 14(4), 309–320.

Aggarwal, R. K., & Dave, M. (2012). Integration of multiple acoustic and language models for improved Hindi speech recognition system. International Journal of Speech Technology, 15(2), 165–180.

Aggarwal, R. K., & Dave, M. (2015). An Empirical Approach for Optimization of Acoustic Models in Hindi Speech Recognition Systems. 8th International Conference on Natural Language Processing (ICON) , IIT Kharagpur, 1–9.

Aggarwal, R. K. & Dave, M. (2013). Performance evaluation of sequentially combined heterogeneous feature streams for Hindi speech recognition system. Telecommunication Systems, 52(3), 1457–1466.

Aggarwal, R. K. (2012). Improving Hindi speech recognition using filter bank optimization and acoustic model refinement. PhD Thesis, National Institute of Technology, Kurukshetra, Haryana, India.

Agrawal, S. S., Sinha, S., Singh, P., & Olsen, J. O. (2012). Development of Text and Speech database for Hindi and Indian English specific to Mobile Communication environment. Proceeding of International Conference on the Language Resources and Evaluation Conference, 3415–3421.

Alam, M. J., Kenny, P., & O’Shaughnessy, D. (2013). Low-variance Multitaper Mel-frequency Cepstral Coefficient Features for Speech and Speaker Recognition Systems. Cognitive Computation, 5(4), 533–544.

Alam, M. J., Kinnunen, T., Kenny, P., & Ouellet, P. (2013). Multitaper MFCC and PLP features for speaker verification using i-vectors. Speech Communication, 55(2), 237–251.

AlHanai, T., Hsu, W. N., & Glass, J. (2016). Development of the MIT ASR system for the 2016 Arabic multi-genre broadcast challenge. IEEE Spoken Language Technology Workshop (SLT), 299–304.

Allauzen, C., Riley, M., Schalkwyk, J., Skut, W., & Mohri, M. (2007). OpenFst: A general and efficient weighted finite-state transducer library. International Conference on Implementation and Application of Automata, 11–23.

Arora, S., Saxena, B., Arora, K., & Agrawal, S. S. (2010). Hindi ASR for Travel Domain. Proceedings of O-COCOSDA, 7(6), 233-241.

Atal, B. S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. The Journal of the Acoustical Society of America, 55(6), 1304–1312.

Aubert, X. L. (2002). An overview of decoding techniques for large vocabulary continuous speech recognition. Computer Speech & Language, 16(1), 89–114.

Baggenstoss, P. M. (2001). A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces. IEEE Transactions of Speech and Audio Processing, 9(4), 411–416.

Bansal, S., Sharan, S., & Agrawal, S. S. (2015). Corpus design and development of an annotated speech database for Punjabi. International Conference on Oriental COCOSDA held jointly with Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 32–37.

Beaufays, F., Sankar, A., Williams, S., & Weintraub, M. (2003). Learning linguistically valid pronunciations from acoustic data. Eighth European Conference on Speech Communication and Technology, 2593–2596.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. Journal of Machine Learning Research, 1089–1105.

Bharali, S. S., & Kalita, S. K. (2015). A comparative study of different features for isolated spoken word recognition using HMM with reference to Assamese language. International Journal of Speech Technology, 18(4), 673–684.

Bhowmik, T., Mukherjee, S., & Mandal, S. K. D. (2015). Detection of attributes for bengali phoneme in continuous speech using deep neural network. 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 103–108.

Bhuriyakorn, P., Punyabukkana, P., & Suchato, A. (2008). A genetic algorithm-aided hidden markov model topology estimation for phoneme recognition of thai continuous speech. Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, (SNPD), 475–480.

Boe, O., Krstulovic, S., Charlet, D., Fohr, D., & Mella, O. (2006). Optimizing the coverage of a speech database through a selection of representative speaker recordings, 48, 1319–1348.

Campbell, J. P. (1997). Speaker recognition: A tutorial. Proceedings of the IEEE, 85(9), 1437–1462.

Chen, T. Y., Mei, X. D., Pan, J. S., & Sun, S. H. (2004). Optimization of HMM by the tabu search algorithm. Journal of Information Science and Engineering, 20(5), 949–957.

Chen, X., & Cheng, J. (2014). Deep neural network acoustic modeling for native and non-native Mandarin speech recognition. 9th International Symposium on Chinese

