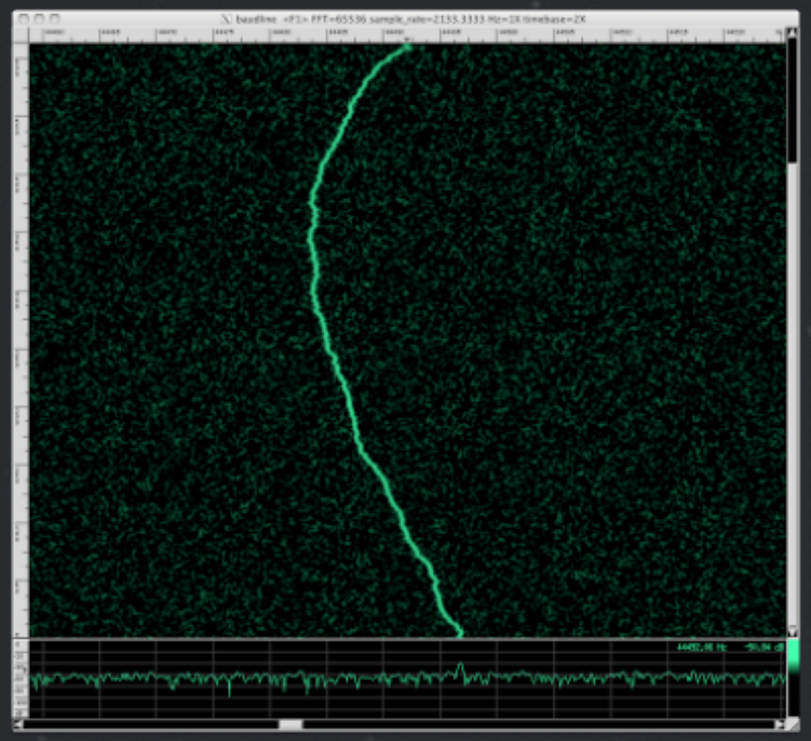
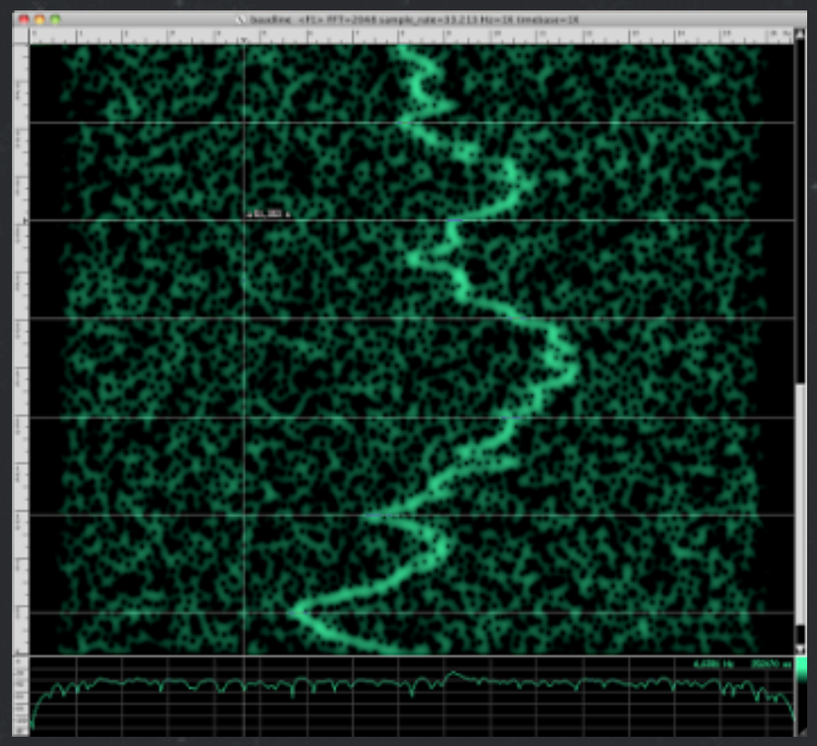
Identifying Known and Unknown Signals from the SETI Dataset

