EMG Spike Decomposition

Shakthi Visagan
804 622 954
M260
Liu

Introduction

- Contracting muscles are evidence of neuromuscular activity
- We can use electromyography (EMG) to measure the and record these signals
- Several electrodes will record the firing of action potentials or motor unit action potentials (MUAPs) and then amplify the signal using filters and amplifiers
- Identifying the specific neurons that fired the spikes will lead to better diagnoses of neuromuscular diseases, and may even contribute to the theory of neuroscience, as evidenced by the discovery of the Rosehip Neuron, where spike sorting played a critical role in identifying the unique Rosehip spikes

Problem

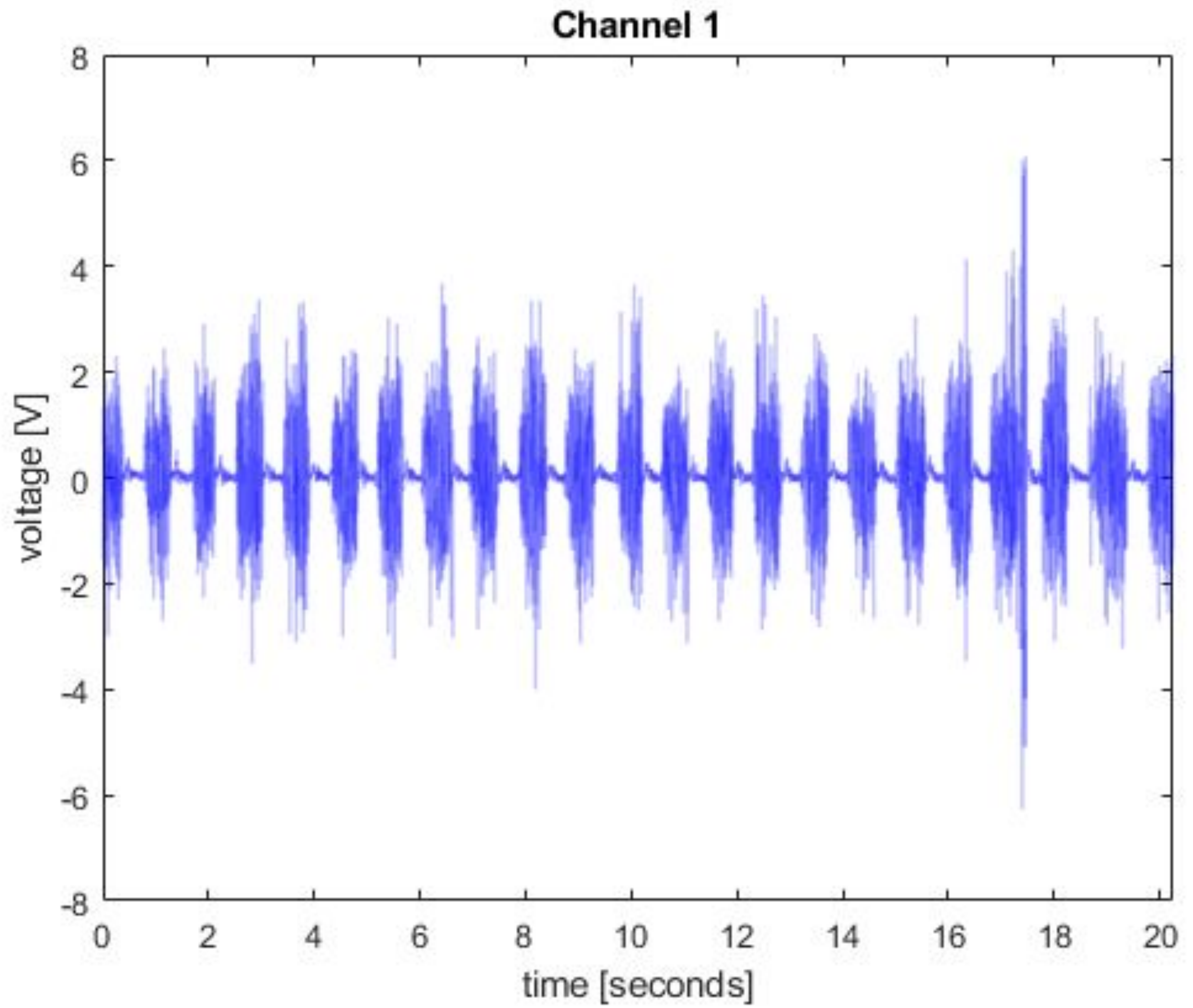
DETECT, SEPARATE, IDENTIFY SEPARATE SPIKES ACCORDING TO THE NEURONS THAT MAY HAVE FIRED THE RESPECTIVE ACTION POTENTIAL.

Workflow

1. Administrative Code
2. Input File
3. Filtering the Signal by Removing Frequencies
4. Detecting Spikes
5. Aligning Spikes
6. t-SNE Latent Space Measurements (Preliminary feature extraction/clustering)
7. PCA Dimensionality Reduction & Feature Extraction
8. Initial *k*-Means Clustering
9. Final *k*-Means Clustering & Analysis

Administrative Code & Input File

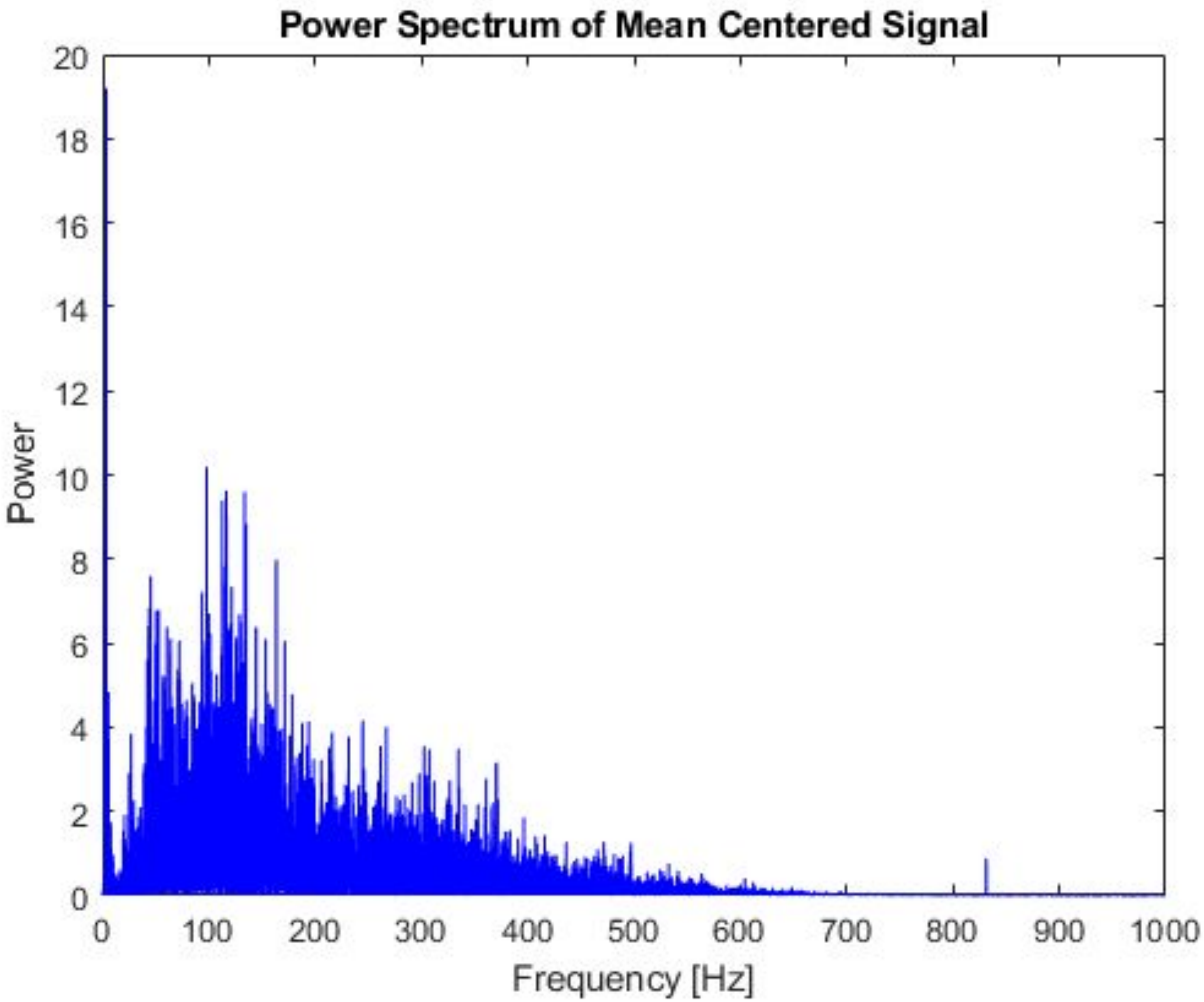
- Most important to establish in this step are the frequencies of the signal
  - Sampling frequency → `freq_samp`
  - Nyquist frequency → `freq_samp/2` → `freq_Nyquist`



