**Department of Electrical, Computer, and Software Engineering**

**Part IV Research Project**

Project Compendium Report

Project Number: 58

The reconstruction of speech/voice with the use of noise-cancelling algorithms and machine learning

Edward Chan

Timothy Aguana Cabrera

Catherine Watson

Yusuke Hioka

**Declaration of Originality**

This report is my own unaided work and was not copied from nor written in collaboration with any other person.

A close-up of a signature

Description automatically generated

Name: Edward Chan

A black line drawing of a fish

Description automatically generated

Name: Timothy Aguana Cabrera

Abstract: This contribution addresses enhancing speech quality in noisy environments with low signal-to-noise ratios (SNR). The approach consists of two key components. First, it leverages codebooks derived from Gaussian mixture models (GMMs) and a Wasserstein Generative Adversarial Network (WGAN) to generate time-varying spectral candidates that closely resemble noise spectra. While the estimation of noise spectra is accurate, the estimation of speech spectra needs to improve, resulting in distorted and unintelligible Wiener-filtered mixtures.

In the realm of audio signal processing, low SNR leads to significantly quieter speech, posing a problem. To tackle this issue, the paper proposes an iterative learning approach that combines two machine learning algorithms for speech enhancement and classification in noisy mixtures. The first algorithm employs a WGAN for end-to-end learning to generate noise spectra that can be used to create enhanced speech. Simultaneously, considering SNR levels, a convolutional neural network (CNN) uses the spectrogram representation of the estimated clean speech to classify between clean and noisy speech.

The primary objective of this paper is to explore whether improvement to the performance of the codebook-based approach can be used with SNR levels to indicate improved speech quality and intelligibility. To evaluate the algorithm's performance, the proposed method utilises computational matrices, including perceptual evaluation of speech (PESQ), short-time objective intelligibility (STOI), and signal distortion ratio (SDR).