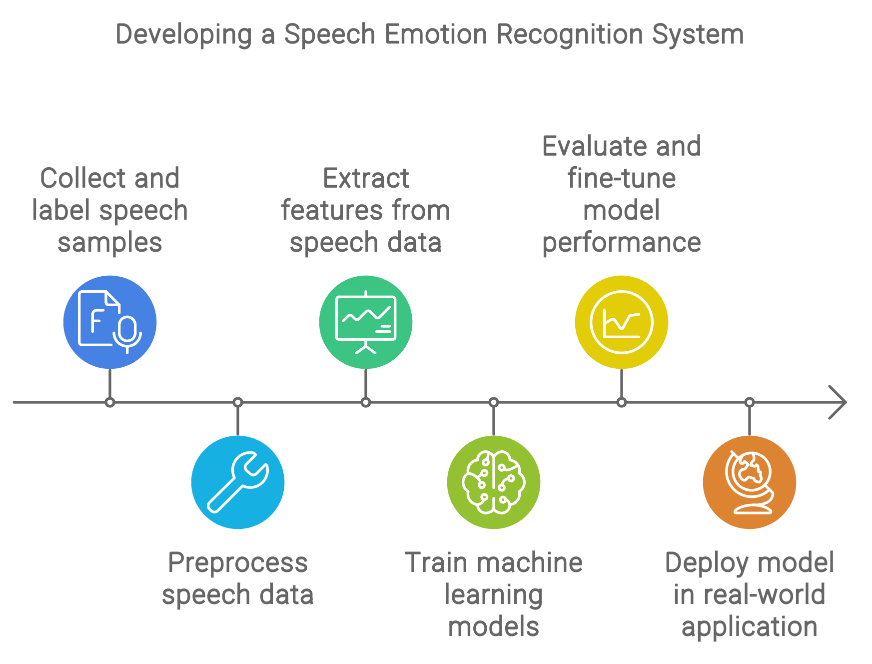
Speech Emotion Recognition Project

Abstract

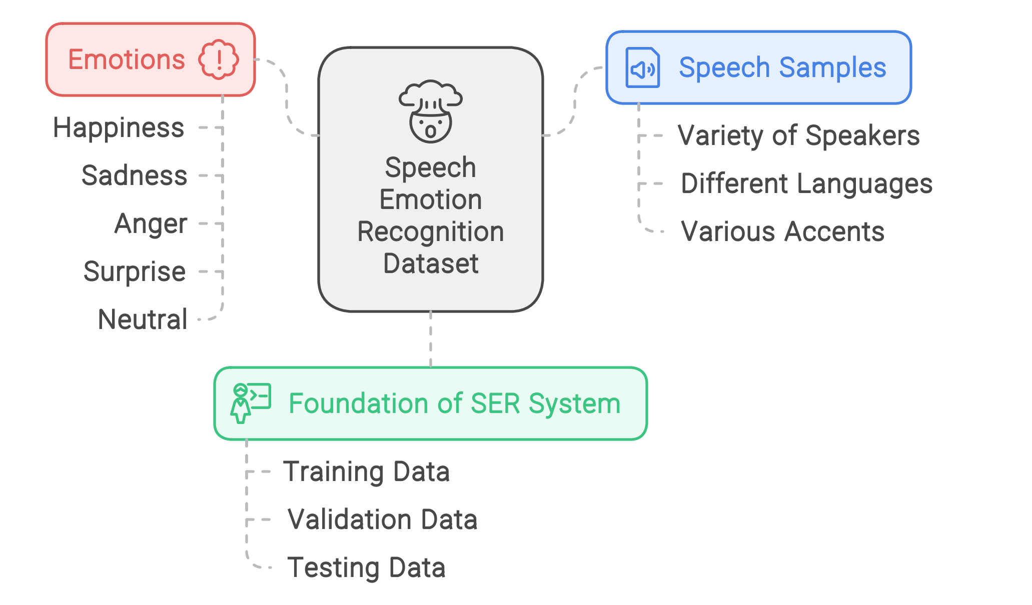
The Speech Emotion Recognition (SER) project aims to develop a system capable of automatically detecting emotions from speech signals. This involves collecting a dataset of speech samples labeled with corresponding emotions, applying preprocessing techniques to clean and standardize the audio data, and using feature extraction methods to capture relevant information. Machine learning models are then trained on these features to classify emotions. The model's performance is evaluated and fine-tuned before being deployed in a real-world application for real-time emotion recognition.