Frank Fan, Kenny Smith, Jason Wang
Advisor: Jeffrey Ullman

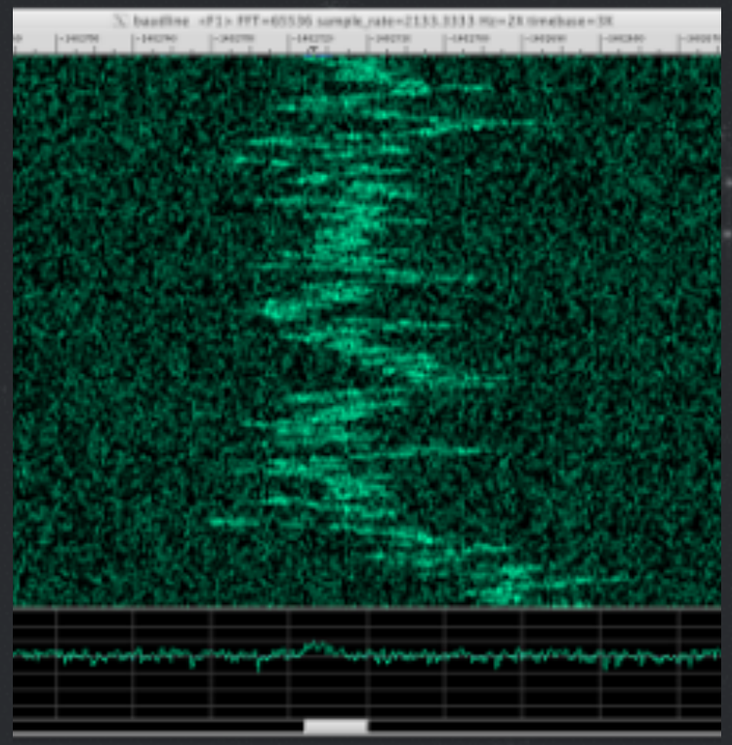
Dataset



Slow Modulation Squiggle



Classic Squiggle



Rapid Modulation Squiggle

>380,000 spectrogram images
833 hand-curated squiggle examples

Question 1: How can we use image processing to convert spectrogram waterfall plots into discrete time series data?

Question 2: How can we build a classifier to identify new “squiggle” signals from incoming spectrogram images?

Approach

1. Spectrogram conversion into discrete time series

2. Discrete Time Fourier Transform (DTFT) to decompose signals into component frequencies

3. Feature extraction using FT & non-FT data

4. Fit a classifier to separate squiggle v.s. non-squiggle

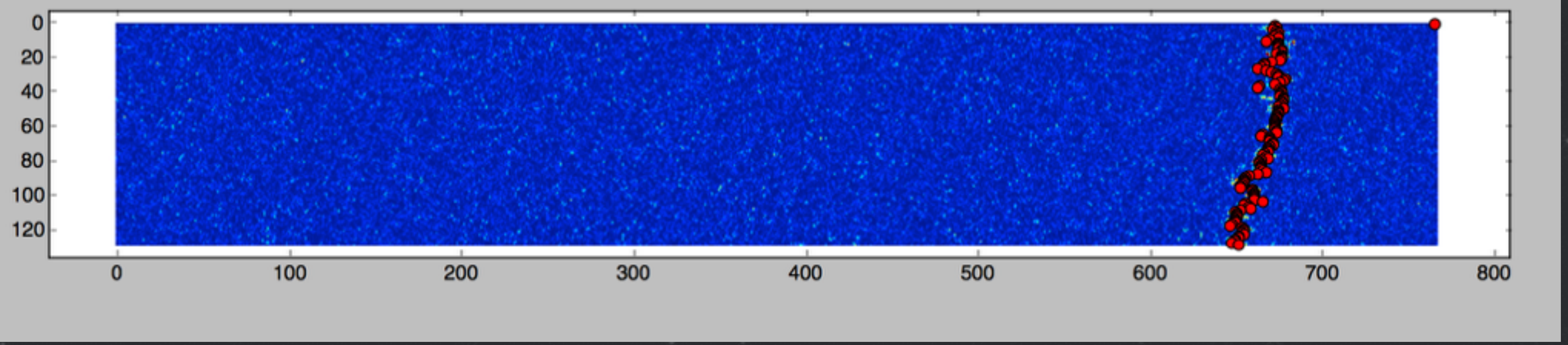
5. Unsupervised learning to identify squiggle subgroups

