

# A Perspective Study on Speech Recognition

Sana Fatema N. Ali<sup>1</sup> and Prof. S. T. Khandare<sup>2</sup>

ME Scholer, Babasaheb Naik College of Engineering, Pusad, Maharashtra, India<sup>1</sup>

Associate Professor, Babasaheb Naik College of Engineering, Pusad, Maharashtra, India<sup>2</sup>

**Abstract:** Emotions play an extremely important role in human mental life. It is a medium of expression of one's perspective or one's mental state to others. Speech Emotion Recognition (SER) can be defined as extraction of the emotional state of the speaker from his or her speech signal. There are few universal emotions including Neutral, Anger, Happiness, and Sadness in which any intelligent system with finite computational resources can be trained to identify or synthesize as required. In this work spectral and prosodic features are used for speech emotion recognition because both of these features contain the emotional information. Mel-Frequency Cepstral Coefficients (MFCC) is one of the spectral features. Fundamental frequency, loudness, pitch and speech intensity and glottal parameters are the prosodic features which are used to model different emotions. The potential features are extracted from each utterance for the computational mapping between emotions and speech patterns. Pitch can be detected from the selected features, using which gender can be classified. The audio signal is filtered using a method known as feature extraction technique. In this article, the feature extraction technique for speech recognition and voice classification is analyzed and also centered to comparative analysis of different types of Mel frequency cepstral coefficients (MFCC) feature extraction method. The MFCC technique is used for deduction of noise in voice signals and also used for voice classification and speaker identification. The statistical results of the different MFCC techniques are discussed and finally concluded that the delta-delta MFCC feature extraction technique is better than the other feature extraction techniques..

**Keywords:** Digital Signal Processing, MATLAB, Machine Learning, voice modulation

## REFERENCES

- [1]. Thiang and Suryo Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robots", in Proceedings of International Conference on Information and Electronics Engineering (IPCSIT).
- [2]. Ms.Vimala.C and Dr.V.Radha, "Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM", in Proceedings International Conference on Communication Technology and System Design 2020, Procedia Engineering 30 ISSN: 1877-7058, 13March 2020, pp.1097-1102.
- [3]. Cini Kuriana and Kannan Balakrishnan, "Development & evaluation of different acoustic models for Malayalam continuous speech recognition", in Proceedings of International Conference on Communication Technology and System Design 2020 Published by Elsevier Ltd, December 2020, pp.1081-1088.
- [4]. Suma Swamy and K.V Ramakrishnan, "An Efficient Speech Recognition System". Computer Science & Engineering: An International Journal (CSEIJ), Vol.3, No.4, and DOI: 10.512 1/cseij.2019.3403 August 2021, pp.21-27.
- [5]. Annu Choudhary, Mr. R.S. Chauhan and Mr. Gautam Gupta et.al. "Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language By Using Hidden Markov Model Toolkit (HTK)", in Proceedings of International Conference on Emerging Trends in Engineering and Technology, 03.AETS.2013.3.234, 22-24th February 2020, pp.244– 252.
- [6]. P. Sharma, V. Abrol, A. Sachdev and A. D. Dileep, et.al. "Speech emotion recognition using kernel sparse representation based classifier," in 2021 24th European Signal Processing Conference (EUSIPCO), pp. 374-377, 2021.
- [7]. Linhui Sun, Yiqing Huang, Qiu Li and Pingan Li, et.al. Multi-classification speech emotion recognition based on two-stage bottleneck features selection and MCJD algorithm, Signal Image and Video Processing, 10.1007/s11760-021-02076-0, 2022.

- [8]. Yu Wang, Research on the Construction of Human-Computer Interaction System Based on a Machine Learning Algorithm, Journal of Sensors, 10.1155/2022/3817226 2022, Vol 2022, pp. 1-11.
- [9]. Sandeep Kumar Pandey, Hanumant Singh Shekhawat and S.R.M Prasanna, et.al. Attention gated tensor neural network architectures for speech emotion recognition, Biomedical Signal Processing and Control □ 10.1016/j.bspc.2021.103173, 2022, Vol 71pp. 103173.
- [10]. Jason C. Hung and Jin-Che Chen, Construction and Research of E-sports Speech Emotion Recognition Model, Lecture Notes in Electrical Engineering - Innovative Computing, 10.1007/978-981-16-4258-6\_4, 2022, pp. 23-31.