CSE 848: Project Proposal
Evolutionary Computation for Speech Enhancement

Vince Fasburg and Josh Thomas
Dept of Computer Science and Engineering, Michigan State University
East Lansing, MI, USA
{fasburgv thom1212} @msu.edu

1. PROBLEM DEFINITION

Speech enhancement aims to improve the quality of digital recorded or streaming speech in realistic, often noisy environments. The quality of speech can be defined in multiple ways for different applications, but the objective is usually to improve clarity, intelligibility, and pleasantness. This is typically accomplished by first identifying, then removing the background noise.

Applications of speech enhancement include microphones in cars that are used to make phone calls and communications from airline pilots to air traffic control. Removal of any noise in such situations is extremely useful in order to focus on the voice of the driver or pilot. This project will focus on the speech enhancement of several clips in which background noise has been added to clean speech in order to simulate some of these common speech enhancement scenarios.

The future findings of this project should reveal an optimal strategy to employ some of the existing speech enhancement algorithms to recover a clean speech signal from noisy speech by determining the optimal parameter values to be used within the algorithms as well as the best way to combine the results of these algorithms for the cleanest overall signal.

2. APPROACH

The team will approach this problem by starting with the simplest form of the strategy and then build upon it. This section describes both the initial simplest version of the proposed strategy, and how it will be improved upon to create the finished product.

2.1 Evolutionary Strategy Design

An evolutionary strategy approach will be used to evolve several real-valued parameters that are internal to the algorithms discussed in the next section, as well as parameters representing how the algorithms are applied, such as the order in which each algorithm will be applied and coefficients representing the amplitude of each algorithm output to be used to produce the final signal.

The fitness function used to determine the quality of each individual in the population will be a simple difference between the desired signal (the clean speech to which the noise was added) and the output of the speech enhancement with the parameters from that individual. A different test case will initially be chosen randomly for each evaluation. If this does not prove to be a robust method of determining fitness, other strategies will be considered such as using multiple test clips on each fitness evaluation.

The team intends to implement several variations of the evolutionary strategy at each stage of the project to determine what methodologies produce the best results in a reasonable amount of time. Parameters that will be varied include the μ and λ values, replacement strategy, and number of generations allowed. If time allows, this strategy will be improved further through the use of Covariance Matrix Adaptation.

2.2 Candidate Algorithms

This section briefly describes some of the common forms of speech enhancement algorithms and how they may be used and optimized as part of our evolutionary strategy.

2.2.1 Spectral Subtraction

Spectral subtraction algorithms come in many flavors, however, the basic idea is always the same. The goal of spectral subtraction is to estimate energy of the noise present in the signal at different frequency bands. Using a Fourier Transform, the input signal is converted into the frequency domain, the estimated spectrum of the noise is subtracted from the total signal, and the inverse Fourier Transform then returns the signal to its original form. There are several different opportunities for tuning parameters for best performance in this algorithm, depending on exactly what variation of the algorithm is used. These differences depend mainly on how the spectrum of the noise is estimated to how "musical tones" (tonal-sounding artifacts that are created from subtracting the noise spectrum when the estimate is imperfect and the timing is imprecise).

2.2.2 Wiener Filtering

The Wiener filter can be used to filter out the noise from a corrupted signal and provide an estimate of the underlying signal. It is based on a statistical approach to minimize the mean square error between the estimated underlying signal and the desired signal. The specific Wiener filter that the team will be using tracks the signal to noise ratio to determine how much noise is in the current signal. Gains are then calculated to be proportional to the signal to noise ratio then are applied to the noisy signal. The initial noise spectrum is computed at the beginning where the program assumes that the initial signal has only sound. The window of time that the noise spectrum is measured can be varied in the code.

2.2.3 Noise Gating

Noise gating is a simple technique by which quiet sections of the signal are attenuated while louder sections are unaffected. This technique works best for noise that is relatively quiet compared to the desired signal, but may work well in combination with the other algorithms. There are several parameters that could be tuned on a noise gate including the following.

1) Threshold – The amplitude below which the signal will be attenuated.
2) Ratio – The ratio of dB attenuation/dB under the threshold.
3) Attack Time – The time it takes for attenuation to occur once the signal has gone below the amplitude threshold.
4) Release Time - The time it takes for attenuation to cease once the signal has gone above the amplitude threshold.

3. PREVIOUS ART

The speech enhancement algorithms discussed in this proposal and their many variations are thoroughly documented in the book “Speech Enhancement Theory and Practice 2nd Edition” by Philipos C. Loizou, which provides in depth discussions of the development and performance of these algorithms.

Evolutionary Strategies (ES) are based on optimization techniques such as adaptation and evolution. The primary search operators are mutation and selection. These operators are applied for multiple generations or loops until an end condition is met. Mutation is typically performed by adding a normally distributed random value to each vector component, and selection is based on the fitness rankings of individuals. Some of the initial ES techniques were developed in the 1960s and 1970s by Ingo Rechenberg and Hans-Paul Schwefel. They applied this technique to several challenging problems such as aerodynamic wing design as well as other applications related to fluid dynamics.