Project Overview

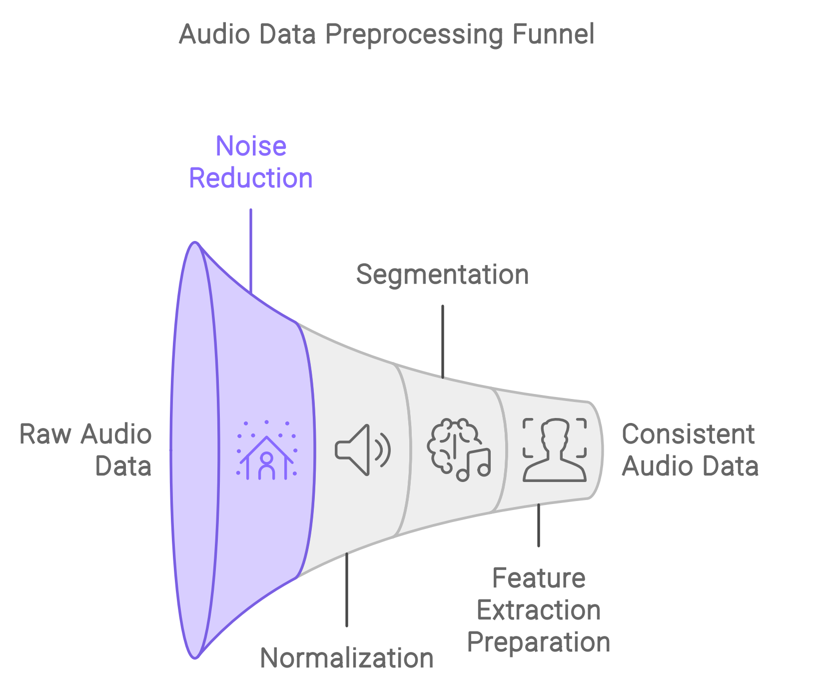
Data Collection

We start by collecting a dataset of speech samples that are labeled with corresponding emotions. This dataset forms the foundation of our SER system.



