ELEC5305 Project Proposal

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

3. Project Overview (provisional)

This project focuses on the recognition of **non-speech acoustic events**, such as alarms, dog barks, door knocks, sirens, and other environmental sounds. While speech recognition has been studied extensively, the detection and classification of non-speech audio is equally important for enabling **intelligent systems that interact with real-world environments**. For example, a smart home assistant could respond not only to spoken commands but also detect an alarm clock, a smoke detector, or a baby crying, thereby improving both safety and user convenience. Similarly, urban monitoring systems that can automatically detect sounds like car horns or sirens could support **transportation safety and emergency response**.

The goal of this project is to design and implement a system capable of automatically categorizing short audio clips into predefined sound classes. The approach combines **signal processing techniques** (e.g., MFCCs, spectral features, and spectrogram representations) with **machine learning models** (both traditional classifiers such as SVMs and modern deep learning methods such as CNNs). Standard open datasets like **ESC-50** and **UrbanSound8k** will be used for training and evaluation, ensuring reproducibility.

Evaluation metrics such as **accuracy, precision, recall, F1-score, and confusion matrices** will provide a quantitative measure of performance, while qualitative analysis will help identify typical misclassifications and areas for improvement. Ultimately, this project aims to contribute a **working prototype system** with documented performance benchmarks and insights into the trade-offs between classical and deep learning approaches for non-speech sound recognition.

4. Background and Motivation

Recent work in audio recognition has demonstrated that machine learning techniques originally designed for speech can be effectively applied to non-speech sounds. For example, MFCCs and spectrogram representations have been used to identify environmental sounds with encouraging results.

This project is motivated by the increasing use of context-aware systems. A smart speaker that not only responds to voice commands but also recognizes an alarm sound or a glass break could

improve safety and user experience. Similarly, monitoring urban environments (e.g., detecting sirens) has applications in transportation and emergency response.

I chose this topic because it balances feasibility and innovation:

- Feasibility: datasets and baseline implementations are readily available.
- Innovation: system design choices (feature extraction, classifier, augmentation methods) leave space for personal contribution.

5. Proposed Methodology

Tools and Platforms: MATLAB, Python, GitHub

Signal Processing Techniques:

Feature Extraction: MFCCs, Chroma features, Spectral Centroid, Mel-spectrograms …

Data Augmentation: time-shifting, adding background noise, pitch shifting …

Data Sources:

ESC-50 dataset (50 environmental classes, 2000 clips, 5 seconds each) and UrbanSound8k (8732 labeled clips across 10 classes) …

6. Expected Outcomes

A working prototype classifier that can recognize non-speech sounds from test audio clips.

Comparative analysis of classical ML vs CNN-based approaches.

Performance benchmarks presented through graphs and confusion matrices.

GitHub documentation and demo

7. Timeline

Week 6–7: Literature review, dataset collection, GitHub repo setup.

Week 8–9: Baseline implementation (feature extraction + classical classifiers).

Week 10–11: CNN model training and optimization (data augmentation, hyperparameter tuning).

Week 12: Evaluation, error analysis, comparison of methods.

Week 13: Final report writing, GitHub documentation, project video preparation.