Real-Time Speech Enhancement for Online Communication Using Deep Learning

Full Name: Jiahao Xu
SID: 530332675
GitHub Username: chlorisdraco
GitHub Project Link: https://chlorisdraco.github.io/elec5305-project-530332675/

1. Project Overview

Online meetings are now common. Platforms such as Zoom or Teams are used every day. But speech quality often drops when there is noise. Sounds like traffic, keyboard clicks, or a fan can make speech hard to understand.

This project will design a system to improve speech clarity in real time. The system will use deep learning and signal processing. The aim is to remove noise but keep the speech natural. A final Python prototype will show the result with both audio examples and measured improvements.

2. Background and Motivation

Speech enhancement has been studied for many years. Classic methods such as Wiener filtering or spectral subtraction work well only for simple noise. They fail when noise changes quickly.

Noise in speech is a big problem in daily life. People take online classes, join meetings, or talk with friends using mobile devices. In all these cases, background sounds reduce clarity. A small fan in a room or cars outside a window can change the listener's experience. When speech is not clear, people get tired and lose focus. This shows why speech enhancement is important not only for science but also for people's lives.

Classic methods cannot solve this fully. They assume the noise is fixed and stable. But in reality, noise changes fast. A door may close, a dog may bark, or a child may shout. Deep learning can adapt better to these changes. The challenge is that many deep models are too heavy. They need strong computers and cannot run in real time. This project will explore a lighter model that still gives good results. It can help many users without strong hardware.

Deep learning has made big progress in this field. Models such as convolutional recurrent networks give much better results. But many models are too large and slow. They are not practical for real-time use.

This project is motivated by the need for efficient systems. A small but effective model can help in real online meetings. It fits well with the course goal, which is to turn theory in acoustics and signal processing into useful tools for people.

3. Proposed Methodology

The project will follow five main steps. First, preprocessing will use the Short-Time Fourier Transform (STFT) to convert speech into a time–frequency form, with framing and normalisation of the input. Second, a small deep network such as a CRN or SE-ResNet will be built, taking a noisy spectrogram as input and producing a clean speech mask as output. Third, training will use the DNS Challenge dataset for noisy speech and the TIMIT dataset for clean speech, to cover a wide range of noise and speaker conditions. Fourth, implementation will be carried out in Python with PyTorch for network training, while MATLAB will be used to test classic baseline methods. Finally, performance will be evaluated using PESQ, STOI, and SNR, which together will show both speech quality and intelligibility improvements.

4. Expected Outcomes

The project will not only create a prototype but also compare methods in detail. The model will be tested with both simple and complex noise. This will show how well it works in realistic conditions. The results will include audio samples so others can listen and judge the difference. A table of metrics such as PESQ, STOI, and SNR will also be given. This will make the evaluation more complete.

The GitHub repository will be easy to use. It will have a README file, sample scripts, and test files. This will allow other students and researchers to try the system. They can also extend it for their own work. By sharing the code and report, the project will add value to the learning community.

5. Timeline (Weeks 6–13)

TASKS：	SIX	SEVEN	EIGHT	NINE	TEN	ELEVEN	TWELVE	THIRTEEN
Review research and collect datasets	▬	▬						
Build first model and test baseline methods			▬	▬				
Improve model and run evaluation					▬	▬		
Write report, prepare GitHub and demos							▬	▬

References

1. Loizou, P. C. (2013). Speech Enhancement: Theory and Practice. CRC Press.

2. Reddy, C. K. et al. (2021). "DNS Challenge: Improving Noise Suppression Models." Proc. Interspeech.

3. Ephraim, Y., & Malah, D. (1984). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator." IEEE Trans. Acoustics, Speech, and Signal Processing, 32(6), 1109–1121.

4. Tan, K., & Wang, D. (2019). "Learning Complex Spectral Mapping with a Convolutional Recurrent Network for Speech Enhancement." IEEE/ACM Trans. Audio, Speech, and Language Processing, 27(12), 1996–2008.