Evolutionary Computation for Speech Enhancement

By: Vince Fasburg and Josh Thomas

Overview

- Introduction
- Background
  - CMA-ES
  - Weiner Filter
  - Spectral Subtraction
  - Noise Gate
- Evolutionary Strategy
- Results
- Conclusions
- Future Work

Introduction

Problem

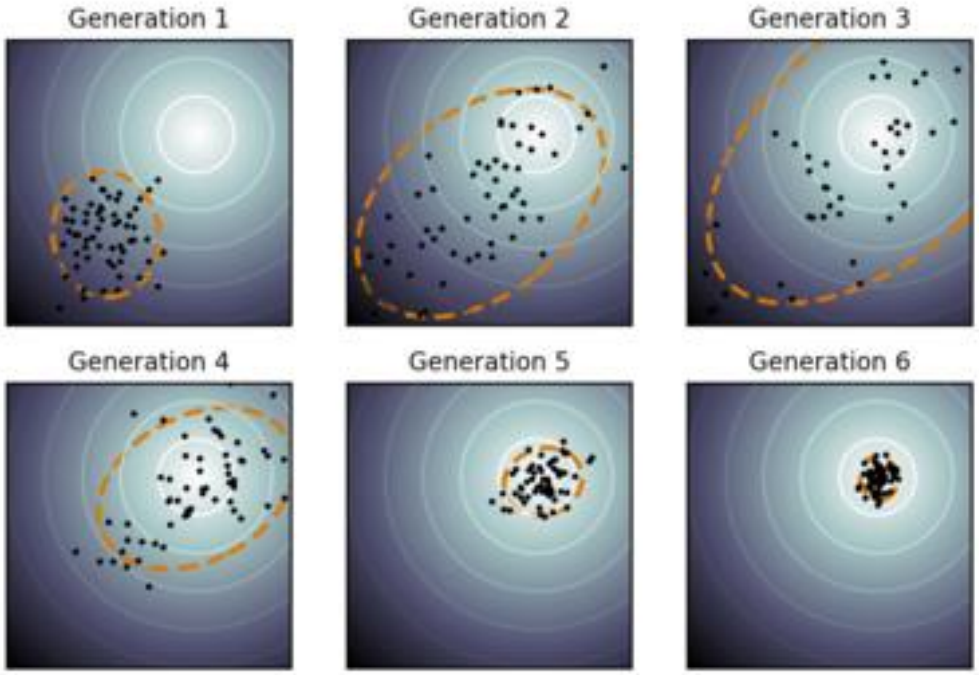
Recover clean speech signal from audio corrupted by noise, such as road noise.

Approach

Combine several existing de-noising algorithms, using evolutionary strategies to tune parameters to recreate clean signal.

Background: CMA-ES

- Real-valued parameters make up individuals
- Random mutation from mean every gen
- Mutation in direction of eigenvectors of covariance matrix



Background: Spectral Subtraction

- **Basic Algorithm:**
- Subtract power spectrum of noise only from that of noisy signal

$$X = (Y^2 - N^2)^{1/2}$$

- **Improvements:**
- Y values smoothed
- Speech/Noise decision with hold times
- Parameters: Noise Margin, Hang Over, Smoothing Factor

