

A Comprehensive Review of Machine Learning Approaches for Speech Emotion Recognition

Trupti Dilip Kalokhe and Prof. Rashmi Kulkarni

Department of Computer Engineering
Siddhant College of Engineering, Pune, India
trupti.bhase18@gmail.com

Abstract: *Speech Emotion Recognition (SER) has become integral to enhancing human-computer interaction, leveraging advanced signal processing and machine learning techniques to analyze emotions in vocal expressions. This review highlights key methods such as Mel Frequency Cepstral Coefficients (MFCCs), Linear Predictive Cepstral Coefficients (LPCCs), and Perceptual Linear Predictive Coefficients (PLPCs) for feature extraction, alongside classification models like Support Vector Machines (SVM), Gaussian Mixture Models (GMM), and deep learning approaches such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Recent advancements include hybrid models like Deep Belief Networks (DBN)-SVM and multimodal approaches combining speech, text, and facial features to improve accuracy. Applications of SER range from adaptive learning systems and mental health monitoring to real-time safety solutions. Despite progress, challenges such as noise resilience, limited dataset diversity, and emotion overlaps persist. Addressing these with strategies like transfer learning, autoencoders, and ensemble methods, the field continues to evolve toward creating scalable and reliable SER systems. Future research focuses on multimodal integration and refined architectures to enhance generalization and applicability in diverse scenarios.*

Keywords: Speech Emotion Recognition (SER), human-computer interaction, signal processing, machine learning, Mel Frequency Cepstral Coefficients (MFCCs), Linear Predictive Cepstral Coefficients (LPCCs), Perceptual Linear Predictive Coefficients (PLPCs), Support Vector Machines (SVM)