Channel 1
voltage [V]
time [seconds]
8
6
4
2
0
-2
-4
-6
-8
0
2
4
6
8
10
12
14
16
18
20

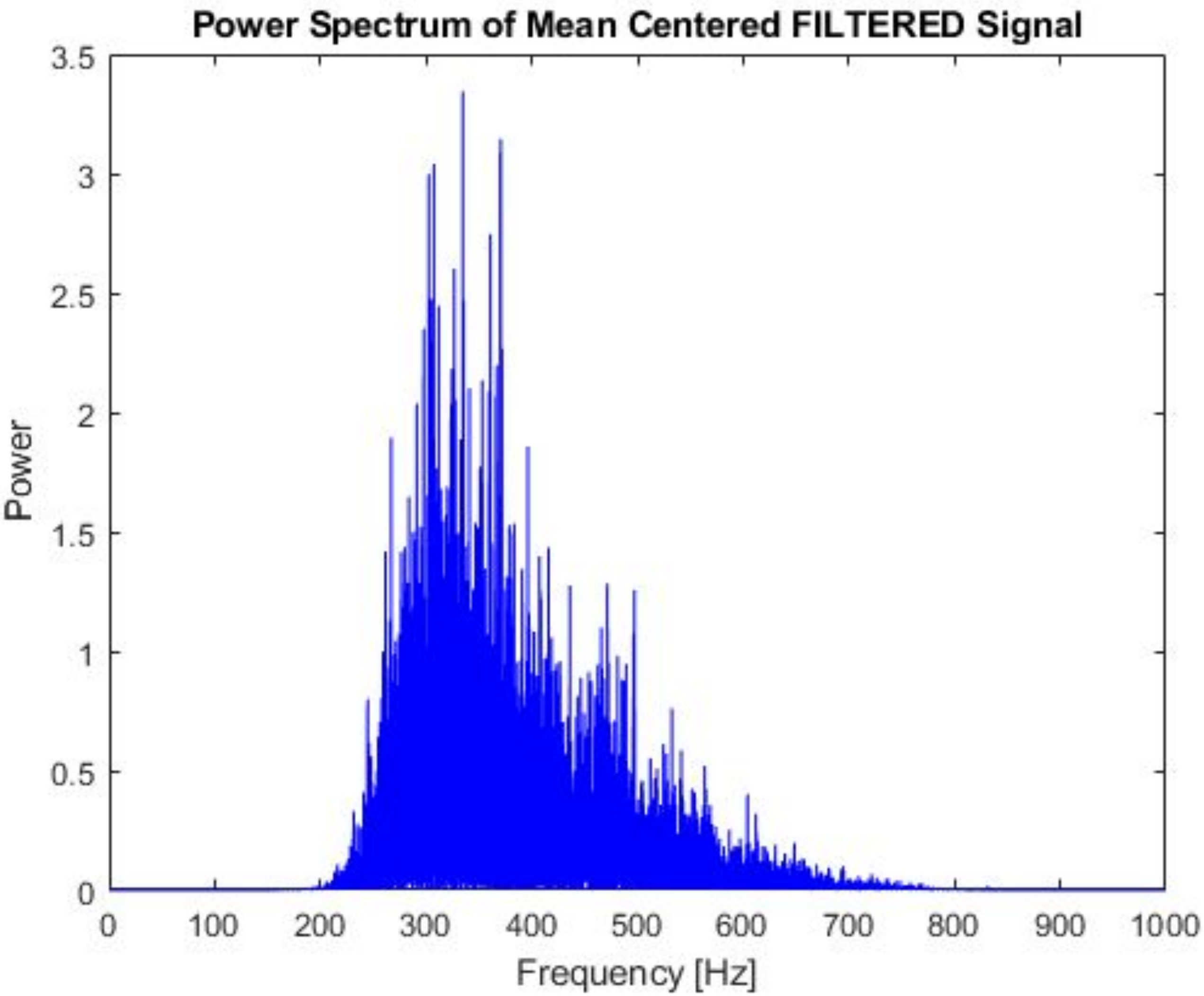
Filtering the Signal by Removing Frequencies

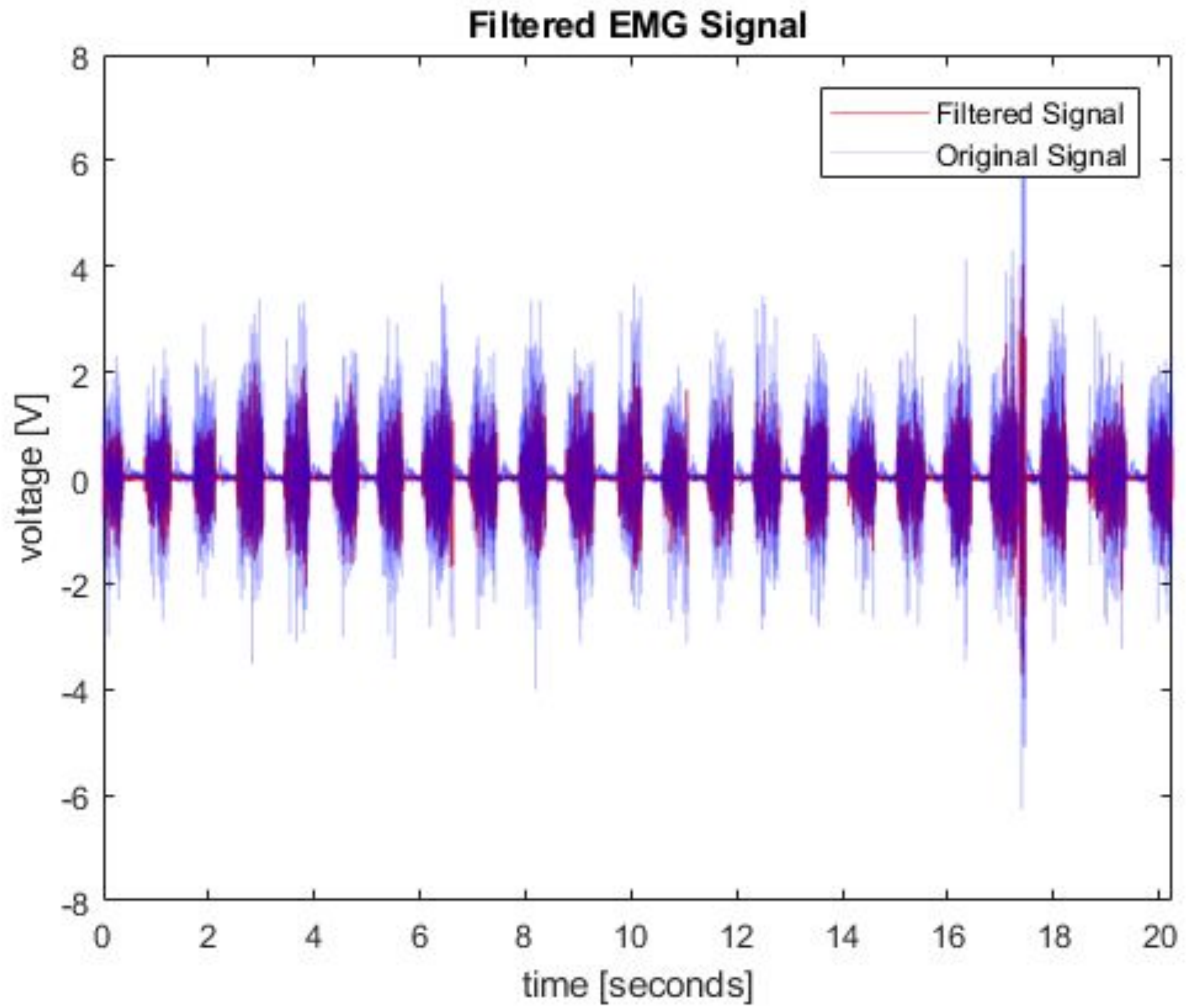
- Perform a Fourier Analysis
  - Requires that we shift our signal by its mean
  - Take the FFT (Fast Fourier Transform) of the signal to identify the frequency components
  - Evaluate the *power* of the signal
- Problems
  - Our signal is nonstationary (the underlying distribution and the parameters of the signal changes with time)
  - Highly nonlinear signal
  - Better to have used empirical mode decomposition, Hilbert Huang Transform, or wavelet decomposition