Background: Wiener Filter

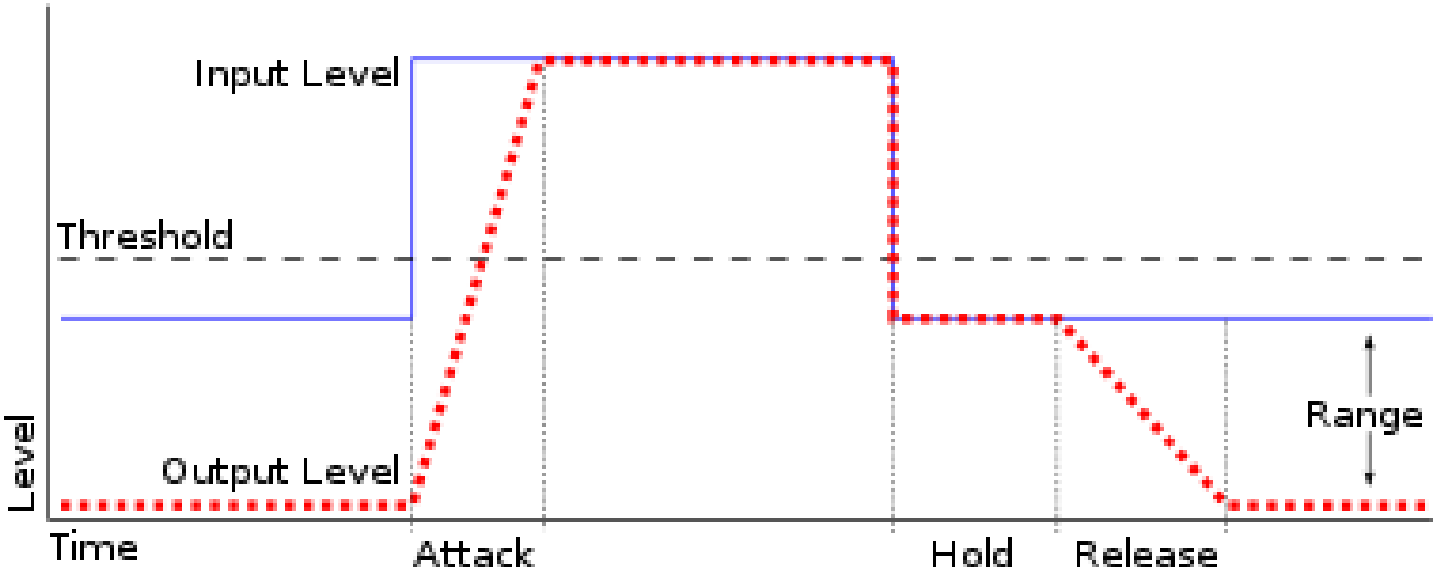
- Estimate SNR $=$ Y(f)/N(f)
- Use SNR to find best gains for FIR filter
- Apply filter to eliminate noise
- Parameters: SNR Smoothing Factor

$$y(t) = (h \star x)(t) + n(t)$$

$$G(f) = \frac{1}{H(f)}\left[\frac{|H(f)|^2}{|H(f)|^2 + \frac{N(f)}{Y(f)}}\right]$$

$$\hat{x} = ifft(G(f)Y(f))$$

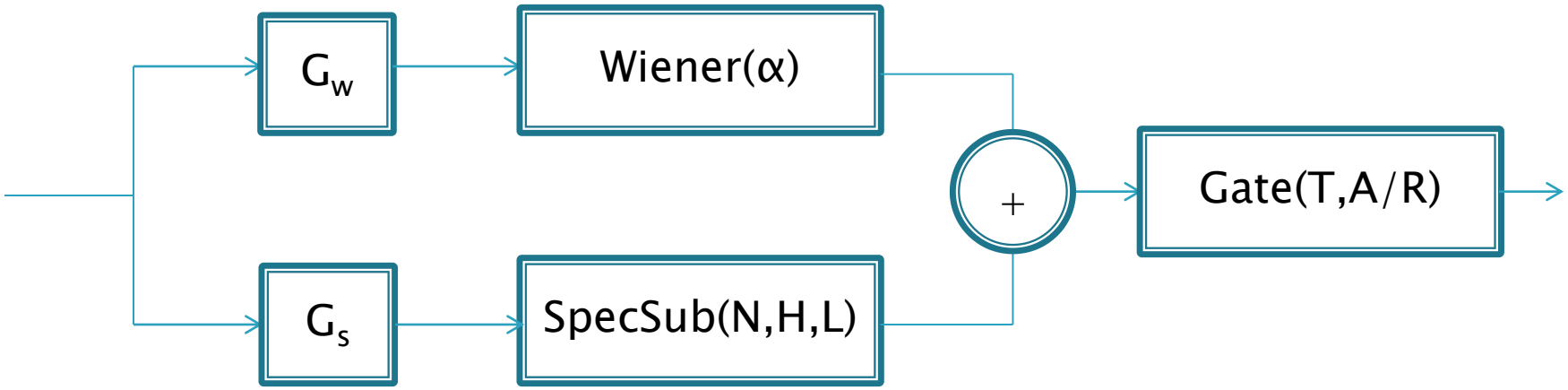
Background: Noise Gate



- Parameters: Threshold, Attack, Release

Evolutionary Strategy

- (μ, λ) CMA-ES with weighted recombination
- μ = 6,12,24 parents, λ = 12,24,48 offspring
- Each offspring contains 8 parameters



