EEC 201 Final Project

Zekai Yang Siheng Wu

Abstract

This system leverages digital signal processing techniques to extract Mel-Frequency Cepstral Coefficients (MFCCs) from speech files, applies vector quantization (VQ) using the Linde-Buzo-Gray (LBG) algorithm to build speaker models, and utilizes various pre- and post-processing steps (including normalization, lowpass filtering, and notch filtering) to enhance recognition performance. The project includes data preprocessing, feature extraction, clustering, and recognition tests with added noise and various speech distortions. The performance is compared with human recognition accuracy, and the impact of different utterances on the recognition rate is discussed.

1. Introduction

Speaker recognition is the task of automatically identifying or verifying a speaker's identity using unique characteristics in their speech. In this project, the speaker recognition system:

- Reads speech files,
- Preprocesses the signal (normalization, lowpass filtering),
- Extracts MFCC features,
- Trains speaker-specific models via vector quantization (using the LBG algorithm),
- Matches test utterances against trained codebooks.

Apart from that, experiments were performed to analyze the effect of different processing parameters (e.g., FFT window sizes, notch filtering) and to classify different speech phrases such as "Zero" versus "Twelve" and "Five" versus "Eleven."

2. Methodology

2.1 Data Access and Preprocessing

Using Google Colab, the speech database is mounted from Google Drive. The code reads audio files (in WAV format) for training and testing:

- **Loading and Normalization:**
  The audio files are read at a sampling rate of 16 kHz. The signal is then normalized by subtracting its mean and dividing by its standard deviation.

- **Filtering:**
  A lowpass Butterworth filter (with a default cutoff at 3000 Hz) is applied to remove high-frequency noise. In later experiments, notch filters are used to test robustness against frequency-specific distortions.

2.2 Feature Extraction

The key feature extraction process involves:

- **Short-Time Fourier Transform (STFT):**
  The STFT is computed to analyzce the spectral content of the speech signal, and the periodogram is used to visualize energy distribution.

- **MFCC Computation:**
  The function compute_mfcc computes 26 MFCC coefficients using a specified FFT length (typically 1024) and hop length (256 samples). The system also provides comparisons between different window sizes (e.g., 128, 256, 512) to observe the effect on MFCC features.

- **Mel Filterbank Visualization:**
  A mel-spaced filterbank is generated and plotted to verify the triangular filter responses that capture perceptually relevant frequency information.

2.3 Model Training Using Vector Quantization

For each speaker, the extracted MFCC features are used to train a VQ codebook:

- **LBG Algorithm:**
  Starting with an initial single-vector codebook (the centroid of the training data), the codebook is iteratively split using a small perturbation (epsilon = 0.01) until a predefined number of clusters (e.g., 16) is reached. The code uses the scipy.cluster.vq.kmeans method to refine the codebook at each iteration.

2.4 Testing and Speaker Matching

The system matches test MFCC features against the trained codebooks using Euclidean distance:

- For each test utterance, the average minimum distance (using the scipy.spatial.distance.cdist function) between the MFCC vectors and each speaker's codebook is computed.
- The speaker (or phrase) associated with the minimum average distance is assigned as the predicted label.

3. Experiments and Results

3.1 Initial Training and Recognition

- **Training Data:**
  Eleven training files (s1.wav to s11.wav) were processed to extract MFCC features. The MFCC features were stored per speaker, and a corresponding VQ codebook was built.

- **Test Matching:**
  Test files were processed similarly, and the matching function predicted the speaker by finding the codebook with the lowest average distortion. The matching results were printed with predicted speaker IDs.

3.2 Analysis of Signal Processing Steps

- **Waveform and Spectrogram Plots:**
  The filtered waveforms and STFT spectrograms were plotted for each file. These visualizations confirmed that lowpass filtering and normalization effectively reduced unwanted noise and aligned the signals in time.