Spoken Language Processing (ISCSLP), 6–9.
Cheng, J., Chen, X., & Metallinou, A. (2015). Deep neural network acoustic models for spoken assessment applications. Speech Communication, 73, 14–27.
Cheshomi, S., Rahati-Q, S., & Akbarzadeh-T, M. R. (2010). Hybrid of Chaos Optimization and Baum-Welch Algorithms for HMM Training in Continuous Speech Recognition. International Conference on Intelligent Control and Information Processing (ICICIP), 83–87.
Chiu, Y. H. B., & Stern, R. M. (2014). Minimum variance modulation filter for robust speech recognition. International Conference on Acoustics, Speech and Signal Processing, (ICASSP), 3917–3920.
Chourasia, V., Samudravijaya, K., & Chandwani, M. (2005). Phonetically Rich Hindi Sentence Corpus for Creation of Speech Database. Proc. O-Cocosda, 132–137.
Chunwijitra, V., Chotimongkol, A., & Wutiwiwatchai, C. (2016). A hybrid input-type recurrent neural network for LVCSR language modeling. Eurasip Journal on Audio, Speech, and Music Processing, (1), 1–12.
Crystal, D. (2000). Language Death. Cambridge University Press.
Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. Journal of the Acoustical Society of America, 24(6), 637–642.
Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366.
Day, P., & Nandi, A. K. (2007). Robust text-independent speaker verification using genetic programming. IEEE Transactions on Audio, Speech, and Language Processing, 15(1), 285–295.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1997). Maximum likliehood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 1–38.
Deng, L., & O'Shaughnessy, D. (2003). Speech processing: a dynamic and optimization-oriented approach. CRC Press.
Deng, L., Droppo, J., & Acero, A. (2005). Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion. IEEE Transactions on Speech and Audio Processing, 13(3), 412–421.
Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8599–8603.
Dey, A., Zhang, W., & Fung, P. (2014). Acoustic modeling for hindi speech recognition in low-resource settings. International Conference on Audio, Language and Image Processing (ICALIP), 891–894.
Dey, A., Lalhminghlui, W., Sarmah, P., Samudravijaya, K., Prasanna, S. M., Sinha, R.,

& Nirmala, S. R. (2018). Mizo Phone Recognition System. Interspeech 2018.
Doddington, G. R. (1985). Speaker recognition – identifying people by their voices. Proceedings of the IEEE, 73(11), 1651–1664.
Dong, Z., Guo, Y., & Zeng, J. (2011). Recurrent hidden Markov models using particle swarm optimization. International Journal of Modelling, Identification and Control, 14(4), 258–264.
Dua, M., Aggarwal, R. K., & Biswas, M. (2017). Discriminative Training using Heterogeneous Feature Vector for Hindi Automatic Speech Recognition System. International Conference on Computer and Applications (ICCA), 158–162.
Dua, M., Aggarwal, R. K., Kadyan, V., & Dua, S. (2012a). Punjabi automatic speech recognition using HTK. International Journal of Computer Science Issues, 9(4), 359–364.
Dua, M., Aggarwal, R. K., Kadyan, V., & Dua, S. (2012b). Punjabi speech to text system for connected words. Fourth International Conference on Advances in Recent Technologies in Communication and Computing, (ARTCom), 206–209.
Dua, M., Aggarwal, R. K., & Biswas, M. (2018). GFCC based discriminatively trained noise robust continuous ASR system for Hindi language. Journal of Ambient Intelligence and Humanized Computing, 1-14.
Dutta, K., & Sarma, K. K. (2012). Multiple feature extraction for RNN-based assamese speech recognition for speech to text conversion application. International Conference on Communications, Devices and Intelligent Systems (CODIS), 600–603.
Engelbrecht, A. P. (2007). Computational intelligence: an introduction. John Wiley& Sons. Inc., West Sussex, England.
Farooq, O. & Datta, S. (2010). Robustness evaluation of wavelet based features for continuous speech recognition. International Journal of Intelligent Systems Technologies and Applications, 9(1), 1–14.
Figielska, E., & Kasprzak, W. (2008). An evolutionary programming based algorithm for HMM training. Computational Intelligence: Methods and Applications, 166–175.
Forney, G. D. (1972). The Viterbi Algorithm. Proceedings of the IEEE, 61(3), 268–278.
Furui, S. (1981). Cepstral Analysis technique for automatic speaker verification. IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(2), 254–272.
Furui, S. (1986). Speaker independent isolated word recognition using dynamix features of speech spectrum. IEEE Transactions on Acoustics, Speech, and Signal Processing, 34(1), 52–59.
Gaida, C., Lange, P., Petrick, R., Proba, P., Malatawy, A., & Suendermann-Oeft, D. (2014). Comparing open-source speech recognition toolkits. Tech. Rep., DHBW Stuttgart.
Gales, M. J. F. (1995). Model-based techniques for noise robust speech recognition. University of Cambridge Dissertation.

Gautama, T., & Van Hulle, M. M. (1999). Self-Organized Feature Extraction Achieved with a Parameterized Filterbank. Neural Processing Letters, 10(2), 131–137.

Gelas, H., Abate, S. T., Besacier, L., & Pellegrino, F. (2011). Quality assessment of crowdsourcing transcriptions for African languages. Twelfth Annual Conference of the International Speech Communication Association, 3065–3068.

Gelly, G., & Gauvain, J. L. (2018). Optimization of RNN-Based Speech Activity Detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(3), 646–656.

Gemello, R., Mana, F., Scanzio, S., Laface, P., & De Mori, R. (2006). Adaptation of hybrid ANN/HMM models using linear hidden transformations and conservative training. IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP), 1189–1192.

Gen, M., & Cheng, R. (2002). Genetic algorithms and engineering optimization.

Georgescu, A. L., Cucu, H., & Burileanu, C. (2017). SpeeD's DNN Approach to Romanian Speech Recognition. International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 1–8.

