Project Report: Speaker Recognition Using Gaussian Mixture Models (GMMs)

Introduction

This project implements a speaker recognition system using Gaussian Mixture Models (GMMs) combined with Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction from audio samples. The goal of the system is to accurately identify a speaker from a set of known speakers using their voice samples. The system has been designed and tested using a subset of the VoxForge dataset, which contains voice samples from multiple speakers.

Project Overview

The project involves several key steps:

1. **Feature Extraction**: Audio samples are processed to extract MFCC features, which effectively capture the unique aspects of each speaker's voice.
2. **Model Training**: GMMs are trained for each speaker using their respective MFCC features from the training set.
3. **Speaker Recognition**: For a given audio sample, the system identifies the speaker by comparing the extracted features against the trained GMMs and selecting the model with the highest likelihood score.

Implementation Details

Feature Extraction

The feature extraction process converts raw audio data into MFCC features. This transformation involves:

- Loading audio files using the pydub library.
- Extracting audio samples and their respective sampling rates.
- Computing the MFCC features using the librosa library, with a specified hop duration and number of coefficients.

The extracted features for each speaker are stored in separate files to facilitate efficient model training.

Model Training

Each speaker's MFCC features are used to train a distinct GMM. The training process involves:

- Initializing a GaussianMixture model from sklearn with a predefined number of components and iterations.
- Fitting the model to the features of a specific speaker.
- Saving the trained models for later use in recognition tasks.

Speaker Recognition

The recognition process uses the trained GMMs to identify the speaker from a new audio sample by:

- Extracting MFCC features from the input audio.

- Calculating the likelihood scores for each trained GMM.
- Identifying the speaker associated with the GMM that gives the highest score.

Testing and Evaluation

The system's performance is evaluated on a test dataset that is not used during the training phase. Recognition accuracy is measured by comparing the predicted speaker identities against the actual identities.

Results

The speaker recognition system achieved a high accuracy rate, demonstrating the effectiveness of GMMs combined with MFCC features for the task of speaker identification. The system successfully differentiated between different speakers in the dataset, handling variations in voice and speech patterns effectively.

Challenges and Future Work

While the system performs well on the provided dataset, challenges such as handling noisy data, more diverse accents, and larger datasets need further investigation. Future work may include:

- Enhancing the robustness of feature extraction to improve system performance in noisy environments.
- Experimenting with more complex models or deep learning approaches to manage larger and more diverse datasets.
- Implementing real-time processing capabilities for live speaker recognition.

Conclusion

This project demonstrates the practical application of GMMs and MFCC for building an effective speaker recognition system. The successful implementation and testing indicate that this approach is viable for applications requiring reliable speaker identification based on audio samples.