- **Window Size Impact:**
  The compare_mfcc function produced MFCC plots using different FFT sizes (128, 256, 512). The experiments demonstrated that while the overall MFCC patterns were similar, finer details varied with window size, affecting the robustness of feature representation.

3.3 Robustness Under Notch Filtering

To simulate real-world distortions:

- Notch filters at various frequencies (50, 100, 200, 500, 1000 Hz) were applied to the test signals.
- The matching function was re-run on the notch-filtered signals.
- The results shows that the system is effected by specific frequency removals.

3.4 Extended Experiments: Phrase and Speaker Classification

3.4.1 Zero vs. Twelve

- **Data Collection:**
  Training and testing files were collected from student recordings for the phrases “Zero” and “Twelve.”

- **Dual Classification:**
  Separate codebooks were built: one for speaker identification and another for phrase classification.
- **Results:**
  The system printed the predicted speaker IDs and phrases. Accuracy metrics were computed for both tasks, with speaker and phrase accuracies reported separately.

  Matching Results:

  TEST 0: ID: 10 | Predicted ID: 10 | True phrase: Zero | Predicted phrase: Zero

  TEST 1: ID: 4 | Predicted ID: 4 | True phrase: Zero | Predicted phrase: Zero

  TEST 2: ID: 13 | Predicted ID: 11 | True phrase: Zero | Predicted phrase: Zero

  TEST 3: ID: 1 | Predicted ID: 1 | True phrase: Zero | Predicted phrase: Zero

  TEST 4: ID: 7 | Predicted ID: 2 | True phrase: Zero | Predicted phrase: Zero

  TEST 5: ID: 6 | Predicted ID: 6 | True phrase: Zero | Predicted phrase: Zero

  TEST 6: ID: 3 | Predicted ID: 3 | True phrase: Zero | Predicted phrase: Twelve

  TEST 7: ID: 9 | Predicted ID: 9 | True phrase: Zero | Predicted phrase: Zero

  TEST 8: ID: 11 | Predicted ID: 11 | True phrase: Zero | Predicted phrase: Zero

  TEST 9: ID: 12 | Predicted ID: 12 | True phrase: Zero | Predicted phrase: Zero

  TEST 10: ID: 8 | Predicted ID: 8 | True phrase: Zero | Predicted phrase: Zero

  TEST 11: ID: 2 | Predicted ID: 2 | True phrase: Zero | Predicted phrase: Zero

  TEST 12: ID: 14 | Predicted ID: 14 | True phrase: Zero | Predicted phrase: Zero

  TEST 13: ID: 17 | Predicted ID: 17 | True phrase: Zero | Predicted phrase: Zero

  TEST 14: ID: 18 | Predicted ID: 18 | True phrase: Zero | Predicted phrase: Zero

  TEST 15: ID: 15 | Predicted ID: 13 | True phrase: Zero | Predicted phrase: Zero

  TEST 16: ID: 19 | Predicted ID: 19 | True phrase: Zero | Predicted phrase: Zero

  TEST 17: ID: 16 | Predicted ID: 16 | True phrase: Zero | Predicted phrase: Zero

  TEST 18: ID: 3 | Predicted ID: 3 | True phrase: Twelve | Predicted phrase: Twelve

  TEST 19: ID: 16 | Predicted ID: 16 | True phrase: Twelve | Predicted phrase: Twelve

  TEST 20: ID: 4 | Predicted ID: 4 | True phrase: Twelve | Predicted phrase: Twelve

TEST 21: ID: 2 | Predicted ID: 2 | True phrase: Twelve | Predicted phrase: Twelve

TEST 22: ID: 19 | Predicted ID: 19 | True phrase: Twelve | Predicted phrase: Twelve

TEST 23: ID: 8 | Predicted ID: 8 | True phrase: Twelve | Predicted phrase: Twelve

TEST 24: ID: 11 | Predicted ID: 11 | True phrase: Twelve | Predicted phrase: Twelve

TEST 25: ID: 7 | Predicted ID: 7 | True phrase: Twelve | Predicted phrase: Twelve