Fitness Function

- Average of the differences between the clean audio file and the processed audio file.

$$Fitness = \frac{1}{n}\left(\sum_{1}^{n} ABS(Clean - Processed)\right)$$

- Lower fitness is desired since it will have less difference from the clean file.

Evolutionary Strategy

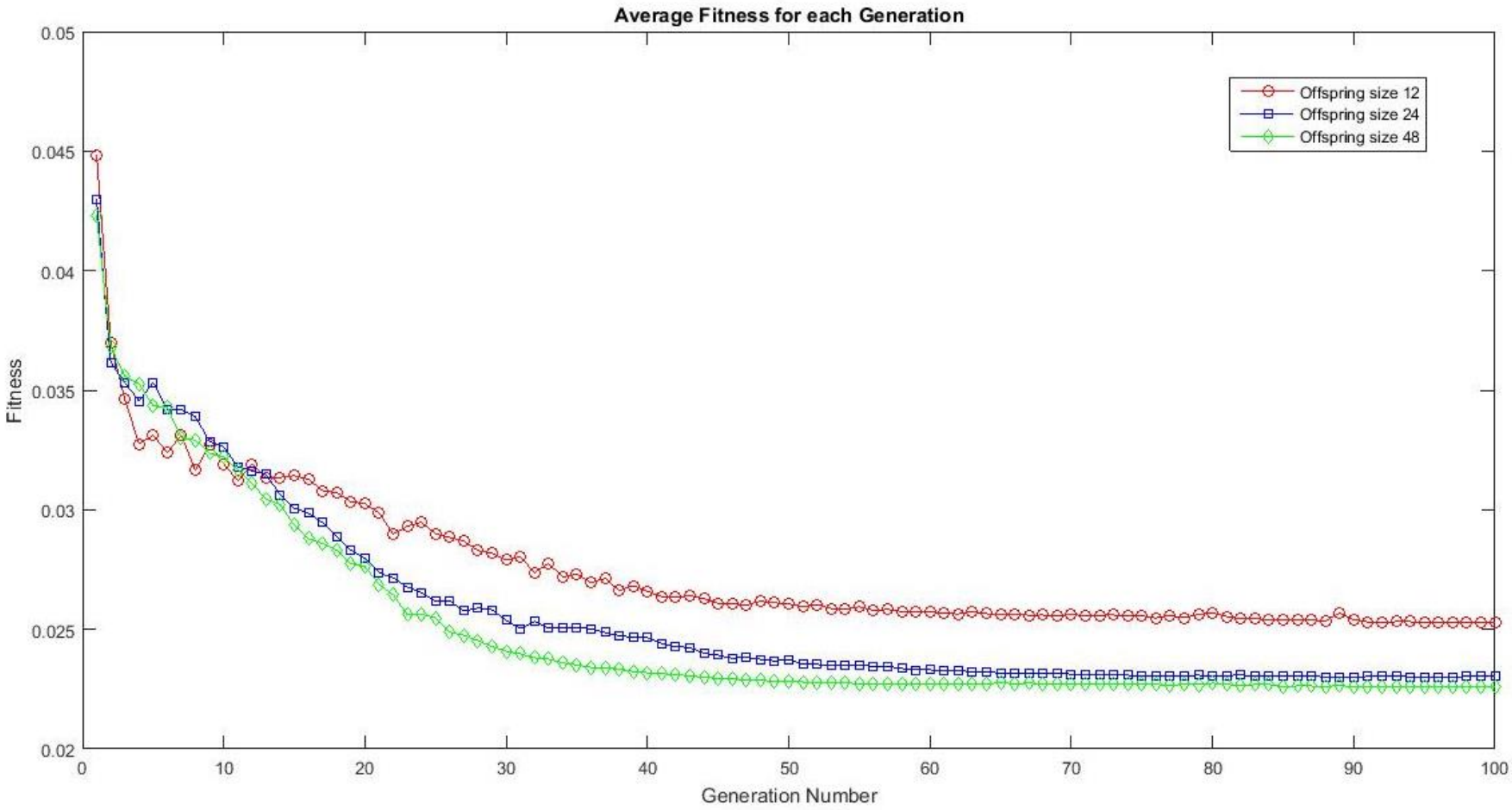
- Each parameter scaled differently from [0,1]