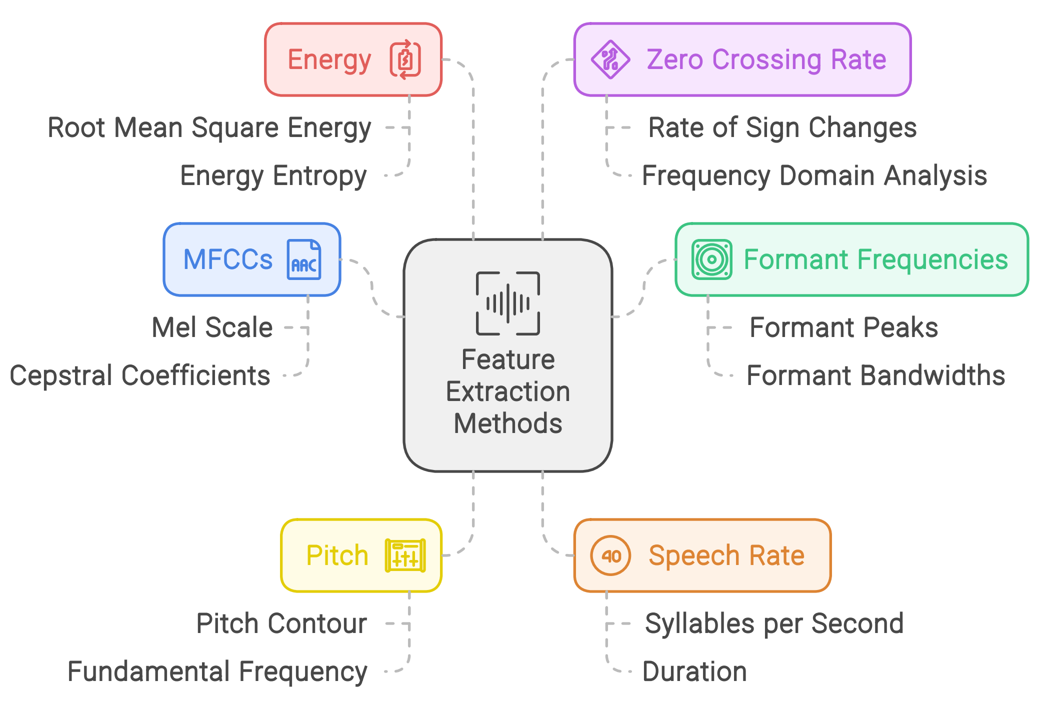
Preprocessing

Preprocessing techniques are applied to clean and standardize the audio data. This step is crucial to ensure that the data is in a consistent format for feature extraction and model training.



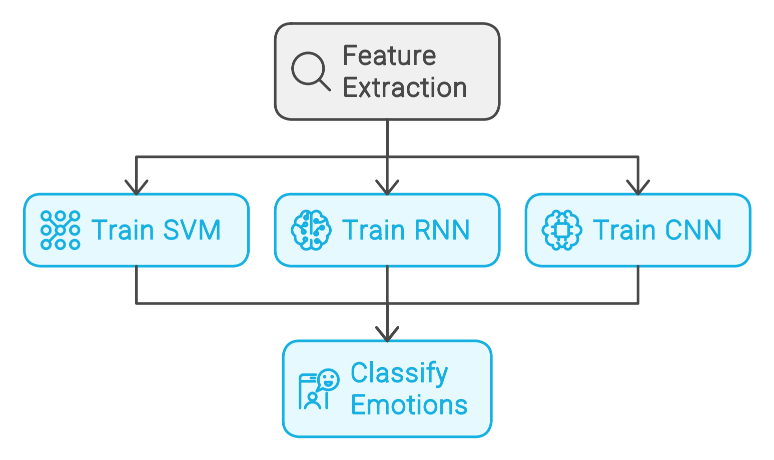
Feature Extraction

Feature extraction methods such as Mel Frequency Cepstral Coefficients (MFCCs) are used to capture relevant information from the speech signals. Other features include formant frequencies, pitch, speech rate, energy, and zero crossing rate.



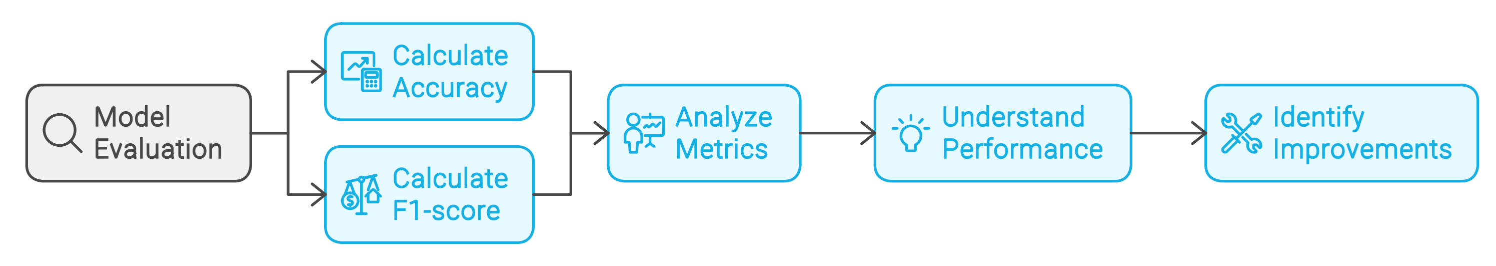
Model Training

Machine learning models, such as Support Vector Machines (SVM) or deep learning architectures like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs), are trained on the extracted features to classify emotions.