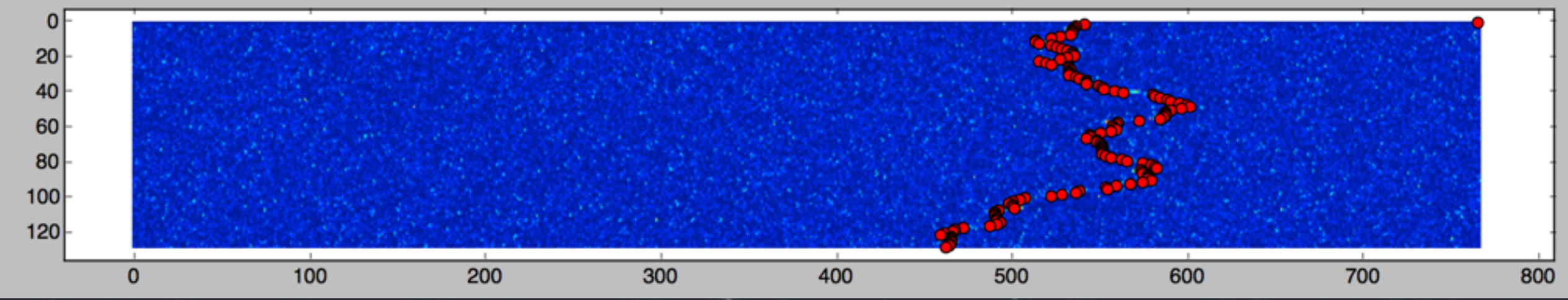
Progress

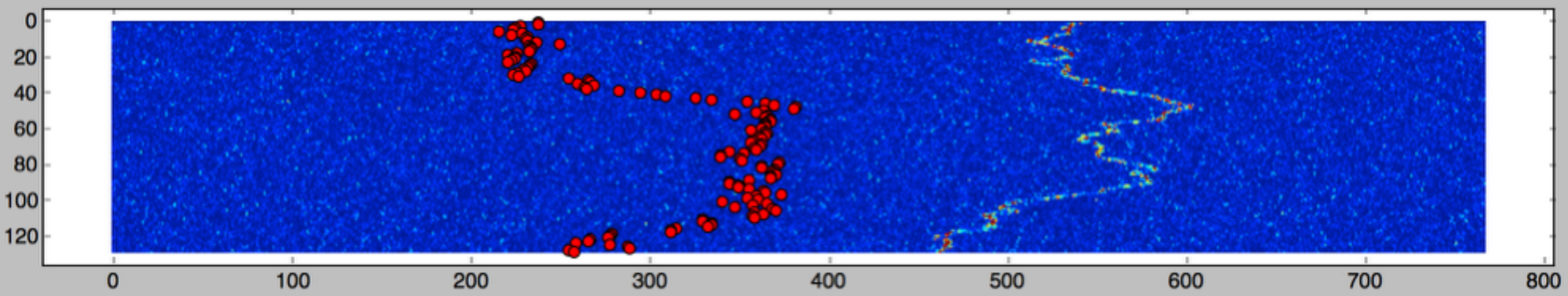
1. Explored generative models: Brownian motion with drift/noise, autoregressive