TEST 26: ID: 18 | Predicted ID: 18 | True phrase: Twelve | Predicted phrase: Twelve

TEST 27: ID: 10 | Predicted ID: 10 | True phrase: Twelve | Predicted phrase: Twelve

TEST 28: ID: 17 | Predicted ID: 17 | True phrase: Twelve | Predicted phrase: Twelve

TEST 29: ID: 12 | Predicted ID: 12 | True phrase: Twelve | Predicted phrase: Twelve

TEST 30: ID: 15 | Predicted ID: 1 | True phrase: Twelve | Predicted phrase: Twelve

TEST 31: ID: 14 | Predicted ID: 14 | True phrase: Twelve | Predicted phrase: Twelve

TEST 32: ID: 13 | Predicted ID: 13 | True phrase: Twelve | Predicted phrase: Twelve

TEST 33: ID: 9 | Predicted ID: 9 | True phrase: Twelve | Predicted phrase: Twelve

TEST 34: ID: 6 | Predicted ID: 13 | True phrase: Twelve | Predicted phrase: Twelve

TEST 35: ID: 1 | Predicted ID: 1 | True phrase: Twelve | Predicted phrase: Twelve

Speaker accuracy: 0.86

Zero accuracy: 0.94

Twelve accuracy: 1.00

3.4.2 Five vs. Eleven

- **Data Collection:**
  A similar procedure was followed for the phrases “Five” and “Eleven.”
- **Results:**
  Matching results and accuracy rates were computed.

Matching Results:
TEST 0: ID: 1 | Predicted ID: 1 | True phrase: Five | Predicted phrase: Eleven
TEST 1: ID: 2 | Predicted ID: 2 | True phrase: Five | Predicted phrase: Five