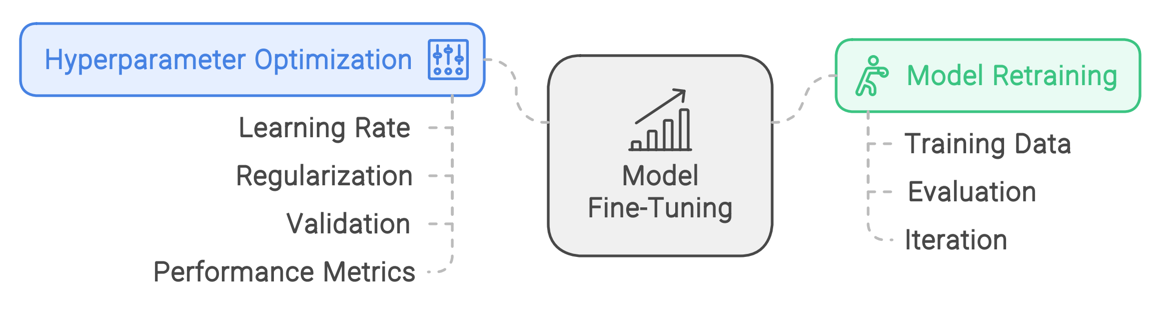
Model Evaluation

The model is evaluated using metrics like accuracy and F1-score to assess its performance. This step helps in understanding how well the model is performing and where improvements are needed.



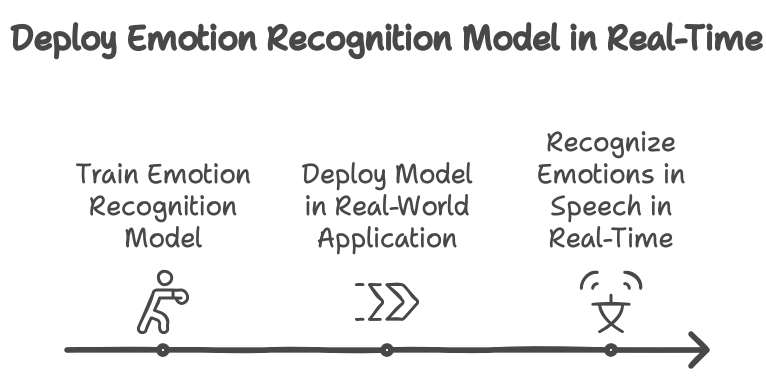
Model Fine-Tuning

We fine-tune the model and optimize hyperparameters to improve its accuracy. This involves adjusting various parameters and retraining the model to achieve better performance.



Deployment

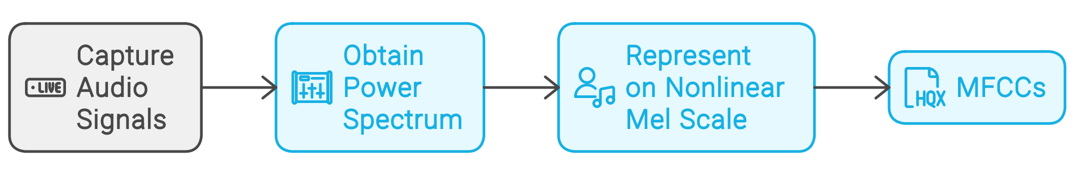
Finally, we deploy the trained model into a real-world application where it can accurately recognize emotions in speech in real-time.



Features

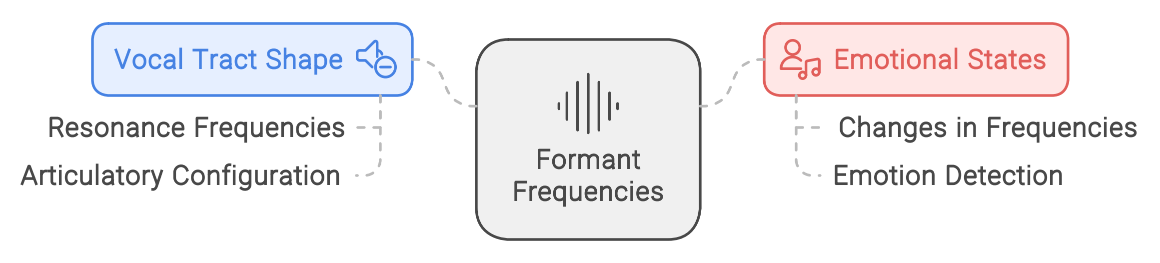
Mel Frequency Cepstral Coefficients (MFCCs)

- Description: MFCCs are widely used in speech processing to capture spectral characteristics.
- Details: They represent the power spectrum of the audio signal on a nonlinear mel scale.