Parameter	Min	Max	“Expert” Value
Wiener Smoothing	0	2	0.98
Wiener Gain	-2	2	N/A
Spec Sub Gain	-2	2	N/A
Noise Margin	0	20	8
Noise Length	0	20	18
Hangover	0	20	3
Threshold	0	0.25	N/A
Attack/Release	0	10	N/A

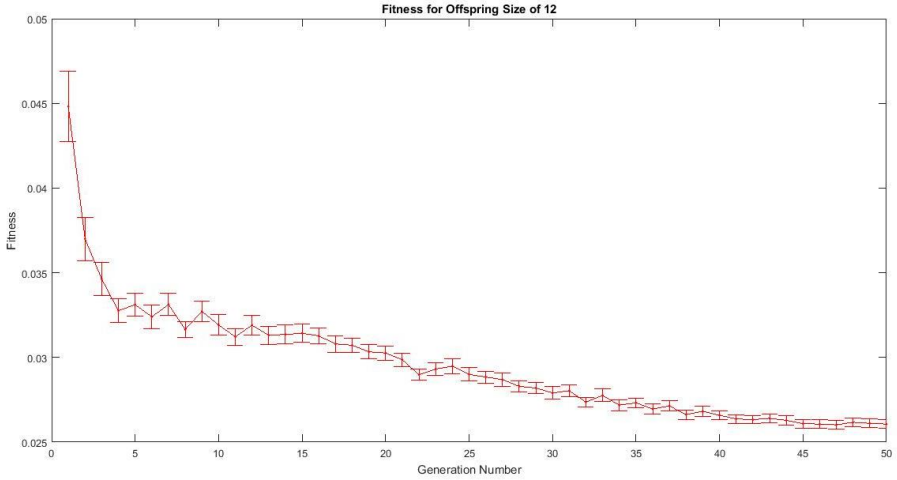
Results

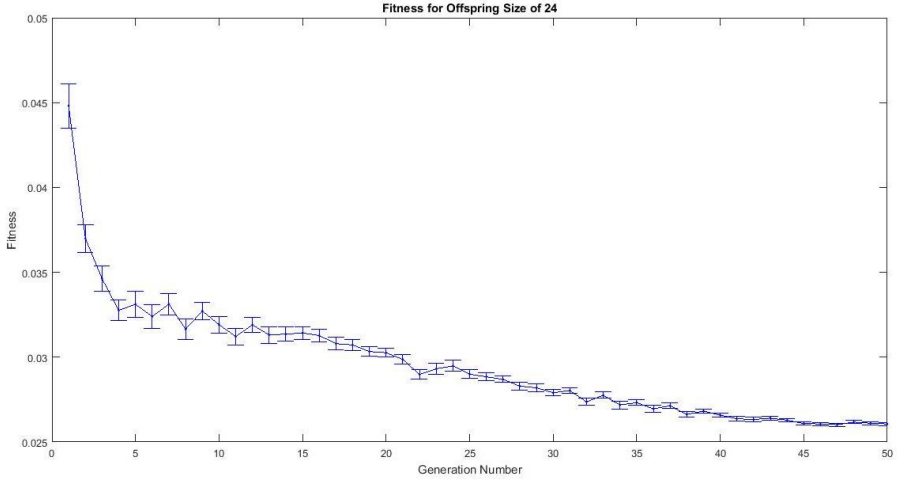
- Three different run configurations
  - 1: Offspring of 12
  - 2: Offspring of 24
  - 3: Offspring of 48
- 50 runs per configuration
- 300 Generations per run