Ghaffarzadegan, S., Bořil, H., & Hansen, J. H. (2017). Deep neural network training for whispered speech recognition using small databases and generative model sampling. International Journal of Speech Technology, 20(4), 1063–1075.

Ghai, W., & Singh, N. (2013). Phone based acoustic modeling for automatic speech recognition for Punjabi language. Journal of Speech Sciences, 1(3), 69–83.

Godambe, T., & Samudravijaya, K. (2011). Speech Data Acquisition for voice based Agricultural Information Retrieval. Proceeding of 39th All India DLA Conference,Punjabi University,Patiala, India, 1–8.

Goffin, V., Allauzen, C., Bocchieri, E., Hakkani-Tur, D., Ljolje, A., Parthasarathy, S., & Saraclar, M. (2005). The AT&T Watson speech recognizer. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP), 1033–1036.

Golik, P., Tüske, Z., Schlüter, R., & Ney, H. (2015). Convolutional neural networks for acoustic modeling of raw time signal in LVCSR. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 26–30.

Grézl, F., Karafiat, M., & Janda, M. (2011). Study of probabilistic and bottle-neck features in multilingual environment. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 359–364.

Guglani, J., & Mishra, A. N. (2018). Continuous Punjabi speech recognition model based on Kaldi ASR toolkit. International Journal of Speech Technology, 21(2), 1–6.

Hansson-Sandsten, M., & Sandberg, J. (2009). Optimal cepstrum estimation using multiple windows. IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP), 3077–3080.

Hansson-Sandsten, M. (2013). Mean square error optimal weighting for multitaper cepstrum estimation. EURASIP Journal on Advances in Signal Processing, (1), 158, 1-12.

Hansson, M., & Salomonsson, G. (1997). A multiple window method for estimation of peaked spectra. IEEE Transactions on Signal Processing1, 45(3), 778–781.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, 51–83.

Hasnat, M., Mowla, J., & Khan, M. (2007). Isolated and continuous bangla speech recognition: implementation, performance and application perspective. Proceedings of International Symposium on Natural Language Processing (SNLP), 1–6.

Hegde, S., Achary, K. K., & Shetty, S. (2012). Isolated word recognition for Kannada language using support vector machine. Wireless networks and computational intelligence, 262–269.

Hermansky, H., & Morgan, N. (1994). RASTA Processing of Speech. IEEE Transactions on Speech and Audio Processing, 2(4), 578–589.

Hermansky, H., Morgan, N., & Hirsch, H. G. (1993). Recognition of speech in additive and convolutional noise based on RASTA spectral processing. In IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP), 83–86.

Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. The Journal of the Acoustical Society of America, 87(4), 1738–1752.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82–97.

Holland, J. H. (1975). Adaptation in natural and artificial systems. Adaptation in Natural and Artificial Systems. Ann Arbor, MI: University of Michigan Press, 439-444.

Hong, Q. Y., & Kwong, S. (2003). A training method for hidden Markov model with maximum model distance and genetic algorithm. International Conference on Neural Networks and Signal Processing, 465–468.

Hsu, W. N., Zhang, Y., & Glass, J. (2016). A prioritized grid long short-term memory rnn for speech recognition. Spoken Language Technology Workshop (SLT), 467–473.

Hu, Y., & Loizou, P. C. (2004). Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Transactions on Speech and Audio Processing, 12(1), 59–67.

Huang, L. X., Evangelista, G., & Zhang, X. Y. (2011). Adaptive bands filter bank optimized by genetic algorithm for robust speech recognition system. Journal of Central South University of Technology, 18(5), 1595–1601.

Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). Spoken Language Processing : A Guide to Theory , Algorithm and System Development. Prentice Hall PTR

Upper Saddle River, NJ, USA.
Huemmer, C., Schwarz, A., Maas, R., Barfuss, H., Astudilló, R. F., & Kellermann, W. (2016). A new uncertainty decoding scheme for DNN-HMM hybrid systems with multichannel speech enhancement. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5760–5764.

Hung, J. W. (2004). Optimization of Filter·Bank to Improve The Extraction of MFCC Features in Speech Recognition. International Symposium on Intelligent Multimedia, Video and Speech Processing, 675–678.

Hwang, J. N., & Li, H. (1992). Interactive query learning for isolated speech recognition. Proceedings of the IEEE-SP Workshop, 93–102.

Imseng, D., Bourlard, H., & Garner, P. N. (2012). Using KL-divergence and multilingual information to improve ASR for under-resourced languages. IEEE International conference on acoustics, speech and signal processing, 4869–4872.

Imseng, D., Bourlard, H., Dines, J., & Garner, P. N. (2011). Improving non-native ASR through stochastic multilingual phoneme space transformations. Idiap.

Jayanna, H. S., & Prasanna, S. M. (2010). Limited data speaker identification. Sadhana, 35(5), 525–546.

Jayanna, H. S. (2009). Limited data speaker recognition. PhD Thesis, Indian Institute Of Technology Guwahati.

Jelinek, F. (1976). Continuous speech recognition by statistical methods. Proceedings of the IEEE, 64(4), 532–556.