Power Spectrum of Mean Centered Signal
Power
Frequency [Hz]
0
2
4
6
8
10
12
14
16
18
20
0
100
200
300
400
500
600
700
800
900
1000

Filtering the Signal by Removing Frequencies

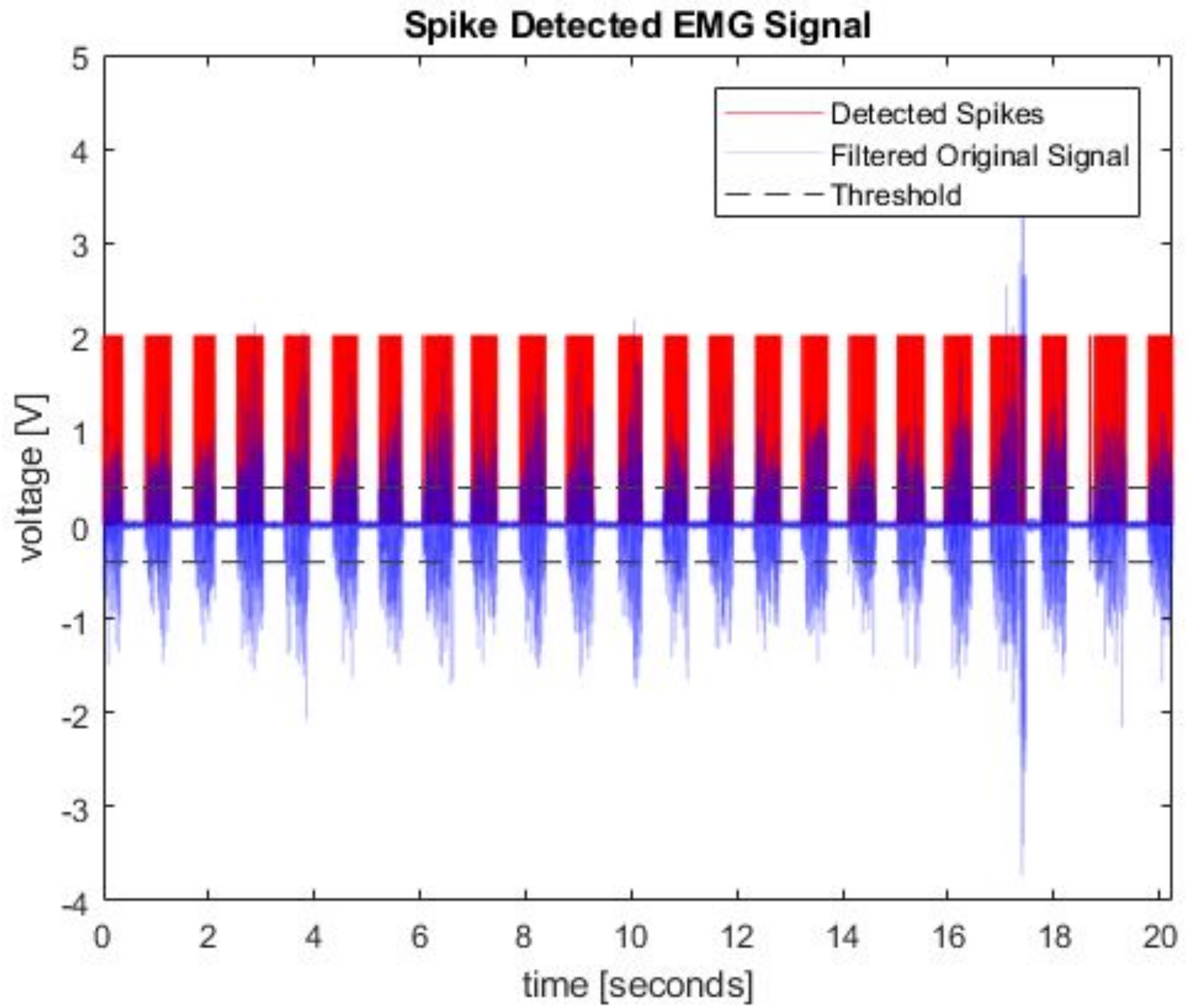
- Use Butterworth *band*pass filter (more of a *low*pass filter)
  - `freq_lowerCutOff = 250; % [Hz]`
  - `freq_upperCutOff = 800; % [Hz]`
  - `[b,a] =`
    - `butter(4, [freq_lowerCutOff/(freq_Nyquist),freq_upperCutOff/(freq_Nyquist)], 'bandpass');`
  - `filt_sig = filtfilt(b, a, test_input);`
- We can evaluate the performance by recalculating the FFT
- We can also compare it to the original signal
  - We see that the signal loses some amplitude after removing some of the noisy lower frequencies
- More could be done if signal was sampled at a higher rate


Power Spectrum of Mean Centered FILTERED Signal
Power
Frequency [Hz]
0
0.5
1
1.5
2
2.5
3
3.5
0
100
200
300
400
500
600
700
800
900
1000

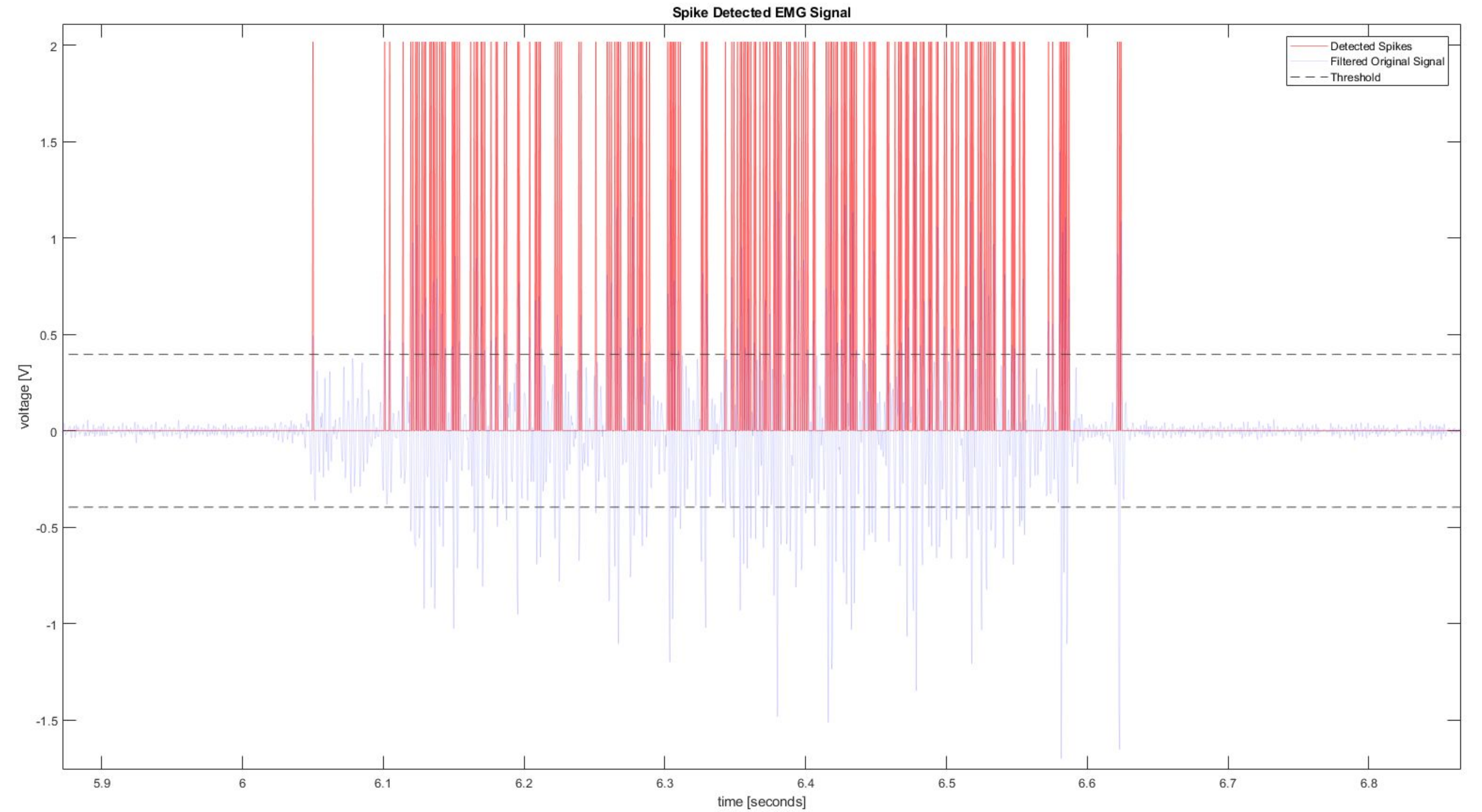

Filtered EMG Signal
Filtered Signal
Original Signal
voltage [V]
time [seconds]
8
6
4
2
0
-2
-4
-6
-8
0
2
4
6
8
10
12
14
16
18
20

