Elec5305 Project Proposal

Professor: Craig Jin

TA: Reza Ghanavi

1. Project Title

Real-Time Speech Enhancement: A Comparative Analysis of Adaptive Filtering and Machine Learning Approaches.

2. Student Information

Full name: Alaa Aldirani

Student ID (SID): 490627356

GitHub Username: Dirani0437

GitHub Project Link: https://github.com/Dirani0437/elec5305-project-490627356

3. Project Overview

This project aims to design and evaluate a real-time system for enhancing speech that has been degraded by background noise, with the primary goal of improving intelligibility in noisy environments. To achieve this, the project will implement and compare two distinct approaches: traditional adaptive filtering techniques and modern machine learning-based methods.

The performance of each approach will be measured and assessed, providing insights into their respective strengths and weaknesses, particularly in terms of speech quality, intelligibility, and feasibility for real-time deployment.

4.Background and Motivation

Speech enhancement has long been a challenge in noisy environments and consequently always evolving to meet the standards of the users. Traditional Digital Signal Processing (DSP) techniques, such as Wiener filtering and adaptive filters like Least Mean Squares (LMS), have widely been used to solve this problem. These methods are computationally efficient and perform reliably in stationary or predictable noise conditions. However, their effectiveness declines when dealing with non-stationary noise sources, such as human babble or traffic, where noise characteristics change rapidly over time.

In recent years, machine learning has emerged as a powerful alternative for speech enhancement. Unlike classical methods that rely on predefined statistical models, machine learning approaches can learn complex signal and noise patterns directly from large datasets, making it better in generalising in non-stationary environments. Despite these advantages, many deep learning models are computationally intensive, making it harder to run in real-time and resource constrained systems.

This trade-off between efficiency and performance motivated the focus of my project. By comparing traditional adaptive filtering techniques with modern machine learning approaches, the project aims to evaluate not only their speech enhancement performance but also their feasibility in real-time applications. My initial interest in this topic was initiated by the project proposal example, which detailed a system for real-time noise suppression on mobile devices.

5.Proposed Methodology

The project will be conducted in two main phases, followed by a final evaluation.

Tools and Platforms:

1. MATLAB: The primary environment for algorithm development, simulation, and real-time processing, utilizing the Audio Toolbox and Deep Learning Toolbox.

2. Data Sources: The NOIZEUS or DEMAND noisy speech corpora will be used for training and testing, as they provide a standardized set of clean speech and diverse noise types.

Phase 1: Adaptive Filtering Implementation:

1.Noise Estimation: A Voice Activity Detector (VAD) will be implemented to distinguish between speech and non-speech segments, allowing for continuous estimation of the noise floor statistics.

2.Frequency-Domain Processing: The Short-Time Fourier Transform (STFT) will be used to analyse the signal in the time-frequency domain, which is essential for spectral modifications.

3.Filtering: Both a Wiener filter and a Normalized Least Mean Squares (NLMS) adaptive filter will be implemented to suppress the estimated noise spectrum.

Phase 2: Machine Learning Implementation

1.Model Architecture: A shallow Convolutional Neural Network (CNN) will be designed using MATLAB's Deep Learning Toolbox.

2.Training: The model will be trained on pairs of noisy and clean speech spectrograms. Its goal will be to predict a time-frequency mask that, when applied to the noisy input, it will reconstruct the clean speech spectrogram.

3.Inference: The trained model will be used to process unseen noisy audio samples for evaluation.

6.Expected Outcomes

The project will culminate in a set of clear deliverables:

- A functional **MATLAB prototype** capable of performing speech enhancement in near real-time using both the adaptive filter and machine learning approaches.
- A comprehensive **comparative analysis** of the two methods, in the following criteria
  - **Signal-to-Noise Ratio (SNR)** improvement
  - **Perceptual Evaluation of Speech Quality (PESQ)** scores.
- A final **written report** on the project.
- A public **GitHub repository** containing well-documented MATLAB code.
- A brief **video demonstration** showcasing the system's performance on various audio samples.

7.Timeline

The project will adhere to the following schedule:

Week	Task
6-7	Literature review; set up MATLAB environment; download and prepare datasets.
8-9	Implement and test the adaptive filtering methods (VAD, STFT, Wiener, NLMS).

10-11	Design, train, and test the machine learning model.
12-12	Finalize comparative evaluation, write the report, and create the demo video.

8.References

3. Loizou, P. C. (2013). Speech Enhancement: Theory and Practice. CRC Press.
4. Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator.
5. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(6), 1109-1121.
6. Wang, D., & Chen, J. (2018). Supervised speech separation based on deep learning: An overview. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(10), 1702-1726.