Jiang, W., Chen, Y., Jin, H., Zheng, R., & Chi, Y. (2016). A novel GPU-based efficient approach for convolutional neural networks with small filters. Journal of Signal Processing Systems, 86(2–3), 313–325.

Jiang, H. (2010). Discriminative training of HMMs for automatic speech recognition: A survey. Computer Speech and Language, 24(4), 589–608.

Johansen, F. T. (2007). A comparison of hybrid HMM architecture using global discriminating training. Fourth International Conference on Spoken Language, (ICSLP), 498–501.

Karanasou, P., Wu, C., Gales, M., & Woodland, P. C. (2017). I-vectors and structured neural networks for rapid adaptation of acoustic models. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(4), 818–828.

Kay, S. M. (1988). Modern spectral estimation. Pearson Education India.

Khelifa, M. O., Elhadj, Y. M., Abdellah, Y., & Belkasmi, M. (2017). Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system. International Journal of Speech Technology, 20(4), 937–949.

Kim, C., & Stern, R. M. (2016). Power-normalized cepstral coefficients (PNCC) for robust speech recognition. IEEE Trans on Audio, Speech, and Language Processing, 24(7), 1315–1329.

Kingsbury, B. (2009). Lattice-based optimization of sequence classification criteria for neural-network acoustic modeling. IEEE International Conference on Acoustics,

Speech and Signal Processing, (ICASSP), 3761–3764.
Kinnunen, T., Saeidi, R., Sandberg, J., & Hansson-Sandsten, M. (2010). What else is new than the Hamming window? Robust MFCCs for speaker recognition via multitapering. Interspeech, 2734–2737.
Kipyatkova, I., & Karpov, A. (2015). Recurrent neural network-based language modeling for an automatic Russian speech recognition system. Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), 33–38.
Kipyatkova, I., & Karpov, A. (2016). DNN-Based Acoustic Modeling for Russian Speech Recognition Using Kaldi. International Conference on Speech and Computer, (SPECOM), (9811), 246–253.
Kiruthiga, S., & Krishnamoorthy, K. (2012). Design Issues in Developing Speech Corpus for Indian Languages – A survey. International Conference on Computer Communication and Informatics, 1–4.
Krishnamoorthy, P. (2009). Combined temporal and spectral processing methods for speech enhancement. Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India.
Kuamr, A., Dua, M., & Choudhary, A. (2014). Implementation and performance evaluation of continuous Hindi speech recognition. International Conference on In Electronics and Communication Systems (ICECS), 1–5.
Kuamr, A., Dua, M., & Choudhary, T. (2014). Continuous Hindi speech recognition using Gaussian mixture HMM. IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 1–5.
Kumar, K., & Aggarwal, R. K. (2011). Hindi Speech Recognition System using HTK. International Journal of Computing and Business Research, 2(2), 1–12.
Kumar, K., Aggarwal, R. K., & Jain, A. (2012). A Hindi speech recognition system for connected words using HTK. International Journal of Computational Systems Engineering, 1(1), 25–32.
Kumar, M., Rajput, N., & Verma, A. (2004). A large-vocabulary continuous speech recognition system for Hindi. IBM Journal of Research and Development, 48(5.6), 703–715.
Kumar, R., & Singh, M. (2011). Spoken isolated word recognition of Punjabi language using dynamic time warp technique. In: Singh C., Singh Lehal G., Sengupta J., Sharma D.V., Goyal V. (eds) Information systems for Indian languages, 139, 301.
Kumar, R., Kishore, S., Gopalakrishna, A., Chitturi, R., Joshi, S., Singh, S., & Sitaram, R. (2005). Development of Indian language speech databases for large vocabulary speech recognition systems. Proceedings of SPECOM, 1–6.
Kumar, Y., & Singh, N. (2017). An automatic speech recognition system for spontaneous Punjabi speech corpus. International Journal of Speech Technology, 20(2), 297–303.
Kumar, R. (2010). Comparison of HMM and DTW for isolated word recognition system

for punjabi language. International Journal of Soft Computing, 5(3), 88–92.

Kundu, S., Mantena, G., Qian, Y., Tan, T., Delcroix, M., & Sim, K. C. (2016). Joint acoustic factor learning for robust deep neural network based automatic speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5025–5029.

Kurian, C. (2015). A Review on Speech Corpus Development for Automatic Speech Recognition in Indian Languages. Int. J, Advanced Networking and Applications, 6(6), 2556–2558.

Kwong, S., Chau, C. W., Man, K. F., & Tang, K. S. (2001). Optimisation of HMM topology and its model parameters by genetic algorithms. Pattern Recognition, 34(2), 509–522.

Kwong, S., He, Q. H., Ku, K. W., Chan, T. M., Man, K. F., & Tang, K. S. (2002). A genetic classication error method for speech recognition. Signal Processing, 82(5), 737–748.

Kwong, S. & Chau, C. W. (1997). Analysis of parallel genetic algorithms on hmm based speech recognition system. IEEE Transactions on Consumer Electronics, 43(4), 1229–1233.

Lasserre, J. A., Bishop, C. M., & Minka, T. P. (2006). Principled Hybrids of Generative and Discriminative Models. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 87–94.

