ELEC5305 Project Proposal

Teacher: Craig Jin

TA: Reza Ghanavi

1. Project Title

Keyword Spotting in Noisy Environments

2. Student Information

- Full Name: Jianing Zhang
- SID: 540101436
- GitHub Username: jthhhh123-wq
- GitHub Project Link: jthhhh123-wq/elec5305-project-540101436: Project proposal for Keyword Spotting in Noisy Environments (ELEC5305)

3. Project Overview

This project aims to investigate keyword spotting (KWS) systems and their robustness in noisy environments. KWS is the task of detecting predefined spoken words from continuous speech. It is a core component in modern voice assistants, mobile devices, and hands-free systems.

The main challenge lies in maintaining reliable performance when background noise is present, such as in cafes, traffic, or public spaces. The goal of this project is to implement a baseline KWS system, evaluate its performance under noisy conditions, and explore strategies to improve noise robustness.

4. Background and Motivation

KWS systems have become increasingly important in daily life, enabling “wake words” like *Hey Siri* or *OK Google*. While existing deep learning models achieve high accuracy in clean conditions, their performance often deteriorates with background noise.

Recent research has explored noise-robust feature extraction (MFCC, spectrograms), data augmentation with artificial noise, and compact neural networks suitable for deployment on mobile and embedded devices. However, the trade-off between robustness and efficiency remains a challenge.

This project was chosen because it combines core signal processing concepts with modern machine learning, and it has strong real-world relevance in ubiquitous computing environments.

5. Proposed Methodology

The project will follow a systematic approach:

- **Tools and Platforms**: Python, PyTorch/TensorFlow, and Librosa for audio preprocessing.
- **Data Sources**: Google Speech Commands dataset for training and testing.
- **Signal Processing Techniques**: MFCCs, log-mel spectrograms, and possibly pre-trained embeddings.
- **Model Architectures**: Lightweight CNNs, RNNs, or Transformer-based models.
- **Noise Handling Strategies**:
    - Additive noise augmentation during training.
    - Controlled testing with different SNR levels.
    - Evaluation of preprocessing (e.g., spectral subtraction, Wiener filtering).

6. Expected Outcomes

By the end of the project, the following outcomes are expected:

- A baseline keyword spotting system trained on a public dataset.
- Quantitative evaluation of performance under clean vs. noisy conditions.
- Identification of techniques that improve robustness to background noise.
- A GitHub repository containing the proposal, code, results, and documentation.
- A short GitHub Pages site showcasing the project summary and outcomes.

7. Timeline (Weeks 5–10)

Week	Task
6–7	Literature review and dataset collection
8–9	Initial implementation and testing
10–11	Optimization and evaluation
12–13	Final report and GitHub documentation

8. References

1. Warden, P. (2018). Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition. Google Research.
2. Loizou, P. C. (2013). *Speech Enhancement: Theory and Practice*. CRC Press.
3. Reddy, C. K. A., et al. (2021). DNS Challenge: Improving Noise Suppression

Models. *Proc. Interspeech*.

4. Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*.