Detecting Spikes

- `% Quiroga et. al threshold`
  - `std_dev_estimate = median(abs(filt_sig)/0.6745);`
  - `MPH_Thr = 6*std_dev_estimate;`
  - `MPD = 5e-3; % width of a usual pulse or the minimum peak spread [seconds]`
- We can define an estimate of the noise based on the median absolute amplitude of the filtered signal
- Then use `findpeaks` to find the peaks and the indices of the signal based on the noise estimate
- We can plot the peaks using a binary *Dirac* impulse at the found locations of the signal to contain spikes

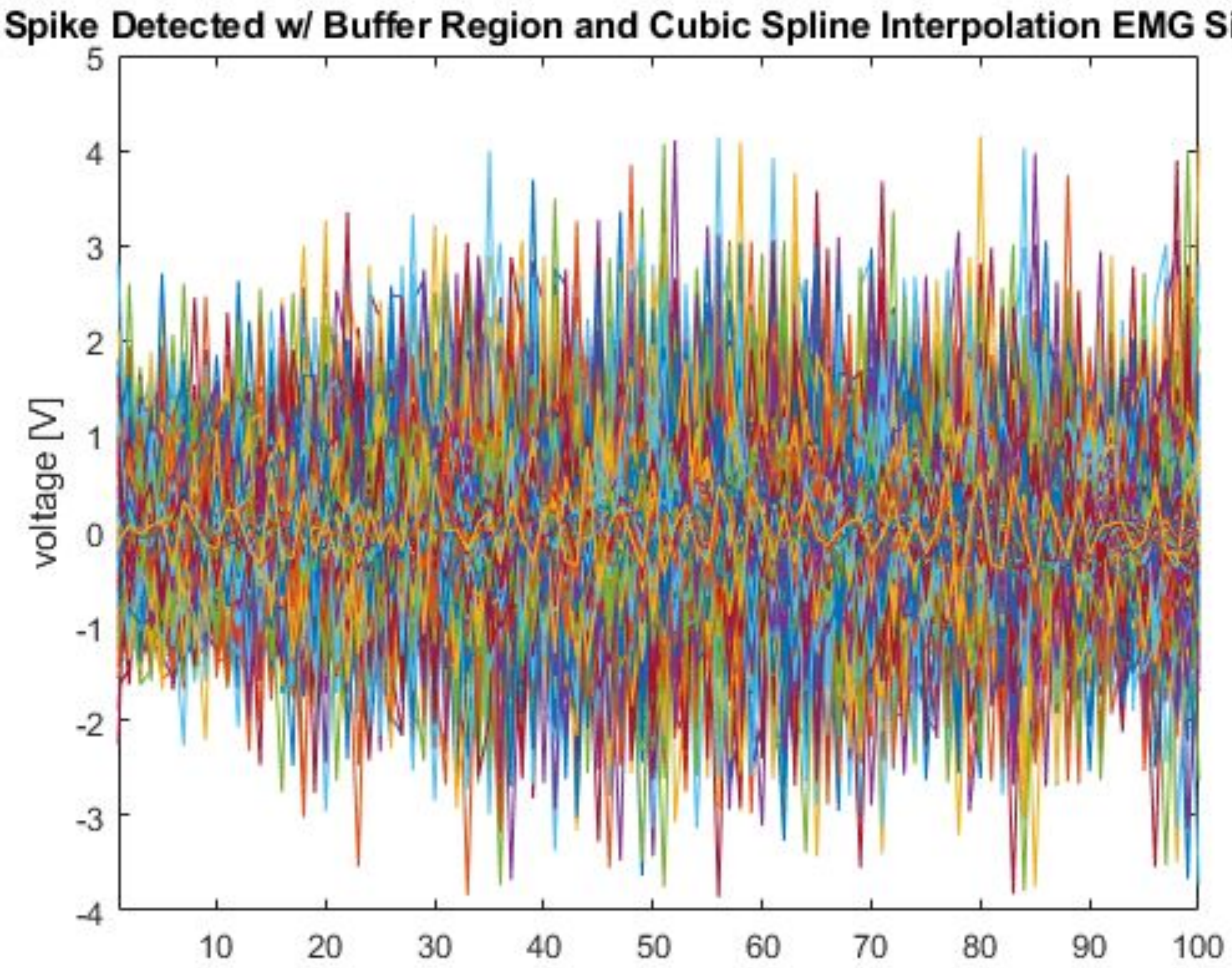


Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
time [seconds]
5
4
3
2
1
0
-1
-2
-3
-4
0
2
4
6
8
10
12
14
16
18
20


Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
time [seconds]
2
1.5
1
0.5
0
-0.5
-1
-1.5
5.9
6
6.1
6.2
6.3
6.4
6.5
6.6
6.7
6.8

Aligning Spikes

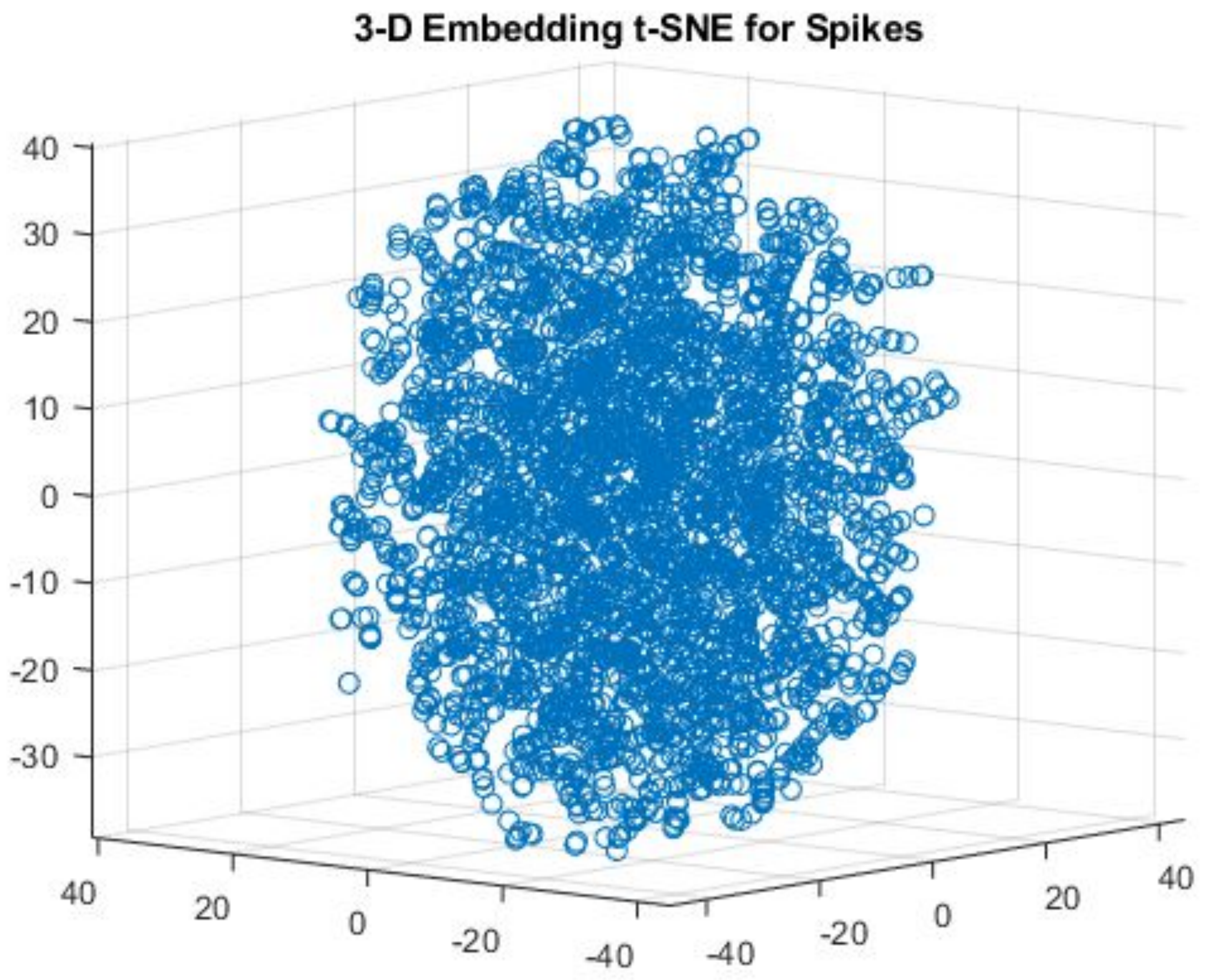
- Need a buffer region around the peaks to obtain the area of the spikes
- Can use *cubic spline interpolation*
  - We want 100 points around the peak to determine the “spike”
  - Sample originally 400 points around the peak
  - Align by the max peak in that spike region
  - Take the cubic spline interpolation of 100 points in the same region
- Problems
  - Issues arise when peaks are at the beginning or the end of the signal due to large window sizes
  - Peaks are not always aligned by max amplitude since sampling more points might alias higher amplitude points