Lata, S., & Arora, S. (2013). Laryngeal tonal characteristics of Punjabi—An experimental study. IEEE International Conference on Human Computer Interactions (ICHCI), 1–6.

Lee, A., Kawahara, T., & Shikano, K. (2001). Julius - an Open Source Real-Time Large Vocabulary Recognition Engine. EUROSPEECH 7th European Conference on Speech Communication and Technology, 1691–1694.

Lee, C. H., Gauvain, J. L., Pieraccini, R., & Rabiner, L. R. (1993). Large vocabulary speech recognition using subword units. Speech Communication, 13(3–4), 263–279.

Lee, S. J., Kang, B. O., Chung, H., & Park, J. G. (2015). A useful feature-engineering approach for a LVCSR system based on CD-DNN-HMM algorithm. 23rd European Signal Processing Conference (EUSIPCO), 1421–1425.

Lee, S. M., Fang, S. H., Hung, J. W., & Lee, L. S. (2001). Improved MFCC feature extraction by PCA-optimized filter-bank for speech recognition. IEEE Workshop on Automatic Speech Recognition and Understanding, (ASRU), 49–52.

Li, B., & Sim, K. C. (2010). Comparison of discriminative input and output transformations for speaker adaptation in the hybrid NN/HMM systems. Eleventh Annual Conference of the International Speech Communication Association, 526–529.

Li, Q., & Huang, Y. (2010). Robust speaker identification using an auditory-based feature. IEEE International Conference on Acoustics Speech and Signal

Processing, (ICASSP), 4514–4517.

Li, Q., & Huang, Y. (2011). An Auditory-Based Feature Extraction Algorithm for Robust Speaker Identification Under Mismatched Conditions. IEEE Transactions on Audio, Speech, and Language Processing, 19(6), 1791–1801.

Li, Q., Soong, F. K., & Siohan, O. (2000). A high-performance auditory feature for robust speech recognition. Sixth International Conference on Spoken Language Processing, 51–54.

Li, Y., Feng, J., & Hu, J. (2016). Covariance and crossover matrix guided differential evolution for global numerical optimization. SpringerPlus, 5(1), 1–22.

Li, Y. X., Kwong, S., He, Q. H., He, J., & Yang, J. C. (2010). Genetic algorithm based simultaneous optimization of feature subsets and hidden Markov model parameters for discrimination between speech and non-speech events. International Journal of Speech Technology, 13(2), 61–73.

Li, Z., & Gao, Y. (2016). Acoustic feature extraction method for robust speaker identification. Multimedia Tools and Applications, 75(12), 7391–7406.

Liporace, L. (1982). Maximum likelihood estimation for multivariate observations of Markov sources. IEEE Transactions on Information Theory, 28(5), 729–734.

Liu, S., & Sim, K. C. (2014). On combining DNN and GMM with unsupervised speaker adaptation for robust automatic speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 195–199.

Maas, A. L., Qi, P., Xie, Z., Hannun, A. Y., Lengerich, C. T., Jurafsky, D., & Ng, A. Y. (2017). Building DNN acoustic models for large vocabulary speech recognition. Computer Speech and Language, 41, 195–213.

Macaš, M., Novák, D., & Lhotská, L. (2006). Constraints in particle swarm optimization of hidden Markov models. International Conference on Intelligent Data Engineering and Automated Learning, 1399–1406.

Maganti, H. K., & Matassoni, M. (2010). An Auditory Based Modulation Spectral Feature for Reverberant Speech Recognition. Eleventh Annual Conference of the International Speech Communication Association, 570–573.

Makhoul, J. (1975). Linear Prediction: A Tutorial Review. Proceedings of the IEEE, 63(4), 561–580.

Maldonado, Y. P., Morales, S. O. C., & Ortega, R. O. C. (2012). GA approaches to HMM optimization for automatic speech recognition. Mexican Conference on Pattern Recognition, 313–322).

Mandal, P., Jain, S., Ojha, G., & Shukla, A. (2015). Development of Hindi speech recognition system of agricultural commodities using deep neural network. Sixteenth Annual Conference of the International Speech Communication Association, 1241–1245.

Mantena, G. V., Rajendran, S., Rambabu, B., Gangashetty, S. V., Yegnanarayana, B., & Prahallad, K. (2011). A Speech-Based Conversation System for Accessing Agriculture Commodity Prices in Indian Languages. Joint Workshop on Hands-

