ELEC5305 Project Feedback 2

Teacher: Craig Jin

TA: Reza Ghanavi

1. Project Title

Nonspeech signal recognition

2. Student Information

- Full Name: Pengcheng Wang
- Student ID (SID): 530534486
- GitHub Username: David-W-Pc
- GitHub Project Link: https://github.com/David-W-Pc/elec5305-project-530534486.git

1. Project Overview

This project focuses on developing a system capable of classifying non-speech environmental sounds such as dog barks, alarms, footsteps, and sirens. Unlike speech recognition, environmental sound classification presents additional challenges due to diverse acoustic textures, irregular durations, and background noise. The project explores and compares both classical machine learning approaches (e.g., SVM with MFCC features) and deep learning models (e.g., CNN using spectrogram representations).

2. Background and Motivation

Environmental sound recognition has growing applications in smart home devices, urban monitoring, assistive technologies, and safety systems. Existing literature demonstrates that techniques commonly used in speech processing—such as MFCCs and Mel-Spectrograms—can be adapted to environmental audio when combined with appropriate classification models.

The motivation for this project is to investigate the trade-offs between computational efficiency and classification accuracy using different feature extraction and learning approaches.

3. Dataset and Literature Review

The project utilizes the ESC-50 dataset, a well-established benchmark for environmental sound classification:

50 balanced sound classes

2000 audio clips, each 5 seconds, sampled at 44.1 kHz

Categories include: animals, natural sounds, human activities, interior/exterior events, machinery

From the literature, two primary strategies emerge:

Classical pipelines using handcrafted features (MFCC, Spectral Centroid, Zero-Crossing Rate)

Deep learning pipelines using 2D time-frequency images (Mel-Spectrograms, Log-Mel, CQT) with CNNs

This comparative framework forms the experimental basis of the project.

4. Feature Extraction Methodology

Two feature representations have been selected:

MFCC (Mel-Frequency Cepstral Coefficients)

Used for classical machine learning models. MFCCs capture spectral envelope characteristics and are effective in compressing audio information into a low-dimensional feature vector.

Mel-Spectrogram

Used for deep learning models. Treated as an image representation, enabling the use of convolutional neural networks for temporal and spectral pattern recognition.

Initial implementation has been developed in MATLAB using mfcc and melSpectrogram functions. The pipeline also includes normalization and statistical pooling (mean/std) for MFCC-based models.

5. Current Progress

Project repository has been established and structured for MATLAB development

Dataset metadata reviewed and prepared for feature extraction

MFCC-based feature extraction implemented successfully

Baseline model (e.g., SVM/ECOC) is being developed for initial benchmarking

Mel-Spectrogram generation pipeline designed for future CNN implementation

6. Next Steps

Train and evaluate baseline models using MFCC + SVM

Implement Mel-Spectrogram CNN pipeline (e.g., shallow CNN, CRNN, or VGG-style)

Perform comparative analysis using accuracy and confusion matrices

Document experimental results and prepare visualizations for presentation/report