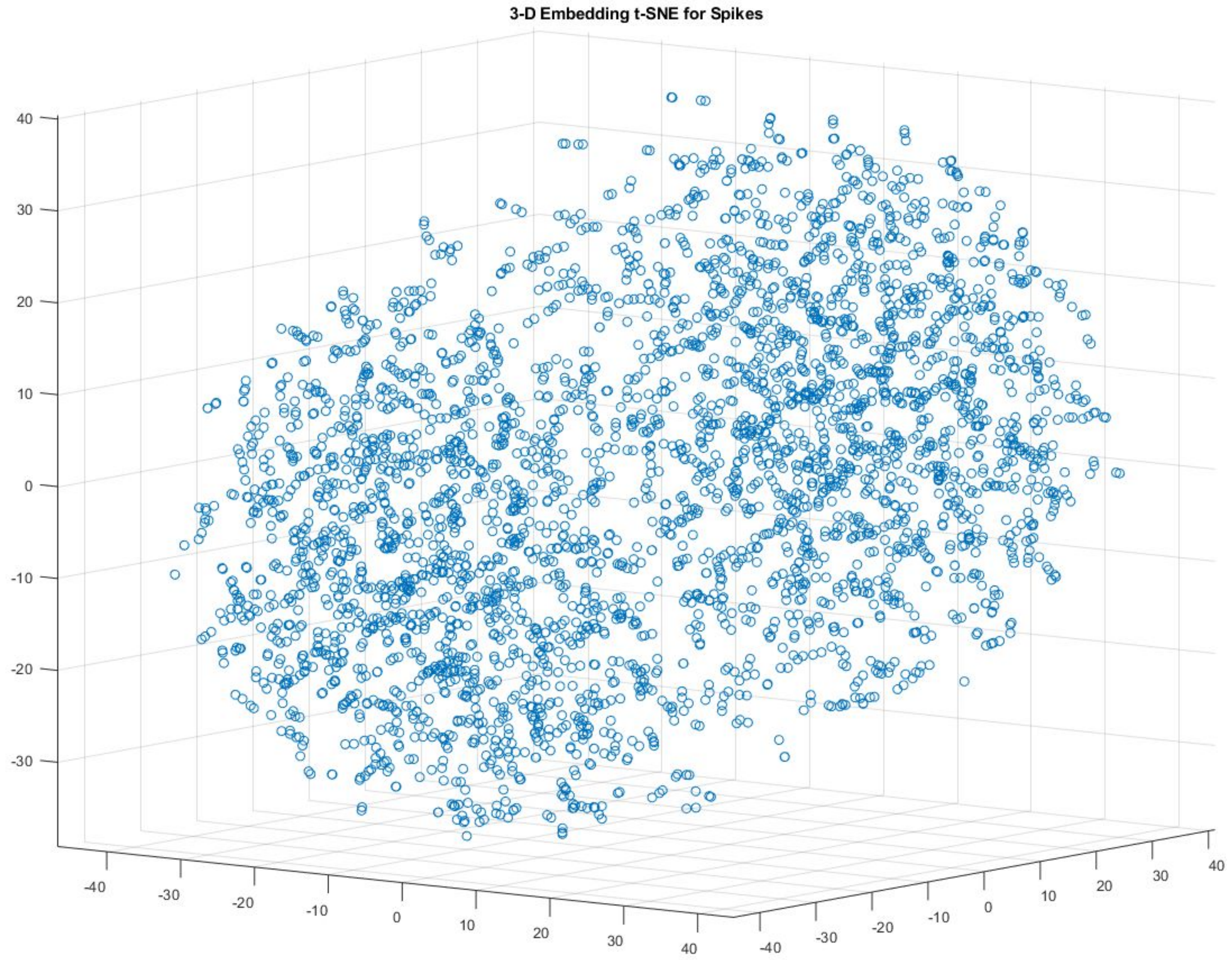


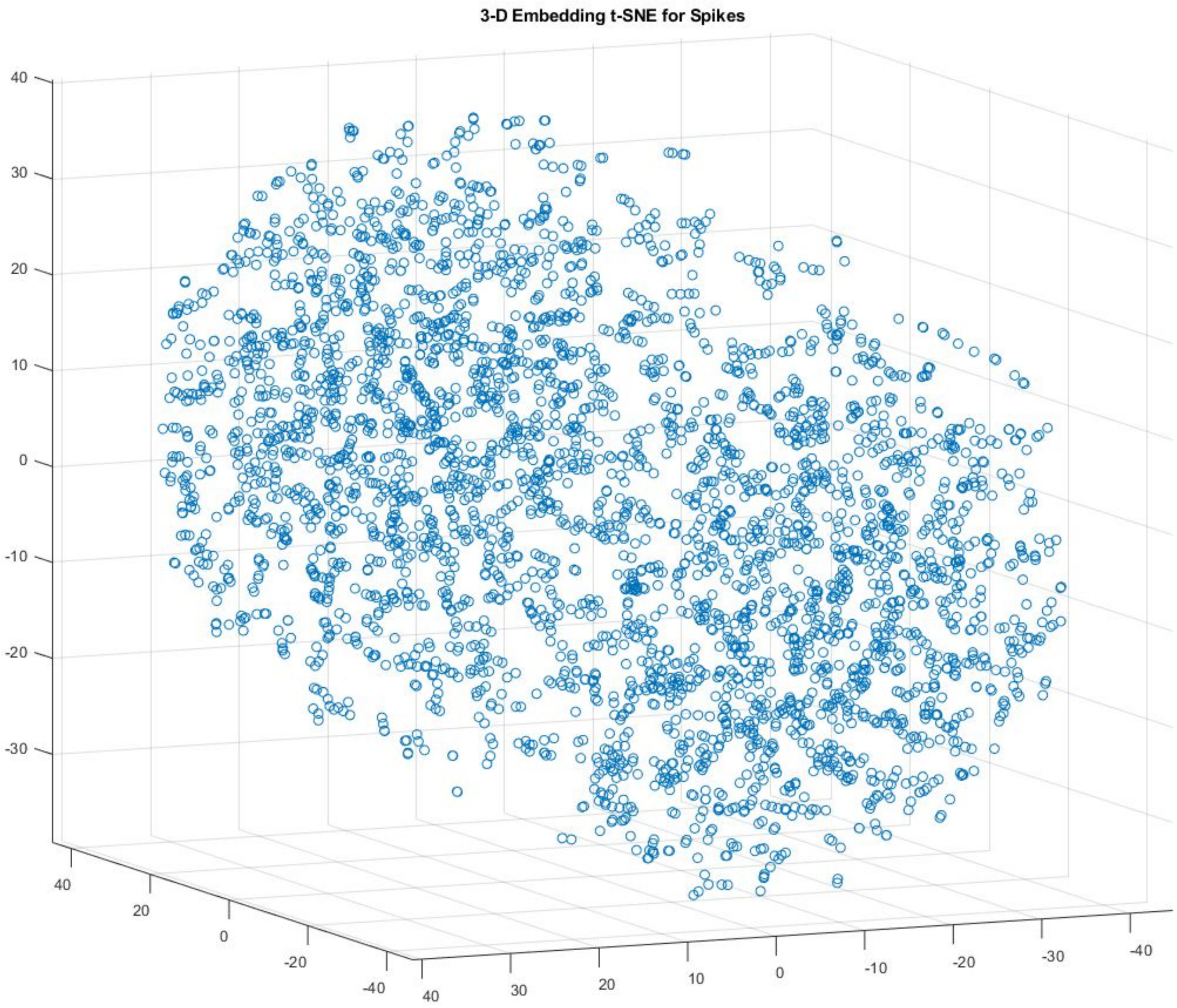
Spike Detected w/ Buffer Region and Cubic Spline Interpolation EMG Signal
voltage [V]
5
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100