Free Speech Communication and Microphone Arrays, 153–154.
Mao, R., Zhu, H., Zhang, L., & Chen, A. (2006). A new method to assist small data set neural network learning. Sixth International Conference on Intelligent Systems Design and Applications, (ISDA), 17–22.
Mason, J. S., & Zhang, X. (1991). Velocity and acceleration features in speaker recognition. International Conference on Acoustics, Speech, and Signal, (ICASSP), 3673–3676.
Mehla, R. & Aggarwal, R. K. (2013). Improving HMM Training Process Using Hybrid Baum-Welch Genetic Algorithm. International Conference on Emerging Research in Computing, Information, Communication and Applications (ERCICA), 1–7.
Meriem, F., Farid, H., Messaoud, B., & Abderrahmene, A. (2017). New front end based on multitaper and gammatone filters for robust speaker verification. In M. Chadli, S. Bououden, & I. Zelinka (Eds.), Recent Advances in Electrical Engineering and Control Applications, 411, 344–354.
Milner, B., & Darch, J. (2011). Robust acoustic speech feature prediction from noisy mel-frequency cepstral coefficients. IEEE Transactions on Audio, Speech, and Language Processing, 19(2), 338–347.
Minh, V. D., & Lee, S. (2004). PCA-based human auditory filter bank for speech recognition. International Conference on Signal Processing and Communications, (SPCOM), 393–397.
Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., Saltzman, E., & Tiede, M. (2017). Hybrid convolutional neural networks for articulatory and acoustic information based speech recognition. Speech Communication, 89, 103–112.
Mitra, V., Wang, W., Franco, H., Lei, Y., Bartels, C., & Graciarena, M. (2014). Evaluating robust features on Deep Neural Networks for speech recognition in noisy and channel mismatched conditions. Annual Conference of the International Speech Communication Association, INTERSPEECH, 895–899.
Mittal, S., & Sharma, R. K. (2014). Development of Phonetic Engine for Punjabi Language. M.Tech. Thesis, Thapar University, Punjab, India.
Mittal, T., & Sharma, R. K. (2016). An improved SVM using predator pre optimization and hook's jeeve method. Journal of Engineering Research, 4(1), 1–20.
Mittal, T. & Sharma, R. K. (2016). Integrated search technique for parameter determination of SVM for speech recognition. Journal of Central South University, 23(6), 1390–1398.
Mizuta, S., & Nakajima, K. (1992). A discriminative training method for continuous mixture density HMMs and its implementation to recognize noisy speech. Journal of the Acoustical Society of Japan (E), 13(6), 389–393.
Mohamed, A., & Nair, K. R. (2012). HMM/ANN hybrid model for continuous Malayalam speech recognition. Procedia Engineering, 30, 616–622.
Moon, S. & Hwang, J. N. (1997). Robust speech recognition based on joint model and feature space optimization of hidden Markov models. IEEE Transactions on Neural

Networks, 8(2), 194–204.

Morgan, N., & Hermansky, H. (1992). RASTA extensions: Robustness to additive and convolutional noise. Speech Processing in Adverse Conditions, 115–118.

Nahid, M. M. H., Purkaystha, B., & Islam, M. S. (2017). Bengali speech recognition: A double layered LSTM-RNN approach. 20th International Conference of Computer and Information Technology (ICCIT), 1–6.

Najkar, N., Razzazi, F., & Sameti, H. (2010). A novel approach to HMM-based speech recognition systems using particle swarm optimization. Mathematical and Computer Modelling, 52(11–12), 1910–1920.

Neal, R. M., & Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan, M.I. (Ed.), Learning in Graphical Models. Kluwer Academic Publishers.

Ney, H. (1984). The use of a one-stage dynamic programming algorithm for connected word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, (ICASSP), 188–196.

Ni, H., Yi, J., Wen, Z., Liu, B., & Tao, J. (2016). Improving accented Mandarin speech recognition by using recurrent neural network based language model adaptation. 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), 1–5.

P1: Punjabi Speech Corpus [http://tdil-dc.in/index.php?option=com_download&task=viewtoolsbyParam&categoryid=84&lang=en] available on 15 Oct 2015 10:30 p:m.

Patil, H. A., & Basu, T. K. (2008). Development of speech corpora for speaker recognition research and evaluation in Indian languages. International Journal of Speech Technology, 11(1), 17-32.

Palaz, D., Magimai.-Doss, M., & Collobert, R. (2015). Analysis of CNN-based speech recognition system using raw speech as input. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 11–15.

Pan, S. T., Chen, C. F., Chang, W. D., & Tsai, Y. H. (2011). Performances comparison between improved DHMM and Gaussian mixture HMM for speech recognition. 4th International Congress on Image and Signal Processing (CISP), 2426–2430.

Pandey, A., Srivastava, B. M. L., & Gangashetty, S. V. (2017). Adapting monolingual resources for code-mixed hindi-english speech recognition. International Conference on Asian Language Processing (IALP), 218–221.

Paramonov, P., & Sutula, N. (2016). Simplified scoring methods for HMM-based speech recognition. Soft Computing, 20(9), 3455–3460.

Parthasarathi, S. H. K., Hoffmeister, B., Matsoukas, S., Mandal, A., Strom, N., & Garimella, S. (2015). FMLLR based feature-space speaker adaptation of DNN acoustic models. Sixteenth Annual Conference of the International Speech Communication Association, 3630–3634.

Povey, D., Burget, L., Agarwal, M., Akyazi, P., Kai, F., Ghoshal, A., & Rose, R. C. (2011). The subspace Gaussian mixture model—A structured model for speech recognition. Computer Speech & Language, 25(2), 404–439.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., & Silovsky, J. (2011). The Kaldi speech recognition toolkit. In IEEE 2011 workshop on automatic speech recognition and understanding (No. EPFL-CONF-192584). IEEE Signal Processing Society.