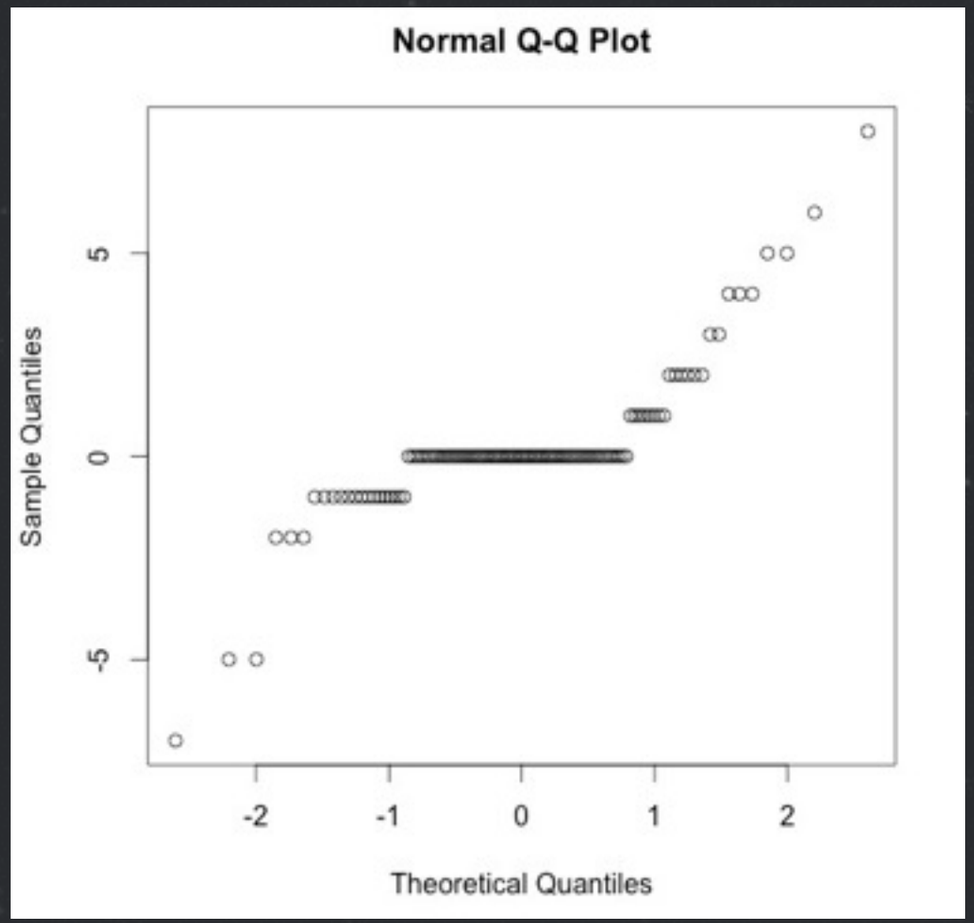
2. Derived dynamic programming algorithm & loss function for time series discretization.

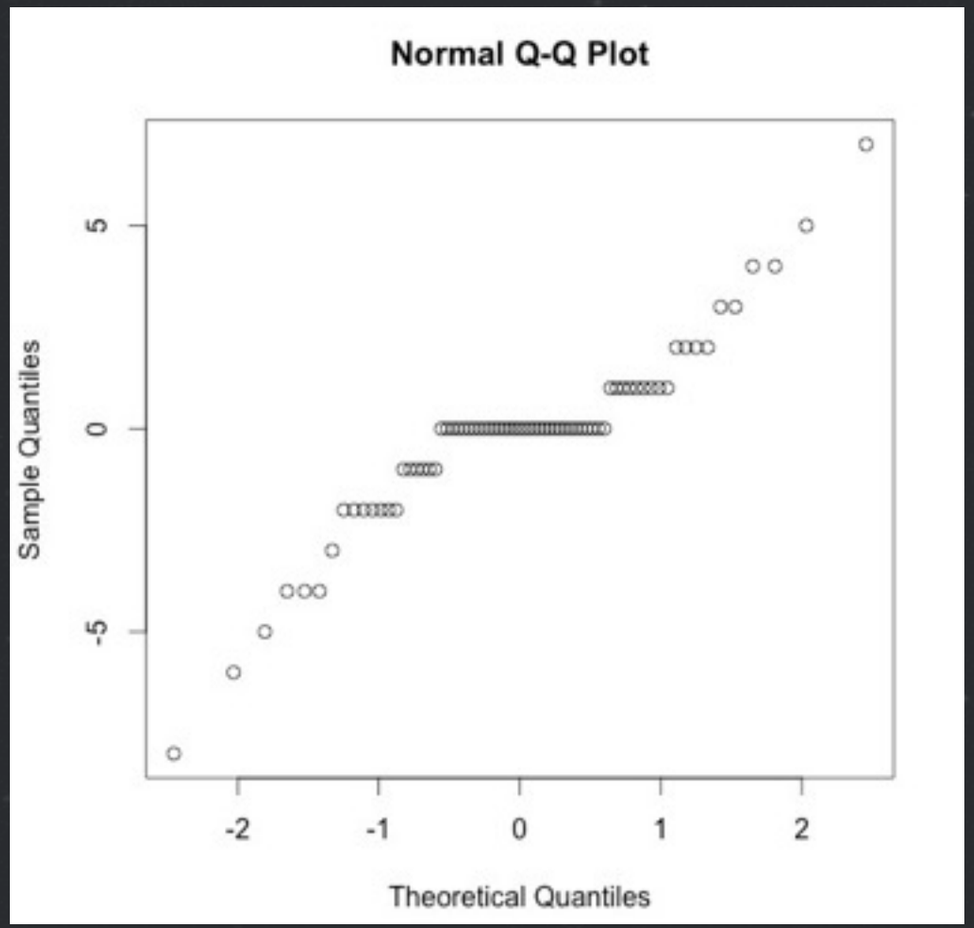
3. Applied DTFT on squiggle and non-squiggle images