TEST 2: ID: 3 | Predicted ID: 3 | True phrase: Five | Predicted phrase: Five
TEST 3: ID: 4 | Predicted ID: 4 | True phrase: Five | Predicted phrase: Five
TEST 4: ID: 5 | Predicted ID: 5 | True phrase: Five | Predicted phrase: Five
TEST 5: ID: 6 | Predicted ID: 6 | True phrase: Five | Predicted phrase: Five
TEST 6: ID: 7 | Predicted ID: 7 | True phrase: Five | Predicted phrase: Five
TEST 7: ID: 8 | Predicted ID: 8 | True phrase: Five | Predicted phrase: Five
TEST 8: ID: 9 | Predicted ID: 9 | True phrase: Five | Predicted phrase: Five
TEST 9: ID: 10 | Predicted ID: 10 | True phrase: Five | Predicted phrase: Eleven
TEST 10: ID: 11 | Predicted ID: 11 | True phrase: Five | Predicted phrase: Five
TEST 11: ID: 12 | Predicted ID: 12 | True phrase: Five | Predicted phrase: Five
TEST 12: ID: 13 | Predicted ID: 13 | True phrase: Five | Predicted phrase: Five
TEST 13: ID: 14 | Predicted ID: 14 | True phrase: Five | Predicted phrase: Eleven
TEST 14: ID: 15 | Predicted ID: 15 | True phrase: Five | Predicted phrase: Five
TEST 15: ID: 16 | Predicted ID: 16 | True phrase: Five | Predicted phrase: Five
TEST 16: ID: 17 | Predicted ID: 17 | True phrase: Five | Predicted phrase: Five
TEST 17: ID: 18 | Predicted ID: 18 | True phrase: Five | Predicted phrase: Five
TEST 18: ID: 19 | Predicted ID: 19 | True phrase: Five | Predicted phrase: Five
TEST 19: ID: 20 | Predicted ID: 20 | True phrase: Five | Predicted phrase: Five
TEST 20: ID: 21 | Predicted ID: 21 | True phrase: Five | Predicted phrase: Five
TEST 21: ID: 22 | Predicted ID: 22 | True phrase: Five | Predicted phrase: Five
TEST 22: ID: 23 | Predicted ID: 23 | True phrase: Five | Predicted phrase: Five
TEST 23: ID: 1 | Predicted ID: 1 | True phrase: Eleven | Predicted phrase: Eleven
TEST 24: ID: 2 | Predicted ID: 2 | True phrase: Eleven | Predicted phrase: Eleven
TEST 25: ID: 3 | Predicted ID: 3 | True phrase: Eleven | Predicted phrase: Eleven
TEST 26: ID: 4 | Predicted ID: 4 | True phrase: Eleven | Predicted phrase: Eleven
TEST 27: ID: 5 | Predicted ID: 5 | True phrase: Eleven | Predicted phrase: Eleven
TEST 28: ID: 6 | Predicted ID: 6 | True phrase: Eleven | Predicted phrase: Eleven
TEST 29: ID: 7 | Predicted ID: 7 | True phrase: Eleven | Predicted phrase: Eleven
TEST 30: ID: 8 | Predicted ID: 8 | True phrase: Eleven | Predicted phrase: Eleven
TEST 31: ID: 9 | Predicted ID: 9 | True phrase: Eleven | Predicted phrase: Eleven
TEST 32: ID: 10 | Predicted ID: 10 | True phrase: Eleven | Predicted phrase: Eleven
TEST 33: ID: 11 | Predicted ID: 11 | True phrase: Eleven | Predicted phrase: Eleven
TEST 34: ID: 12 | Predicted ID: 12 | True phrase: Eleven | Predicted phrase: Eleven
TEST 35: ID: 13 | Predicted ID: 13 | True phrase: Eleven | Predicted phrase: Eleven
TEST 36: ID: 14 | Predicted ID: 14 | True phrase: Eleven | Predicted phrase: Eleven
TEST 37: ID: 15 | Predicted ID: 15 | True phrase: Eleven | Predicted phrase: Eleven
TEST 38: ID: 16 | Predicted ID: 16 | True phrase: Eleven | Predicted phrase: Eleven
TEST 39: ID: 17 | Predicted ID: 17 | True phrase: Eleven | Predicted phrase: Eleven
TEST 40: ID: 18 | Predicted ID: 18 | True phrase: Eleven | Predicted phrase: Eleven
TEST 41: ID: 19 | Predicted ID: 19 | True phrase: Eleven | Predicted phrase: Eleven
TEST 42: ID: 20 | Predicted ID: 20 | True phrase: Eleven | Predicted phrase: Eleven
TEST 43: ID: 21 | Predicted ID: 21 | True phrase: Eleven | Predicted phrase: Eleven
TEST 44: ID: 22 | Predicted ID: 22 | True phrase: Eleven | Predicted phrase: Eleven
TEST 45: ID: 23 | Predicted ID: 23 | True phrase: Eleven | Predicted phrase: Five
Speaker accuracy: 1.00
Five accuracy: 0.87
Eleven accuracy: 0.96

The experiments allowed a direct comparison of phrase-specific recognition performance across different sets.

4. Discussion

The experiments demonstrated that:

- **Preprocessing and Feature Extraction:**
  Normalization and filtering are critical for minimizing variability in the raw signals. The MFCC extraction process successfully captured the unique spectral characteristics of each speaker.

- **Model Training and Matching:**
  The LBG algorithm provided a straightforward yet effective way to build speaker-specific codebooks. The matching results—based on Euclidean distance—correlated well with the expected speaker identities under clean conditions.

- **Robustness and Sensitivity:**
  While the system achieved high accuracy under normal conditions, the performance degraded with frequency-specific distortions (notch filtering). The extended experiments on phrase classification also revealed the challenges in simultaneously modeling speaker identity and spoken content.

5. Conclusion

The project integrated standard DSP techniques (normalization, lowpass filtering, STFT) with MFCC feature extraction and vector quantization for speaker and phrase identification. Experimental results indicate that while the system performs reliably under controlled conditions, its robustness under distortion and mixed phrase conditions can be further improved.