

References

- [1] H. Xuedong, A. Alex, H. Hsiao-Wuen, and R. Raj, *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice Hall PTR Upper Saddle River, 2001.
- [2] S. Karpagavalli and E. Chandra, “A Review on Automatic Speech Recognition Architecture and Approaches,” *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 9, no. 4, pp. 393–404, 2016, doi: 10.14257/ijsp.2016.9.4.34.
- [3] D. Jurafsky and H. M. James, *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Pearson/Prentice Hall, 2009.
- [4] S. Y. N.Wada, Y. Miyana, and N. Yoshida, “A consideration about an extraction of features for isolated word speech recognition in noisy environments,” in *in ISPACS2002, DSP2002-33*, 2002, pp. 19–22.
- [5] Y. W. Chit, “Review on Different Approaches for Continuous Speech Recognition System,” *Int. J. Sci. Res.*, vol. 6, no. 2, pp. 2051–2055, 2017.
- [6] B. H. Juang and L. R. Rabiner, “Automatic Speech Recognition – A Brief History of the Technology Development - Juang, Rabiner - Unknown.pdf,” 2004.
- [7] R. Haridy, “Microsoft’s speech recognition system is now as good as a human,” 2017. .
- [8] V. Radha and Vimala, “A review on speech recognition challenges and approaches,” *doaj. org*, 2(1), pp. 1–7, 2012.
- [9] M. Forsberg, “Why is Speech Recognition Difficult?,” in *Chalmers University of Technology.*, 2003, pp. 1–10.
- [10] X. Wang and D. O’Shaughnessy, “Improving the Efficiency of Automatic Speech Recognition by Feature Transformation And Dimensionality Reduction,” 2003.
- [11] S. Shahnawazuddin, B. A. N. Adiga, S. Tarun, A. Waquar, and K. K. Hemant, “Developing speaker independent ASR system using limited data through prosody modification based on fuzzy classification of spectral bins,” *Digit. Signal Process.*, vol. 93, pp. 34–42, 2019.
- [12] F. Chen and C. Masi, “Effect of Noise on Automatic Speech Recognition System Error Rate,” 2000.
- [13] T. Valenta and L. Šmídl, “On the Impact of Sentence Length on Recognition

- Accuracy,” in *12th International Conference on Signal Processing (ICSP)*, 2014, pp. 500–504.
- [14] M. A. Siegler and M. R. Stern, “On the Effect of Speech Rate in Large Vocabulary Speech Recognition Systems,” in *International conference on acoustics, speech, and signal processing*, 1995, pp. 612–615.
- [15] C. Vimala and V. Radha, “Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 378–383, 2014.
- [16] D. Gupta, P. Bansal, and K. Choudhary, “The state of the art of feature extraction techniques in speech recognition,” in *Speech and language processing for human-machine communications*, 2018, pp. 195–207.
- [17] P. Saini and P. Kaur, “Automatic Speech Recognition: A Review,” *Int. J. Eng. Trends Technol.*, vol. 4, pp. 132–136, 2013.
- [18] P. V Janse, S. B. Magre, P. K. Kurzekar, and R. R. Deshmukh, “A Comparative Study between MFCC and DWT Feature Extraction Technique,” *Int. J. Eng. Res. Technol.*, vol. 3, no. 1, pp. 3124–3127, 2014.
- [19] N. Upadhyay and H. G. Rosales, “Robust Recognition of English Speech in Noisy Environments Using Frequency Warped Signal Processing,” *Natl. Acad. Sci. Lett.*, vol. 41, no. 1, pp. 15–22, 2018.
- [20] C. Kim and R. M. Stern, “Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition,” *IEEE/ACM Trans. audio, speech, Lang. Process.*, vol. 24, no. 7, pp. 1315–1329, 2016.
- [21] G. Y. Li and X. Wang, “Modified MFCC Methods Based on Differential Power Spectrum and Power Law for Robust Speech Recognition,” in *Computer Science and Technology: Proceedings of the International Conference*, 2017, pp. 696–702.
- [22] P. Harshita and A. R. Adiga, “Speech Recognition with Frequency Domain Linear Prediction,” in *International Conference on Communication and Signal Processing*, 2018, pp. 630–634.
- [23] S. Nayak, D. B. Shashank, S. Bhati, K. Bramhendra, and K. S. Murty, “Instantaneous Frequency Features for Noise Robust Speech Recognition,” in *National Conference on Communications (NCC)*, 2019, pp. 1–5.
- [24] K. K. Tomchuk, “Spectral Masking in MFCC Calculation for Noisy Speech,” in *Proc. of Wave Electronics and its Application in Information and Telecommunication Systems*, 2018, pp. 1–4.
- [25] A. Garg and P. Sharma, “Survey on Acoustic Modeling and Feature Extraction for Speech Recognition,” in *3rd International Conference on*



- Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 2291–2295.
- [26] S. A. Alim and N. K. Rashid, “Some commonly used speech feature extraction algorithms,” *From Nat. to Artif. Intell. Appl.*, 2018.
- [27] H. K. Maganti and M. Matassoni, “A perceptual masking approach for noise robust speech recognition,” *EURASIP J. Audio, Speech, Music Process.*, vol. 1, p. 29, 2012.
- [28] Y. Lin and W. H. Abdulla, “Principles of psychoacoustics. In Audio Watermark,” Springer, Cham, 2015, pp. 15–49.
- [29] M. Amazin, A. Gouda, and M. Khedr, “Enhanced Automatic Speech Recognition System Based on Enhancing Power-Normalized Cepstral Coefficients,” *Appl. Sci.*, vol. 9, no. 10, pp. 1–15, 2019.
- [30] R. Hidayat, A. Bejo, S. Sumaryono, and A. Winursito, “Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System,” in *10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2018, pp. 280–284.
- [31] Y. Tian, J. Tang, X. Jiang, H. Tsutsui, and Y. Miyanaga, “Accuracy on Children’s Speech Recognition under Noisy Circumstances,” in *18th International Symposium on Communication and Information Technology*, 2018, pp. 101–104.
- [32] A. Dixit, A. Vidwans, and P. Sharma, “Improved MFCC and LPC Algorithm for Bundelkhandi Isolated Digit Speech Recognition,” in *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 3755–3759.
- [33] H. Rahali, Z. Hajaiej, and N. Ellouze, “Asr systems in noisy environment: Auditory features based on gammachirp filter using the AURORA database,” in *22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 696–700.
- [34] J. Qi, D. Wang, Y. Jiang, and R. Liu, “Auditory features based on gammatone filters for robust speech recognition,” in *International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 305–308.
- [35] A. Kaur and A. Singh, “Power-Normalized Cepstral Coefficients (PNCC) for Punjabi Automatic Speech Recognition using phone based modelling in HTK,” in *2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT*, 2017, pp. 372–375.
- [36] A. Hazrat, Z. Xianwei, and T. Sun, “Comparison of MFCC and DWT features for automatic speech recognition of Urdu,” in *International Conference on Cyberspace Technology*, 2014, pp. 154–158.



- [37] M. A. Imtiaz and G. Raja, "Isolated word Automatic Speech Recognition (ASR) System using MFCC, DTW & KNN," in *Multimedia and Broadcasting (APMediaCast), 2016 Asia Pacific Conference*, 2016, pp. 106–110.
- [38] K. Aida-zade, A. Xocayev, and S. Rustamov, "Speech recognition using Support Vector Machines," in *International Conference on Application of Information and Communication Technologies (AICT)*, 2016, pp. 1–4.
- [39] A. L. Georgescu, H. Cucu, and C. Burileanu, "Speed's DNN approach to Romanian speech recognition," in *International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2017, pp. 1–8.
- [40] M. M. Nahid, B. Purkaystha, and M. S. Islam, "Bengali speech recognition: A double layered LSTM-RNN approach," in *20th International Conference of Computer and Information Technology (ICCIIT)*, 2017, pp. 1–6.
- [41] D. Gupta and K. Gupta, "An analysis on LPC, RASTA and MFCC techniques in Automatic Speech recognition system," in *Conference of Cloud System and Big Data Engineering (Confluence)*, 2016, pp. 493–497.
- [42] S. K. Rao, R. V. Reddy, and S. Maity, *Language Identification Using Spectral and Prosodic Features*. 2015.
- [43] M. M. Pour and F. Farokhi, "An advanced method for speech recognition," 2009.
- [44] G. Athiramenon and V. K. Anjusha, "Analysis of Feature Extraction Methods for Speech Recognition," *Int. J. Innov. Sci. Eng. Technol.*, vol. 4, no. 4, 2017.
- [45] H. Gupta and D. Gupta, "LPC and LPCC method of feature extraction in Speech Recognition System," in *6th International Conference-Cloud System and Big Data Engineering*, 2016, pp. 498–502.
- [46] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp. 1738–1752, 1990.
- [47] D. Alam, M. J. Kinnunen, T. Kenny, P. Ouellet, and P. O'Shaughnessy, "Multitaper MFCC and PLP features for speaker verification using i-vectors," *Speech Commun.*, vol. 55, no. 2, pp. 237–251, 2013.
- [48] N. Dave, "Feature extraction methods LPC, PLP and MFCC in speech recognition," *Int. J. Adv. Res. Eng. Technol.*, vol. 1, no. 6, pp. 1–4, 2013.
- [49] M. D. Narang and S. Gupta, "Speech Feature Extraction Techniques: A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 4, no. 3, pp. 107–114, 2015.
- [50] K. M. Indrebo, R. J. Povinelli, and M. T. Johnson, "Minimum Mean-Squared