Prahallad, K., Black, A. W., Kumar, R., & Sangal, R. (2003). Experiments with unit selection speech databases for Indian languages. National Seminar on Language Technology Tools: Implementation of Telugu, 1–6.

Prasanna, S. M. (2004). Event based analysis of speech. Indian Institute of Technology Madras, India.

Price, K. V. (1996). Differential evolution: a fast and simple numerical optimizer. Biennial Conference of the North American Fuzzy Information Processing Society, (NAFIPS), 524–527.

Pujol, P., Pol, S., Nadeu, C., Hagen, A., & Bourlard, H. (2005). Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system. IEEE Transactions on Speech and Audio Processing, 13(1), 14–22.

Qi, J., Wang, D., Jiang, Y., & Liu, R. (2013). Auditory features based on gammatone filters for robust speech recognition. IEEE International Symposium on Circuits and Systems (ISCAS), 305–308.

Quatieri, T. F. (2006). Discrete-time speech signal processing: principles and practice. Pearson Education.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286.

Rabiner, L. R., & Juang, B. H. (1993). Fundamentals of Speech Recognition. PTR Prentice Hall.

Radha, V. (2012). Speaker independent isolated speech recognition system for Tamil language using HMM. Procedia Engineering, 1097–1102.

Rajagede, R. A., & Dewa, C. K. (2017). Recognizing Arabic letter utterance using convolutional neural network. 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 181–186.

Ranjan, R., Singh, S. K., Shukla, A., & Tiwari, R. (2010). Text-dependent multilingual speaker identification for indian languages using artificial neural network. 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET), 632–635.

Rao, K. S. (2011). Application Prosody model for Developing speech system. International Journal of Speech Technology, 14(1), 19–33.

Ravanelli, M., Brakel, P., Omologo, M., & Bengio, Y. (2016). Batch-normalized joint

training for DNN-based distant speech recognition. Spoken Language Technology Workshop (SLT), 28–34.
Renjith, S., & Manju, K. G. (2017). Speech based emotion recognition in Tamil and Telugu using LPCC and hurst parameters—A comparitive study using KNN and ANN classifiers. International Conference on In Circuit, Power and Computing Technologies (ICCPCT), 1–6.
Renshaw, D., & Hall, K. B. (2015). Long short-term memory language models with additive morphological features for automatic speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5246–5250.
Reynolds, D. A. (2002). An overview of automatic speaker recognition technology. IEEE International Conference on Acoustics, Speech, and Signal Processing, 4, 4072–4075.
Reza, M., Rashid, W., & Mostakim, M. (2017). Prodorshok I: A Bengali Isolated Speech Dataset for Voice-Based Assistive Technologies - A comparative analysis of the effects of data augmentation on HMM-GMM and DNN classifiers. 5th IEEE R10 HTC, 1–4.
Riedel, K. S., & Sidorenko, A. (1995). Minimum bias multiple taper spectral estimation. IEEE Transactions on Signal Processing, 43(1), 188–195.
Rodríguez, L. J., & Torres, I. (2003). Comparative Study of the Baum-Welch and Viterbi Training Algorithms Applied to Read and Spontaneous Speech Recognition. Iberian Conference on Pattern Recognition and Image Analysis, 847–857.
Sahidullah, M., & Saha, G. (2012). Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication, 54(4), 543–565.
Samudravijaya, K., Rao, P. V. S., & Agrawal, S. S. (2000). Hindi Speech Database. Sixth International Conference on Spoken Language Processing (ICSLP), 456-459.
Sandberg, J., Hansson-Sandsten, M., Kinnunen, T., Saeidi, R., Flandrin, P., & Borgnat, P. (2010). Multitaper estimation of frequency-warped cepstra with application to speaker verification. IEEE Signal Processing Letters, 17(4), 343–346.
Schluter, R., Bezrukov, I., Wagner, H., & Ney, H. (2007). Gammatone features and feature combination for large vocabulary speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4, 649–652.
Schmidhuber, J. (2015). Deep learning in neural networks: an overview. Neural Networks, 61, 85–117.
Schwartz, R., Austin, S., Kubala, F., Makhoul, J., Nguyen, L., Placeway, P., & Zavaliagkos, G. (1992). New uses for the N-best sentence hypotheses within the BYBLOS speech recognition system. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP), 1–4.
Seide, F., Li, G., Chen, X., & Yu, D. (2011). Feature engineering in context-dependent

deep neural networks for conversational speech transcription. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 24–29.

Sarma, B. D., Sarmah, P., Lalhminghlui, W., & Prasanna, S. M. (2015). Detection of Mizo Tones. Sixteenth Annual Conference of the International Speech Communication Association (Interspeech), 934-937.

Siddhant, A., Jyothi, P., & Ganapathy, S. (2017). Leveraging Native Language Speech for Accent Identification using Deep Siamese Networks. CoRR, 1–8.

Singh, A., Pandey, D., & Agrawal, S. S. (2015). Analysis of Punjabi tonemes. 2nd International Conference on Computing for Sustainable Global Development (INDIACom), 1694–1697.

