ELEC5305 Project Proposal One

1. Project Title:

Practical Audio Analysis and Filtering with MATLAB

2. Student Information:

Full Name: Xiaoyu Hu

Student ID (SID): 530598444

GitHub Username: Dimmerbulb

GitHub Project Link: https://github.com/Dimmerbulb/elec5305-project-530598444

GitHub Pages Link: https://dimmerbulb.github.io/elec5305-project-530598444/

3. GitHub Project Site:

This project aims to build a simple MATLAB system that can reduce reverberation in recorded speech. The problem is that when we record in real rooms, echoes and reflections make the speech sound blurry and hard to understand. This is important because clear speech is needed for daily communication and also for applications like online meetings or speech recognition. To solve this, I plan to measure the room impulse response using an exponential sine sweep (ESS) and then apply short-time Fourier transform (STFT) filtering methods such as spectral subtraction and Wiener filtering. The goal is to make the speech clearer and easier to understand in practical situations.

4. Background and Motivation

When we record speech in real rooms, there are always echoes and reverberation that make the sound less clear. Some existing solutions, like exponential sine sweep (ESS) for measuring room impulse response [1] and short-time Fourier transform (STFT) for time–frequency analysis [2], are commonly used in research and practice. Filters such as pre-emphasis and methods like spectral subtraction [3] have also been studied for improving speech quality.

I chose this topic because in the lab I already learned how to do audio I/O, run ESS recordings, and use STFT to analyze signals [2]. These gave me a basic understanding of how sound can be recorded, processed, and improved. Working on dereverberation feels like a natural next step, since it lets me apply what I learned in class to a real problem—making sound easier to understand in everyday environments like classrooms or offices.

5. Proposed Methodology

For this project I will mainly rely on MATLAB, because I have already practiced with it in the labs when working on audio I/O, exponential sine sweep (ESS), short-time Fourier transform (STFT), and feature extraction. MATLAB is convenient for signal processing tasks since it provides built-in functions for spectrograms, filtering, and recording, and it also makes it easy to visualize results with clear plots. If needed for evaluation or additional metrics, I may also use Python later, but the core implementation will be in MATLAB.

In terms of signal processing techniques, the first step will be to measure or simulate the room impulse response (RIR). I will do this using the ESS method with inverse filtering, something I already tested in the lab, since it provides a good estimate of how sound behaves in a room. Once I have the RIR, I will process the speech in the STFT domain, because this representation makes it possible to see how reverberation spreads energy across time and frequency. For dereverberation, I plan to start with spectral subtraction and Wiener filtering, as they are well-known baseline methods. If time permits, I will also test a log-MMSE enhancement approach to compare performance. Additionally, I will use pre-emphasis and de-emphasis filters where needed to balance high-frequency content, since these techniques are common in speech processing and I already applied them in previous lab work.

For data sources, I will create my own recordings by speaking in different small rooms and capturing ESS signals to estimate their RIRs. This will provide realistic test material. To supplement these recordings, I will also convolve clean speech from public datasets such as TIMIT or VCTK with the measured RIRs. This way, I can make controlled comparisons between clean, reverberant, and enhanced versions of the same utterance. Finally, if time allows, I might also test the system on larger open datasets like the DNS Challenge dataset, which includes noisy and reverberant speech, to see how robust my methods are under more challenging conditions.

6. Expected Outcomes

At the end of this project, I plan to build a MATLAB program that can not only record audio but also measure the room impulse response (RIR) using an exponential sine sweep. Based on the estimated RIR, the program will apply STFT-based filters, such as spectral subtraction and Wiener filtering, to reduce reverberation and make the speech signal clearer. I expect to present both listening examples and visual evidence of improvement. For example, the processed speech should sound sharper and easier to understand than the original recordings, and the spectrograms should show reduced smearing of energy over time.

In addition to the program, I will prepare a small set of test recordings, including both my own speech and speech samples convolved with measured RIRs, so that the experiments can be reproduced. The results will include before-and-after comparisons, spectrogram plots, and simple objective metrics such as PESQ, STOI, or SI-SDR to support the evaluation. Finally, I will organize all code, data, and the proposal in a GitHub repository, together with a short README and GitHub Pages site, so that others can easily run my scripts, repeat the experiments, and understand the workflow.

7. Reference

[1] Farina, A. (2000). *Simultaneous measurement of impulse response and distortion with a swept-sine technique*. AES Convention.

[2] Oppenheim, A. V., & Schafer, R. W. (1999). *Discrete-Time Signal Processing*. Pearson.

[3] Boll, S. F. (1979). *Suppression of acoustic noise in speech using spectral subtraction*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 27(2), 113–120.