- Error Estimation of Mel-Frequency Cepstral Coefficients using a Novel Distortion Model,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 16, no. 8, pp. 1654–1661, 2008.
- [51] Z. S. Al-Timime, “Signal Denoising Using Double Density Discrete Wavelet Transform,” *Al-Nahrain J. Sci.*, vol. 20, no. 4, pp. 125–129, 2017.
- [52] M. S. G., R. M. Ivanov, and A. N. Popov, “Denoising Speech Signals by Wavelet Transform,” *Annu. J. Electron.*, pp. 712–715, 2009.
- [53] A. S. Raj, D. H. Oliver, Y. Srinivas, and J. Viswanath, “Wavelet denoising algorithm to refine noisy geoelectrical data for versatile inversion,” *Model. Earth Syst. Environment*, vol. 2, no. 1, pp. 1–11, 2016.
- [54] M. Srivastava, C. L. Anderson, and J. H. Freed, “A New Wavelet Denoising Method for Selecting Decomposition Levels and Noise Thresholds,” *IEEE Access*, vol. 4, pp. 3862–3877, 2016.
- [55] I. W. Selesnick, “The Double Density DWT,” in *Petrosian A.A., Meyer F.G. (eds) Wavelets in Signal and Image Analysis. Computational Imaging and Vision, vol 19.*, Springer, Dordrecht, 2001.
- [56] S. B. Magre and R. R. Deshmukh, “Design and Development of Automatic Speech Recognition of Isolated Marathi Words for Agricultural Purpose,” *IOSR J. Comput. Eng.*, vol. 16, no. 3, pp. 79–85, 2014.
- [57] H. Hermansky and N. Morgan, “RASTA processing of speech,” *IEEE Trans. speech audio Process.*, vol. 2, no. 4, pp. 578–89, 1994.
- [58] W. Ghai and N. Singh, “Literature review on automatic speech recognition,” *Int. J. Comput. Appl.*, vol. 41, no. 8, pp. 42–50, 2012, doi: 10.5120/5565-7646.
- [59] M. M. Pour and F. Farok, “An advanced method for speech recognition,” *World Acad. Sci. Eng. Technol.*, vol. 49, 2009.
- [60] D. R. Reddy, “Speech Recognition by Machine: A Review,” 1990, pp. 8–38.
- [61] R. L. K. Venkateswarlu, R. V. Kumari, and G. V. JayaSri, “Speech Recognition By Using Recurrent Neural Networks,” *Int. J. Sci. Eng. Res.*, vol. 2, no. 6, 2011.
- [62] D. C. de Andrade, “Recognizing Speech Commands Using Recurrent Neural Networks with Attention,” 2018. .
- [63] J. E. i Gelabert, “Exploring Automatic Speech Recognition with TensorFlow,” 2018.
- [64] A. Amberkar, P. A. G. Deshmukh, and P. Dave, “Speech Recognition using Recurrent Neural Networks,” in *IEEE International Conference on Current*



Trends toward Converging Technologies, 2018, pp. 1–4.

- [65] D. Britz, “Recurrent Neural Networks Tutorial,” 2015.
- [66] M. Liu, Y. Wang, J. Wang, and X. Xie, “Speech Enhancement Method Based on LSTM Neural Network for Speech Recognition,” in *14th IEEE International Conference on Signal Processing (ICSP)*, 2018, pp. 245–249.
- [67] D. Povey, A. Ghoshal, G. Boulianne, and L. Burget, “The Kaldi speech recognition toolkit,” 2011.
- [68] C. Gaida, P. Lange, R. Petrick, P. Proba, A. Malatawy, and D. Suendermann-Oeft, “Comparing open-source speech recognition toolkits,” 2014.
- [69] D. Pearce and H. G. Hirsch, “The AURORA Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions,” 2000.
- [70] R. Leonard, “A database for speaker-independent digit recognition,” in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1984, pp. 328–331.
- [71] R. F. B.-R. X. Huang, A. Acero, H. W. Hon, “Spoken Language Processing: A Guide to Theory, Algorithm, and System Development,” Prentice hall PTR, 2001.
- [72] T. C. A. Kuamr, M. Dua, “Continuous Hindi Speech Recognition using Gaussian Mixture HMM,” in *IEEE Students’ Conference on Electrical, Electronics and Computer Science*, 2014, pp. 1–5.
- [73] EDUCAB, “Recurrent Neural Networks (RNN).” .
- [74] S. Desai, P. D. Khandekar, and K. J. Raut, “2-D psychoacoustic modeling for automatic speech recognition in noisy environment,” in *Conference on Advances in Signal Processing (CASP)*, 2016, pp. 129–132.
- [75] D. Naveen and A. Jhansi, “Implementation of Psychoacoustic Model in Audio Compression using Munich and Gammachirp Wavelets,” *Int. J. Eng. Sci. ...*, vol. 2, no. 5, pp. 1066–1072, 2010, [Online]. Available: <http://www.doaj.org/doaj?func=abstract&id=707372>.
- [76] P. Dai and Y. Soon, “A temporal frequency warped (TFW) 2D psychoacoustic filter for robust speech recognition system,” *Speech Commun.*, vol. 54, no. 3, pp. 402–413, 2012.
- [77] T. S. Gunawan, “Audio Compression and Speech Enhancement Using Temporal Masking Models,” 2007.
- [78] S. Sahoo and A. Routray, “MFCC feature with optimized frequency range: An essential step for emotion recognition,” in *International Conference on*