Song, X., Zou, Y., Huang, S., Chen, S., & Liu, Y. (2017). Investigating multi-task learning for automatic speech recognition with code-switching between mandarin and english. International Conference on Asian Language Processing (IALP), 27–30.

Stuttle, M. N., & Gales, M. J. F. (2002). Combining a Gaussian mixture model front end with MFCC parameters. 7th International Conference on Spoken Language Processing, 1565–1568.

Takamichi, S. (2017). Modulation spectrum-based speech parameter trajectory smoothing for DNN-based speech synthesis using FFT spectra. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 1308–1311.

Takara, T., Iha, Y., & Nagayama, I. (1998). Selection of the optimal structure of the continuous hmm using the genetic algorithm. Fifth International Conference on Spoken Language Processing, 1–4.

Thalengala, A., & Shama, K. (2016). Study of sub-word acoustical models for Kannada isolated word recognition system. International Journal of Speech Technology, 19(4), 817–826.

Thatphithakkul, N., & Kanokphara, S. (2004). HMM parameter optimization using tabu search. IEEE International Symposium on Communications and Information Technology, (ISCIT), 2, 904–908.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 70(9), 1055–1096.

Tomashenko, N., & Khokhlov, Y. (2015). GMM-derived features for effective unsupervised adaptation of deep neural network acoustic models. Sixteenth Annual Conference of the International Speech Communication Association, 2882–2886.

Tomashenko, N., Khokhlov, Y., & Estève, Y. (2016). A New Perspective on Combining GMM and DNN Frameworks for Speaker Adaptation. International Conference on Statistical Language and Speech Processing, 120–132.

Variani, E., Sainath, T. N., Shafran, I., & Bacchiani, M. (2016). Complex Linear Projection (CLP): A Discriminative Approach to Joint Feature Extraction and Acoustic Modeling. Interspeech, 808–812.

Veisi, H., & Sameti, H. (2013). Speech enhancement using hidden Markov models in Mel- frequency domain. Speech Communication, 52(2), 205–220.

Vesterstrom, J., & Thomsen, R. (2004). A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. Evolutionary Computation, (CEC), 1980–1987.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory, 13(2), 260–269.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., & Woelfel, J. (2004). Sphinx-4: A flexible open source framework for speech recognition. Sun Microsystems, Inc., Mountain View, CA, USA.

Wang, L., Minami, K., Yamamoto, K., & Nakagawa, S. (2010). Speaker identification by combining MFCC and phase information in noisy environments. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 4502–4505.

Xue, S., Abdel-Hamid, O., Jiang, H., Dai, L., & Liu, Q. (2014). Fast adaptation of deep neural network based on discriminant codes for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12), 1713–1725.

Yan, Z., Liu, P., Du, J., Soong, F., & Wang, R. (2006). Training Discriminative HMM by Optimal Allocation of Gaussian Kernels. International Symposium on Chinese Spoken Language Processing, 1–10.

Yang, F., Zhang, C., & Bai, G. (2008). A novel genetic algorithm based on tabu search for HMM optimization. Fourth International Conference on Natural Computation, (ICNC), 57–61.

Yang, F., Zhang, C., & Sun, T. (2008). Comparison of particle swarm optimization and genetic algorithm for HMM training. 19th International Conference on Pattern Recognition, (ICPR), 1–4.

Yao, K., Yu, D., Seide, F., Su, H., Deng, L., & Gong, Y. (2012a). Adaptation of context-dependent deep neural networks for automatic speech recognition. IEEE on Spoken Language Technology Workshop (SLT), 366–369.

Yi, J., Ni, H., Wen, Z., Liu, B., & Tao, J. (2016). CTC regularized model adaptation for improving LSTM RNN based multi-accent Mandarin speech recognition. 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), 1–5.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., & Valtchev, V. (2002). The HTK book. Cambridge University Engineering Department, vol . 3, 1-336.

Yu, D., & Deng, L. (2015). Automatic Speech Recognition - A Deep Learning Approach. Springer-Verlag London.

Zhang, X., Wang, Y., & Zhao, Z. (2008). A hybrid speech recognition training method for HMM based on genetic algorithm and Baum Welch algorithm. Second International Conference on Innovative Computing, Information and Control, (ICICIC), 25–28.

Zhao, T., Zhao, Y., & Chen, X. (2016). Ensemble Acoustic Modeling for CD-DNN-HMM Using Random Forests of Phonetic Decision Trees. Journal of Signal Processing Systems, 82(2), 187–196.

Zhao, X., & Wang, D. (2013). Analyzing noise robustness of MFCC and GFCC features in speaker identification. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 7204–7208.

Zolnay, A., Kocharov, D., Schlüter, R., & Ney, H. (2007). Using multiple acoustic feature sets for speech recognition. Speech Communication, 49(6), 514–525.

Zolnay, A., Schluter, R., & Ney, H. (2005). Acoustic feature combination for robust speech recognition. IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP), 457–460.

Zouhir, Y., & Ouni, K. (2014). A bio-inspired feature extraction for robust speech recognition. SpringerPlus, 1–8.