t-Distributed Stochastic Neighbor Embedding

- PCA requires a number of components to keep in the model after selection running
- t-SNE is an unsupervised dimensionality reduction technique and it helps visualize high dimensional datasets
    - In particular, the space it chooses has little meaning
    - However, if two objects are close to each other (in terms of a distance metric, in this case *cosine* distance, or the *correlation*) then they are similar
    - Objects that are dissimilar will be farther apart
    - Axes have no meaning
    - We choose to embed the collection in a 3D space and present multiple views, the final view suggests that bifurcating the dataset is possible (by visual inspection)

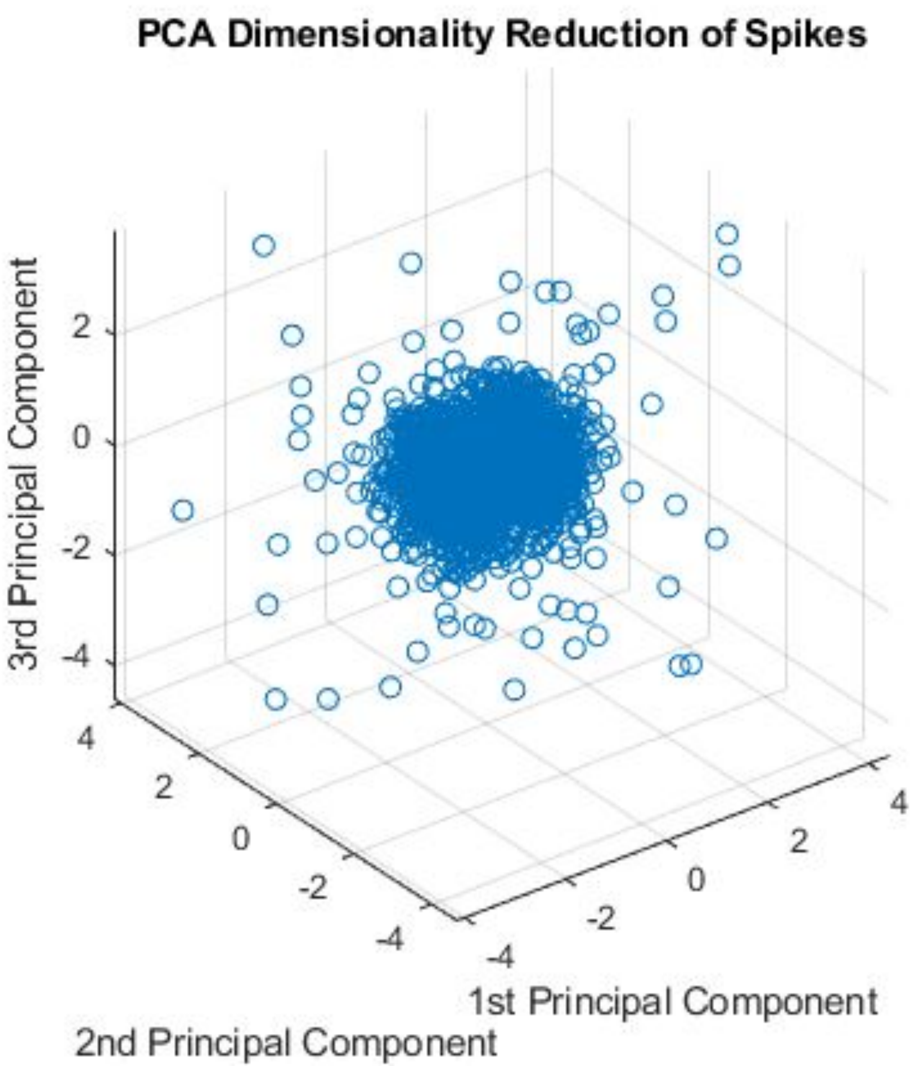

3-D Embedding t-SNE for Spikes
40
30
20
10
0
-10
-20
-30
40
20
0
-20
-40
-40
-20
0
20
40

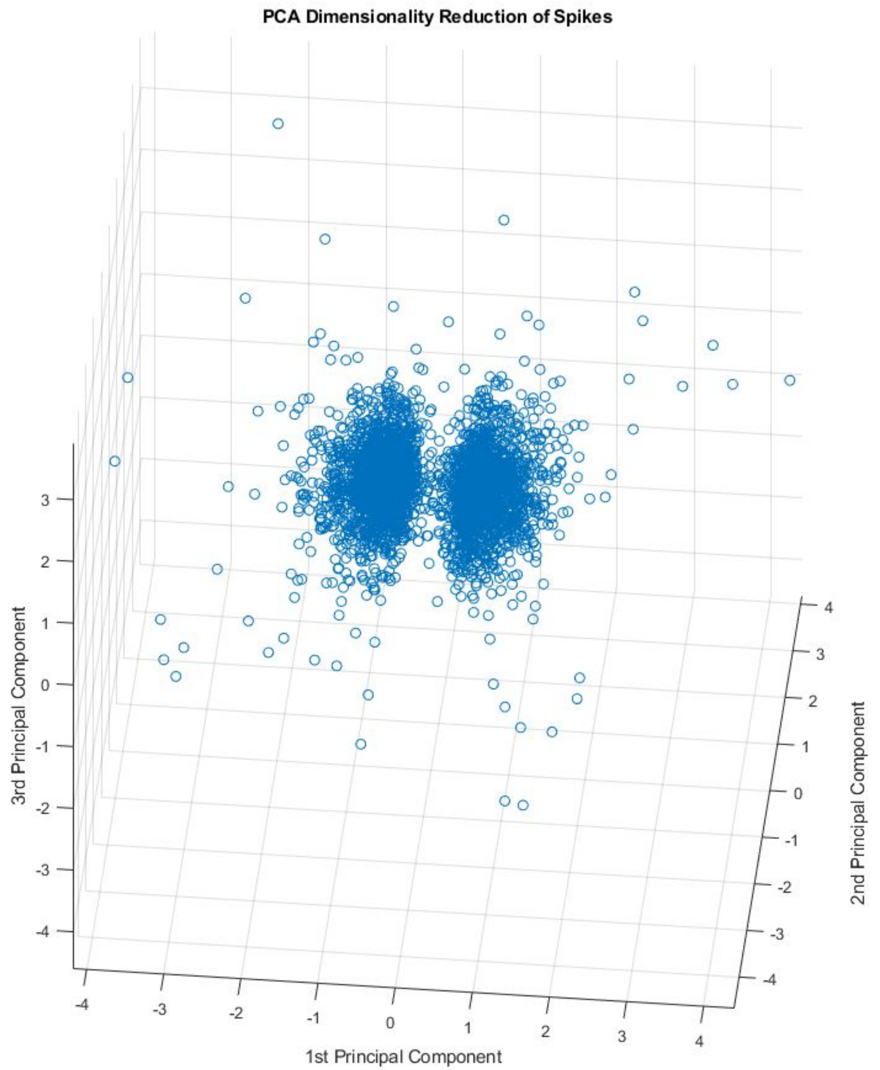

3-D Embedding t-SNE for Spikes
40
30
20
10
0
-10
-20
-30
-40
-30
-20
-10
0
10
20
30
40
-40
-30
-20
-10
0
10
20
30
40


3-D Embedding t-SNE for Spikes
40
30
20
10
0
-10
-20
-30
40
20
0
-20
-40
40
30
20
10
0
-10
-20
-30
-40

