Speech Understanding Programming Assignment – 1

Accent Detection & Spectrogram Analysis

Rishabh Sharma / M23CSA523

February 2, 2025

Outline

Introduction

- **Objective:** Address two key problems in speech processing:
  - Accent Detection (Group Assignment)
  - Spectrogram Analysis with Windowing Techniques (Individual Assignment)
- **Applications:** Speech recognition, personalization, forensic analysis, and language learning.

Accent Detection: Importance & Applications

- **Why Accent Detection?**
  - Enhances speech recognition accuracy.
  - Tailors user interactions.
  - Assists in speaker profiling and forensic applications.
  - Provides feedback for language learning.

Current Approaches in Accent Detection

Traditional Methods:

- Feature extraction (MFCCs, prosodic features)
- Classical classifiers (SVMs, GMMs)

Deep Learning Methods:

- CNNs for spectrogram analysis
- RNNs/LSTMs for temporal modeling
- Transformer and hybrid architectures

Evaluation Metrics & Challenges

- **Metrics:** Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC-AUC.
- **Challenges:**
  - Limited and imbalanced datasets.
  - High variability within accent classes.
  - Noise robustness.
  - Model interpretability.
  - Real-time deployment.

CNN-based Accent Detection Module

- **Data Processing:** Audio converted to Mel-spectrograms.
- **Model:** Convolutional Neural Network implemented in PyTorch.
- **Training:** Standard training loop using cross-entropy loss and the Adam optimizer.
- **Results:** (Insert performance metrics and confusion matrix images here.)

Spectrogram Analysis: Overview

- **Dataset:** UrbanSound8K.
- **Techniques:** Explore various window functions in the STFT:
  - Hann Window
  - Hamming Window
  - Rectangular Window
- **Objective:** Visual comparison and analysis of spectral features.

Windowing Functions and Their Effects

Hann & Hamming:

- Smoother spectral content.
- Reduced spectral leakage.

Rectangular:

- Less smoothing.
- More spectral leakage.

UrbanSound8K Classifier Experiment

- **Approach:** Train a simple CNN classifier on spectrogram features.
- **Comparison:** Evaluate which windowing function yields better classification performance.
- **Results:** (Insert classification accuracy and loss graphs here.)

Comparative Analysis of Music Genres

- **Genres Analyzed:** Rock, Classical, Pop, Jazz.
- **Observations:**
  - Differences in harmonic structure.
  - Variations in transient features.
  - Distinct energy distributions.
- **Visuals:** Spectrogram images for each genre.

Conclusion and Future Directions

- **Summary:**
  - A comprehensive analysis of accent detection and spectrogram analysis.
  - Successful implementation of a CNN for accent detection.
  - Detailed comparison of windowing techniques on the UrbanSound8K dataset.
- **Future Work:**
  - Enhance model robustness and accuracy.
  - Explore advanced architectures.
  - Optimize models for real-time processing.

Q & A

Questions?

References

- Li, J., et al. (2019). *Accent Detection and Its Applications in Speech Recognition.* IEEE/ACM Transactions on Audio, Speech, and Language Processing.
- Zhang, Y., et al. (2020). *Deep Learning Approaches for Accent Classification.* In Proceedings of Interspeech 2020.
- Chung, Y. A., et al. (2021). *Transformers for Speech Recognition.* In Proceedings of ICASSP 2021.
- Salamon, J., et al. (2014). *A Dataset and Taxonomy for Urban Sound Research.* In Proceedings of ACM Multimedia.