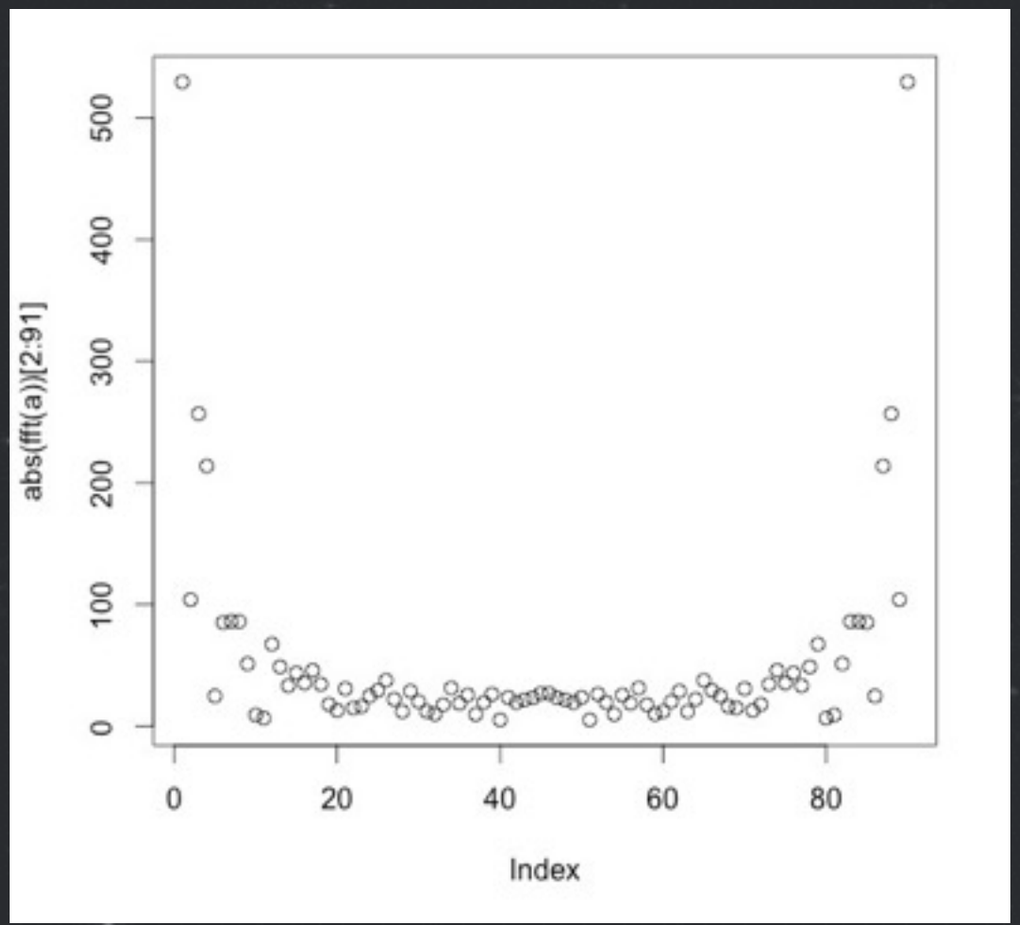

0
20
40
60
80
100
120
0
100
200
300
400
500
600
700
800


0
20
40
60
80
100
120
0
100
200
300
400
500
600
700
800


0
20
40
60
80
100
120
0
100
200
300
400
500
600
700
800


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-5
0
5
-2
-1
0
1
2


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-5
0
5
-2
-1
0
1
2


abs(fft(a))[2:91]
Index
0
100
200
300
400
500
0
20
40
60
80

Future Directions

1. Parallelize discretization algorithm

2. Feature Extraction: parameters of AR model, band intensity, etc.

3. Fit a classifier using logistic regression, linear discriminant analysis, etc. to separate squiggle & non-squiggle

4. Apply K-means, KNN, self-organizing maps, etc. to identify characteristic groups within squiggles