Systems in Medicine and Biology (ICSMB), 2017, pp. 162–165.

- [79] S. Iqbal, T. Mahboob, and M. S. H. Khiyal, “Voice Recognition using HMM with MFCC for Secure ATM,” *Int. J. Comput. Sci. Issues*, vol. 8, no. 6, pp. 297–303, 2011.
- [80] L. Tan and M. Karnjanadecha, “Modified Mel-Frequency Cepstrum Coefficient,” in *Proceedings of the Information Engineering Postgraduate Workshop*, 2003, pp. 127–130.
- [81] M. Russo, M. Stella, M. Sikora, and V. Pekić, “Robust cochlear-modelbased speech recognition,” *Computers*, vol. 8, no. 1, 2019.
- [82] G. K. Liu, “Evaluating gammatone frequency cepstral coefficients with neural networks for emotion recognition from speech,” *arXiv Prepr. arXiv1806.09010*, 2018.
- [83] J. Qi, D. Wang, Y. Jiang, and R. Liu, “Auditory features based on Gammatone filters for robust speech recognition,” in *IEEE International Symposium on Circuits and Systems*, 2013, pp. 305–308.
- [84] Y. Yang, Y. Jiang, R. Liu, and D. Li, “A realtime analysis/synthesis Gammatone filterbank,” in *International Conference on Signal Processing, Communications and Computing, ICSPCC*, 2015, pp. 1–6.
- [85] O. Cheng, W. Abdulla, and Z. Salcic, “Performance evaluation of front-end processing for speech recognition systems,” *The University of Auckland*. 2005.
- [86] B. Gerazov and Z. Ivanovski, “Influence of the filter bank on automatic speech recognition system performance in noise,” 2013.
- [87] T. T. Swee, S. H. S. Salleh, and M. R. Jamaludin, “Speech Pitch Detection Using Short-Time Energy,” in *International Conference on Computer and Communication Engineering (ICCCE)*, 2010, pp. 1–6.
- [88] R. Rehr and T. Gerkmann, “Cepstral noise subtraction for robust automatic speech recognition,” in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 375–378.
- [89] O. Strand and A. Egeberg, “Cepstral mean and variance normalization in the model domain,” in *COST278 and ISCA Tutorial and Research Workshop (ITRW) on Robustness Issues in Conversational Interaction*, 2004, pp. 2–5.
- [90] P. Upadhyaya, S. K. Mittal, Y. V. Varshney, O. Farooq, and M. R. Abidi, “Speaker adaptive model for hindi speech using Kaldi speech recognition toolkit,” in *International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, 2017, pp. 222–226.
- [91] P. Voigtlaender, P. Doetsch, S. Wiesler, R. Schluter, and H. Ney, “Sequence-



discriminative training of recurrent neural networks,” in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2100–2104.

- [92] V. V. Vegesna, K. Gurugubelli, H. K. Vydana, B. Pulugandla, M. Shrivastava, and A. K. Vuppala, “DNN-HMM acoustic modeling for large vocabulary Telugu speech recognition,” in *International Conference on Mining Intelligence and Knowledge Exploration*, 2017, pp. 189–197.
- [93] K. Vesely, A. Ghoshal, and D. Povey, “Karel’s DNN implementation:Kaldi Toolkit.” .
- [94] K. Veselý, A. Ghoshal, L. Burget, and D. Povey, “Sequence-discriminative training of deep neural networks,” in *Interspeech*, 2013, pp. 2345–2349.