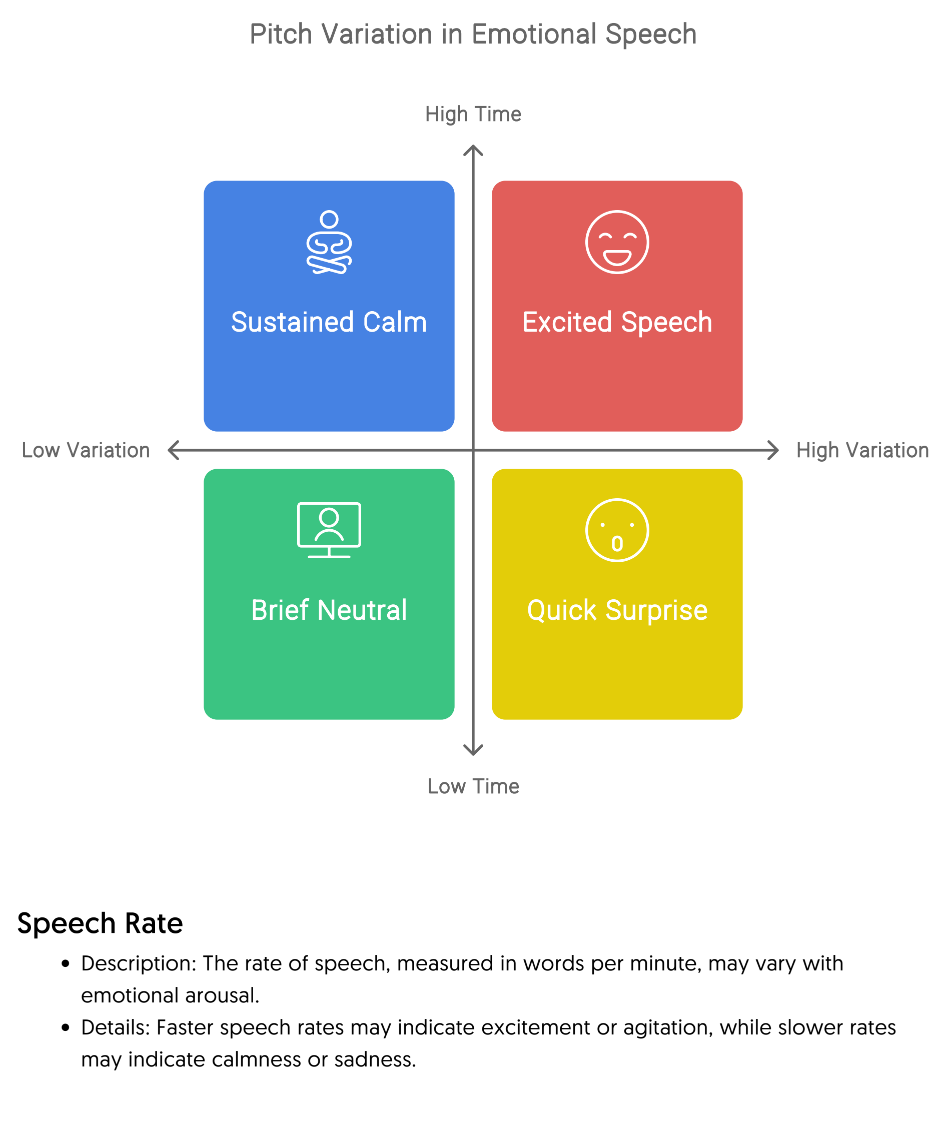
Formant Frequencies

- Description: Formants are resonance frequencies in the speech signal that are related to the shape of the vocal tract.
- Details: Changes in formant frequencies may reflect different emotional states.



