DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF E&ME, NUST, RAWALPINDI

EE232-Signals & Systems

Project Report

Project: "Audio Classification"

SUBMITTED TO:

- **Lec Furqan Haider**
- **LE Aleena**

SUBMITTED BY:

1. **Shaheer Mukhtiar**
   - Reg No: 432017
2. **Wahaaj Nasir**
   - Reg No: 413238
3. **Mohammed Ali Riaz**
   - Reg No: 432103
4. **Syed Adnan Aijaz Bukhari**
   - Reg No: 432028

DE- 44-CE-B

Theoretical Background

1. Introduction

The goal of this project is to classify urban sound signals using the UrbanSound8K dataset. This involves leveraging audio signal processing techniques such as pitch detection, feature extraction, and machine learning algorithms to achieve accurate classification. The output is a robust classification system capable of distinguishing between speech, music, and noise categories. The dataset provides a wide variety of urban sound classes, ensuring a comprehensive evaluation of the system's performance. This project serves as an essential step in advancing the application of machine learning in audio signal analysis, with potential implications in areas such as noise pollution monitoring and urban planning.

2. Signal Processing Techniques

i. **Pitch Detection:**

- Pitch is a key feature in audio analysis, reflecting the fundamental frequency of a signal. It plays a vital role in distinguishing between audio classes like speech and music.
- Techniques used:
    - **Autocorrelation**: Detects periodicity in signals for estimating pitch by analysing repetitive patterns within the signal. This method is particularly useful for steady-state signals like musical tones.
    - **Harmonic Product Spectrum (HPS)**: Combines harmonics to enhance pitch detection accuracy by reducing the impact of noise and harmonics. It is effective for complex signals where multiple frequencies coexist.
    - **Librosa's piptrack Method**: Uses a spectrogram-based approach to track pitch. This technique is well-suited for signals with varying pitches, such as speech.

ii. Feature Extraction:

- **MFCCs (Mel-Frequency Cepstral Coefficients)**: Capture timbral aspects of audio by simulating human auditory perception, making them a critical component for audio classification tasks.
- **Zero Crossing Rate (ZCR)**: Indicates signal complexity by measuring how often the signal changes sign. This feature helps in distinguishing between tonal and noisy sounds.
- **RMS Energy**: Represents signal loudness and is useful for identifying high-energy events such as gunshots or jackhammer noise.
- **Pitch**: Frequency of the fundamental tone, crucial for distinguishing tonal sounds like music from noise.

3. Machine Learning Algorithms

i. Random Forest Classifier:

- Ensemble learning technique that builds multiple decision trees and aggregates their results for robust predictions. It is particularly effective for handling high-dimensional datasets like UrbanSound8K.

- Used for both sound classification and pitch detection comparisons. The algorithm’s ability to rank features by importance provides insights into the most significant audio characteristics for classification.

ii. **Evaluation Metrics:**

- **Accuracy:** Measures the percentage of correctly classified instances, providing an overall performance indicator.
- **Precision:** Reflects the proportion of true positive predictions among all positive predictions, ensuring the reliability of classifications.
- **Recall:** Indicates the proportion of actual positives correctly identified, highlighting the sensitivity of the model.
- **F1-Score:** Combines precision and recall into a single metric, particularly useful for imbalanced datasets.

4. UrbanSound8K Dataset

- **Description:**
  - The dataset contains 8,732 labelled audio recordings of urban sound classes such as car horns, jackhammers, and dog barks. Each recording is up to four seconds long and categorized into one of ten classes.
  - Sounds are stored in corresponding folders (folds) to facilitate cross-validation. This organization ensures robust model training and testing.
- **Application:**
  - Used for training and testing the classification models, with a focus on diverse urban environments. The dataset’s rich variety enhances the model’s generalizability to real-world scenarios.
  - Data preprocessing included noise addition (Rayleigh, Nakagami) to test model robustness and feature extraction to prepare the data for machine learning.