PCA & Feature Extraction

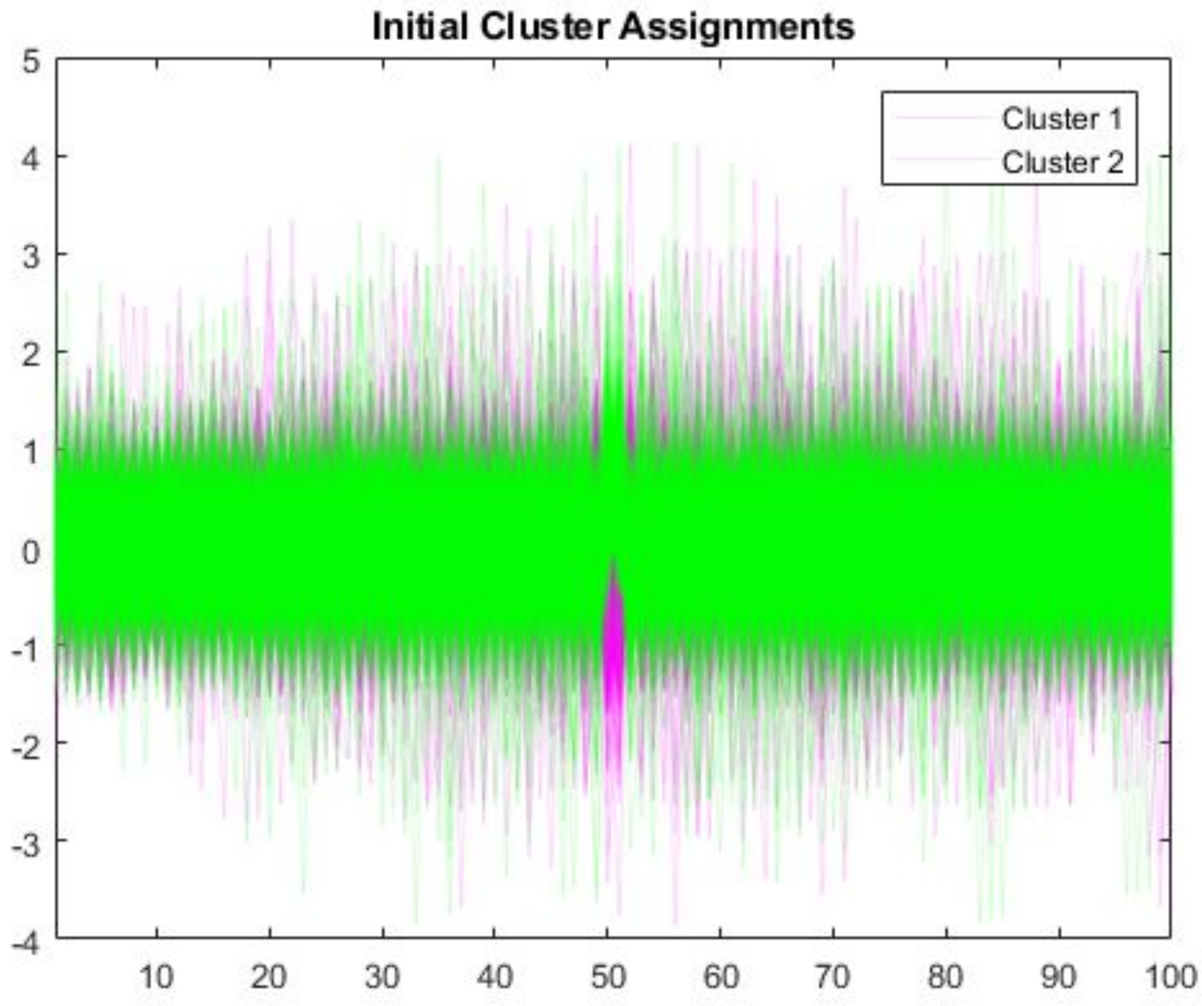
- Knowing that two clusters can possible be easily *shattered* (in terms of VC dimension) gives me confidence that the noisy dataset can be reduced
  - `[coeff,score,latent,~,explained] = pca(detected_spikes_holder);`
  - `pca_score_holder = score(:,1:3);`
  - We present multiple views to highlight the similarity between t-SNE and PCA
- Percent variance explained by the first three components is not high.
  - At best, the first three principal components explain about 10% of the variance
  - This is worrying and remarkably low


PCA Dimensionality Reduction of Spikes
3rd Principal Component
2
0
-2
-4
4
2
0
-2
-4
4
2
0
-2
-4
1st Principal Component
2nd Principal Component

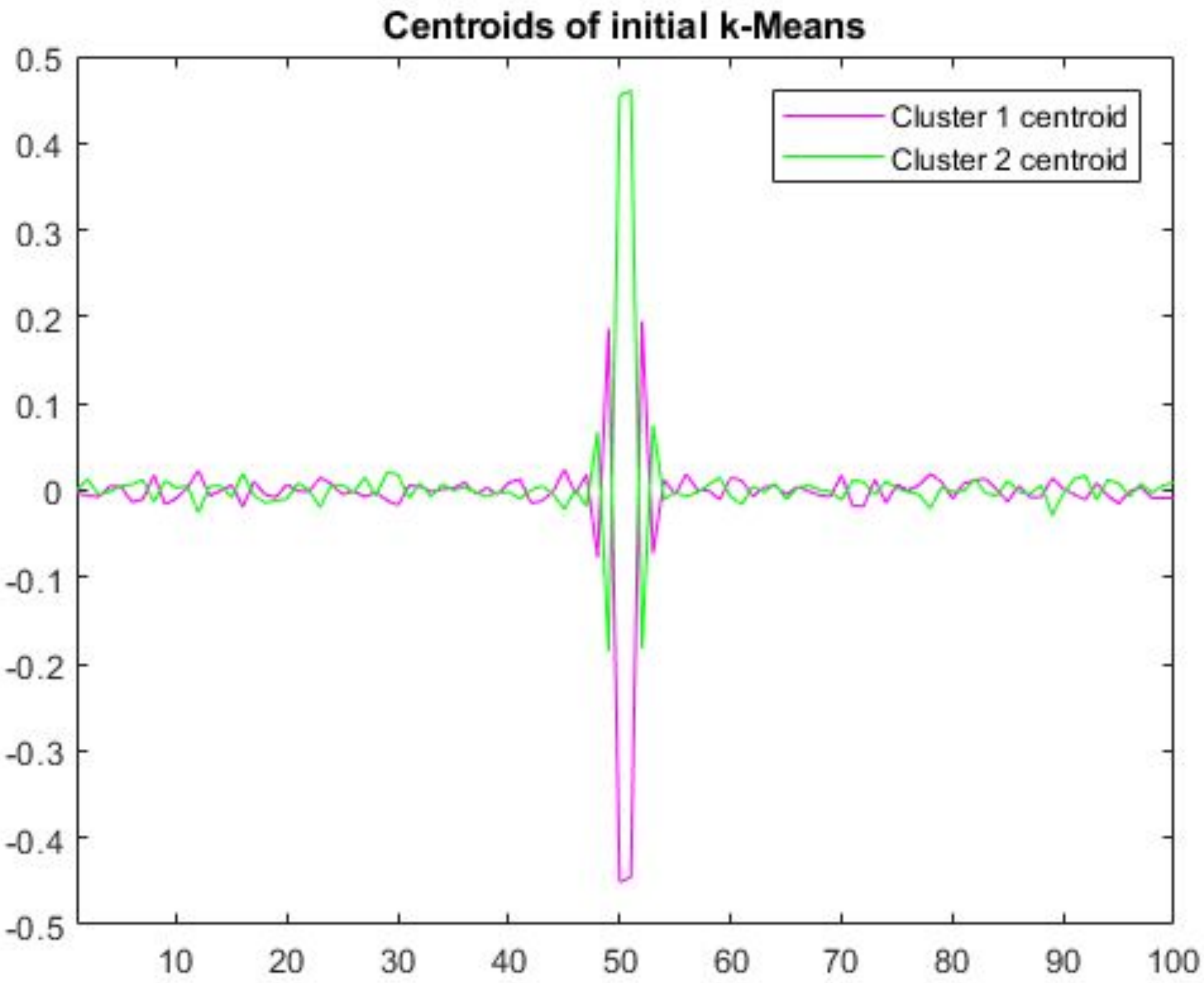

PCA Dimensionality Reduction of Spikes
1st Principal Component
2nd Principal Component
3rd Principal Component
-4
-3
-2
-1
0
1
2
3
4