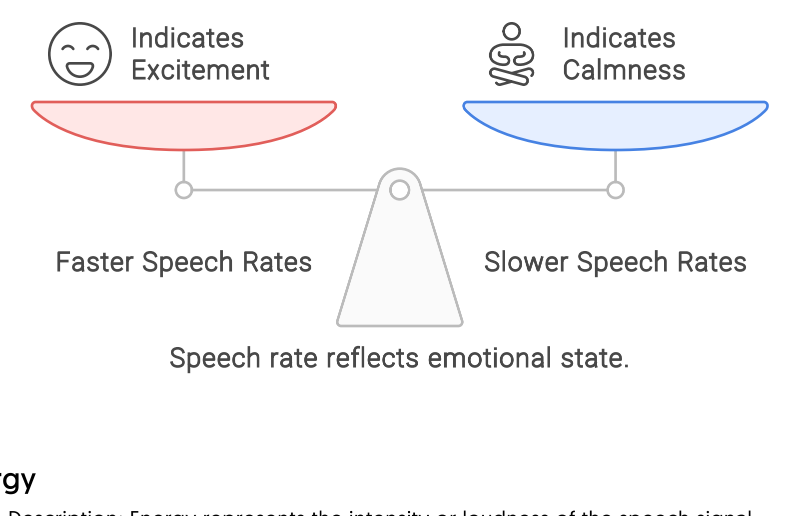
Pitch and Pitch Contour

- Description: Pitch represents the fundamental frequency of the speech signal, which can vary with different emotions.
- Details: Pitch contour analysis can capture variations in pitch over time, which may correlate with certain emotions.



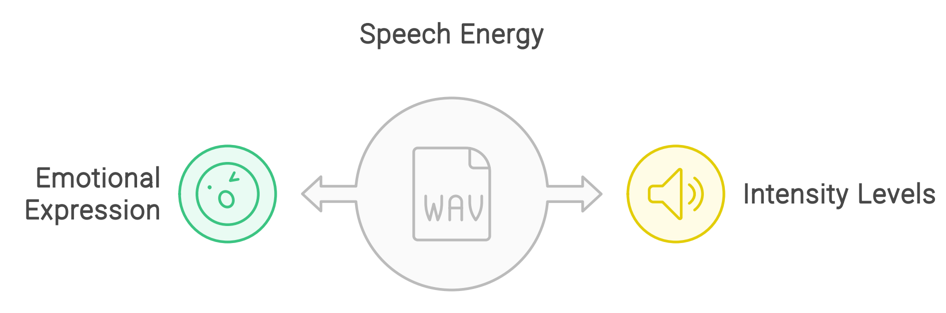
Speech Rate

- Description: The rate of speech, measured in words per minute, may vary with emotional arousal.
- Details: Faster speech rates may indicate excitement or agitation, while slower rates may indicate calmness or sadness.



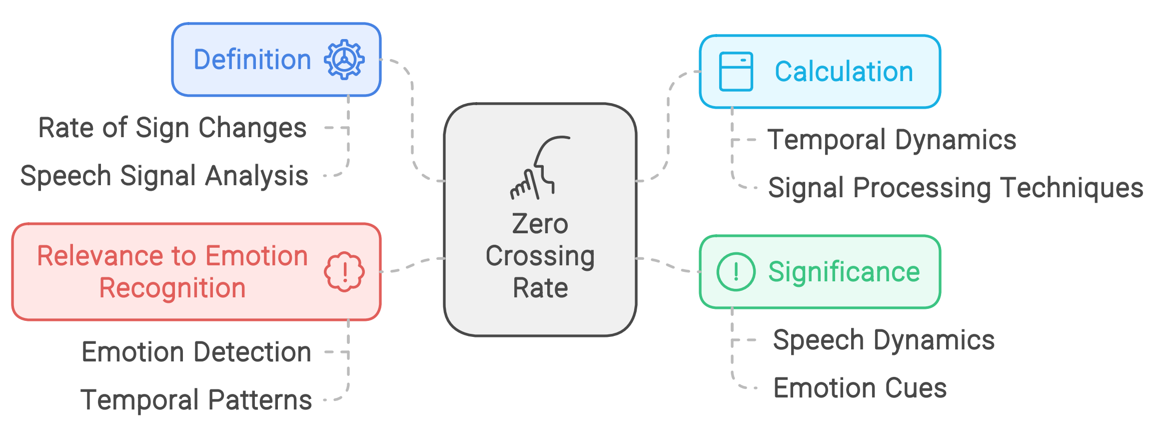
Energy

- Description: Energy represents the intensity or loudness of the speech signal.
- Details: Emotions like anger or excitement may be associated with higher energy levels.



Zero Crossing Rate

- Description: Zero crossing rate measures the rate at which the speech signal changes its sign.
- Details: It can provide information about the temporal dynamics of the speech signal, which may be relevant for emotion recognition.



By leveraging these features and advanced machine learning techniques, the Speech Emotion Recognition project aims to create a robust system capable of accurately identifying emotions from speech in real-time applications.