5. Challenges and Solutions

- **Handling Noise:**
  - Added synthetic noise (Rayleigh, Nakagami) to enhance the model's robustness.
- **Empty Filter Issues:**
  - Addressed warning messages by adjusting sampling rates and frequencies for audio processing.

Abstract

The project aims to develop a robust system to classify audio signals into speech, music, and noise categories. Leveraging the UrbanSound8K dataset and advanced signal processing techniques, this study employs pitch detection methods, such as autocorrelation, harmonic product spectrum (HPS), and piptrack. A Random Forest Classifier evaluates the features extracted from the audio data. Key findings highlight the effectiveness of the HPS method under clean conditions, achieving the highest classification accuracy of 89.98%. This project provides a comparative analysis of pitch detection methods and explores the resilience of the model under various noise conditions, offering insights for real-world applications.

Objectives

1. **Audio Signal Classification:** Develop a system to classify audio signals into speech, music, and noise categories.
2. **Pitch Detection Comparison:** Evaluate the performance of pitch detection methods (autocorrelation, HPS, piptrack).
3. **Noise Impact Analysis:** Analyze the robustness of pitch detection methods and classification models under noisy conditions.

Methodology

1. Dataset Preparation

The UrbanSound8K dataset, consisting of labeled urban audio signals, was used. Preprocessing involved:

- Loading the audio files and handling sampling rates.
- Adding noise (Rayleigh and Nakagami-m) for robustness testing.
- Extracting features such as MFCCs, zero-crossing rate, RMS, energy, and pitch.

2. Algorithms

- **Pitch Detection Methods:**
    - *Autocorrelation*: Captures repetitive patterns in audio for fundamental frequency estimation.
    - *HPS*: Combines harmonics to enhance pitch estimation.
    - *Piptrack*: Extracts pitch contours using a spectrogram-based approach.
- **Random Forest Classifier:**
    - The extracted features were used to train and evaluate the model. Training involved an 80-20 train-test split, with accuracy and performance metrics computed for each pitch detection method.

3. **Evaluation Metrics**
    - **Accuracy:** Percentage of correctly classified samples.
    - **Precision:** Measure of true positive predictions.
    - **Recall:** Measure of correctly identified positives.
    - **F1-Score:** Harmonic mean of precision and recall.

Results & Outputs

1. **Pitch Detection Method Comparison**
- **Without Noise:**
    - Autocorrelation: 89.70%
    - HPS: 89.98%
    - Piptrack: 89.52%
- **With Noise:**
    - Rayleigh Noise: HPS achieved the highest robustness with 86.20% accuracy.
    - Nakagami Noise: HPS again performed the best with 85.98% accuracy.

2. **Classification Performance**
- Speech: 76% Precision, 64% Recall
- Music: 81% Precision, 50% Recall
- Noise: 84% Precision, 95% Recall
  Overall Classification Accuracy: 82.03%

Discussion

- **Best Performing Method:** HPS consistently performed the best under clean and noisy conditions due to its effective harmonic analysis.
- **Challenges:**
    - **Empty Filter Warnings:** Addressed by adjusting sampling rates and filter parameters.
    - **Processing Errors:** Occurred with zero-size arrays in piptrack, mitigated by data cleaning and validation steps.
- **Insights:** Noise addition showed a predictable decline in accuracy, highlighting areas for improving feature extraction under challenging conditions.

Conclusion

The project demonstrated the viability of classifying audio signals using signal processing and machine learning techniques. HPS emerged as the most effective pitch detection method. Future work includes exploring deep learning-based models and enhancing noise resilience.

References

1. **UrbanSound8K Dataset**: Kaggle - UrbanSound8K
2. **Librosa Library**: Python package for audio processing.
3. **Scikit-learn**: Toolkit for machine learning.
4. **Reference Texts**: Signal processing textbooks and online tutorials on pitch detection and digital filtering.