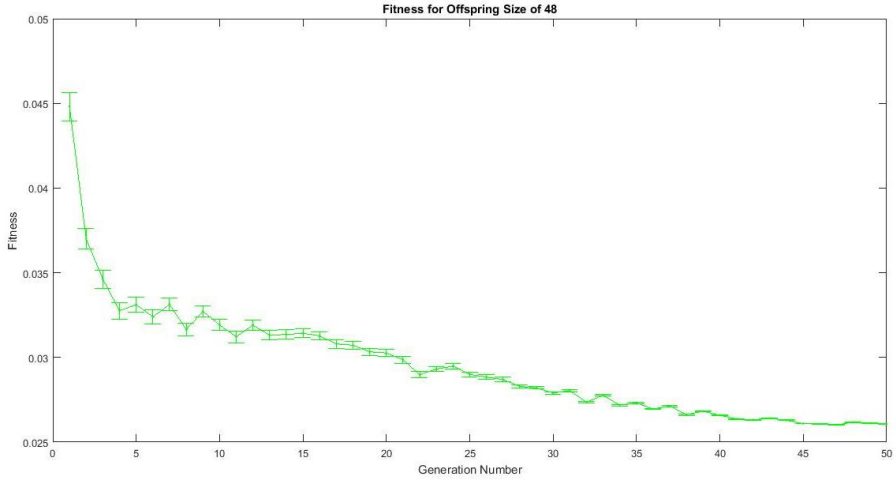
Results


Average Fitness for each Generation
Offspring size 12
Offspring size 24
Offspring size 48
Fitness
Generation Number
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0
10
20
30
40
50
60
70
80
90
100

Results


Fitness for Offspring Size of 12
Fitness
Generation Number
0.05
0.045
0.04
0.035
0.03
0.025
0
5
10
15
20
25
30
35
40
45
50


Fitness for Offspring Size of 24
Fitness
Generation Number
0.05
0.045
0.04
0.035
0.03
0.025
0
5
10
15
20
25
30
35
40
45
50


Fitness for Offspring Size of 48
Fitness
Generation Number
0.05
0.045
0.04
0.035
0.03
0.025
0
5
10
15
20
25
30
35
40
45
50

Results

	Expert Results	Expert Results	Best Pop 12	Best Pop 24	Best Pop 48
Wiener Gain	1	0	1.10	0.97	1.10
Wiener Smoothing	0.98	-	0.95	0.96	0.95
Spec Sub Gain	0	1	0.41	0.45	0.41
Noise Length	-	20	13.65	18.92	0.69
Noise Margin	-	8	18.23	0.78	15.35
Hang Over	-	3	9.33	0.58	18.16
Threshold	-	-	0.03	0.03	0.03
Attack/Release	-	-	1.25	0.96	1.25
Fitness	**0.0301**	**0.0316**	**0.0225**	**0.0223**	**0.0225**

Results

Running the best results with only the Wiener Filter

	Expert Results	Best Results
Wiener Gain	1	1
Wiener Smoothing	0.98	0.96
Spec Sub Gain	0	0
Fitness	**0.0301**	**0.0270**

Results

Running the best results with only Spectral Subtraction

	Expert Results	Best Results
Wiener Gain	0	0
Spec Sub Gain	1	1
Noise Length	20	18.92
Noise Margin	8	0.78
Hang Over	3	0.58
Fitness	**0.0316**	**0.0354**

Results

Clean File 

Dirty File 

Processed File 

Conclusions

- Improve upon the expert results with ES
- Difficult to compare to expert results
- Fitness is improved from expert results
  - Computation time of EC may not be worth it
- Wiener Filter is better than Spectral Subtraction at reducing noise

Future Work

- New fitness function
  - Tuned to human hearing
  - For computer interpretation
- Evolve different combinations of de-noising algorithms alongside parameters
  - Genetic programming
- Variety of audio samples
  - Different voices
  - Different noise situations

Questions?