Cluster Spikes: k-Means Clustering

- We can cluster our PCA reduced dataset into two obvious classes.
- Doing so will remove even more noise, and then we can again cluster on one of the remaining datasets.
- Instead of plotting the PC reduced dataset clustered, I replot the spikes clustered by their respective class
- Doing so, and plotting their means tells us about the nature of the spikes stored in the class
- We believe that the two clusters represent spikes that were -,+,-, in magnitude for one, and +,-,+ for the other

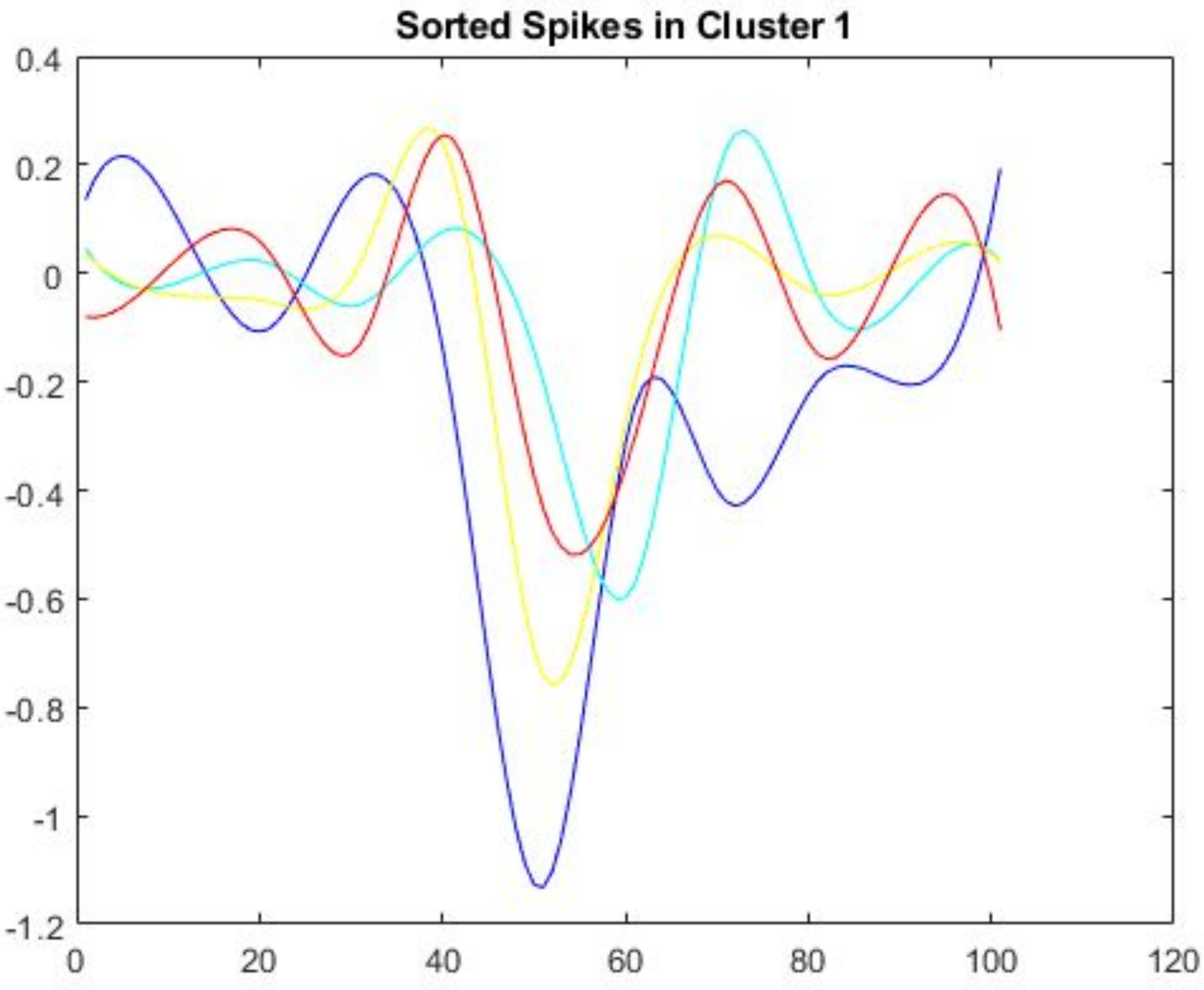


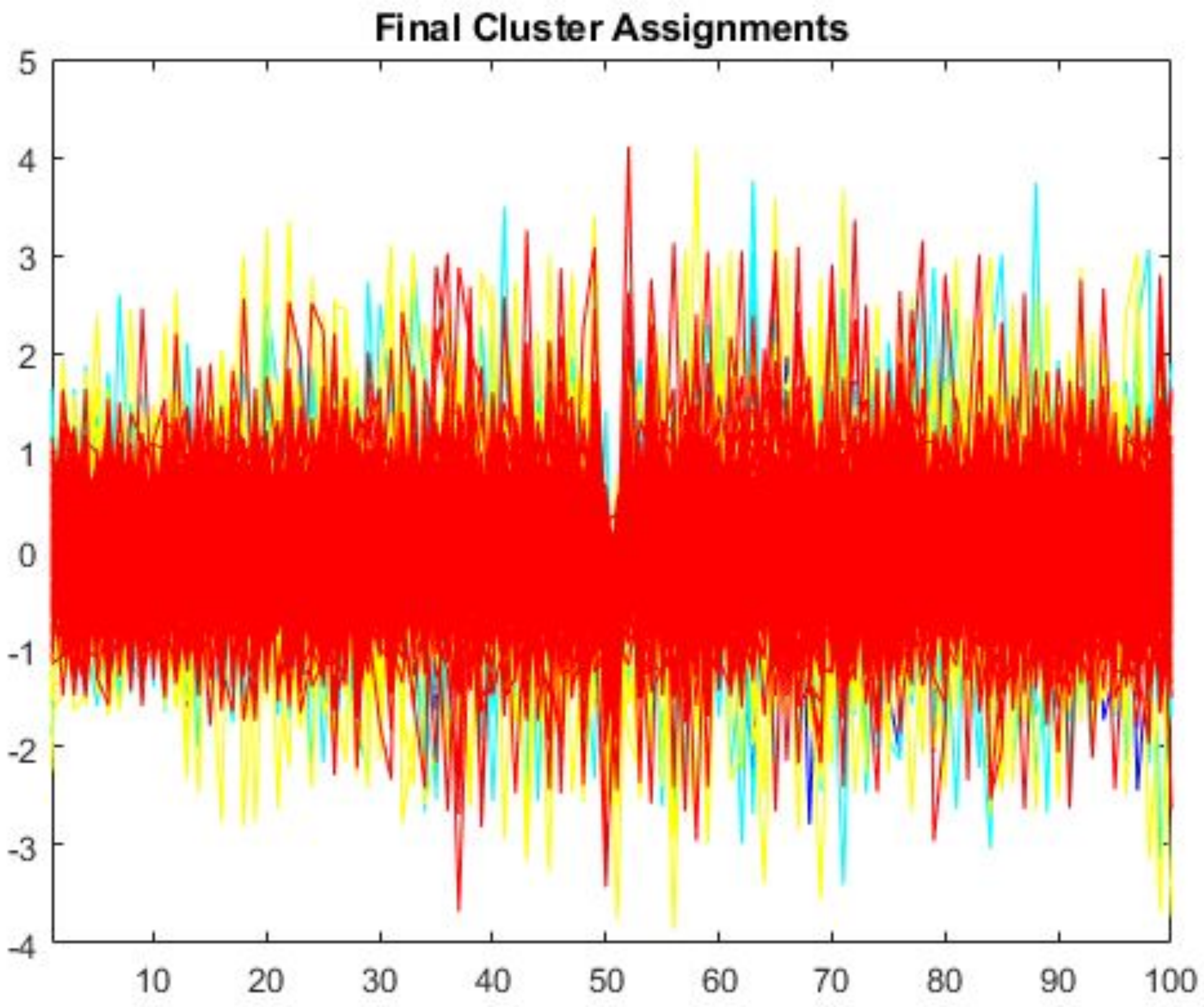
Initial Cluster Assignments
Cluster 1
Cluster 2
5
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100

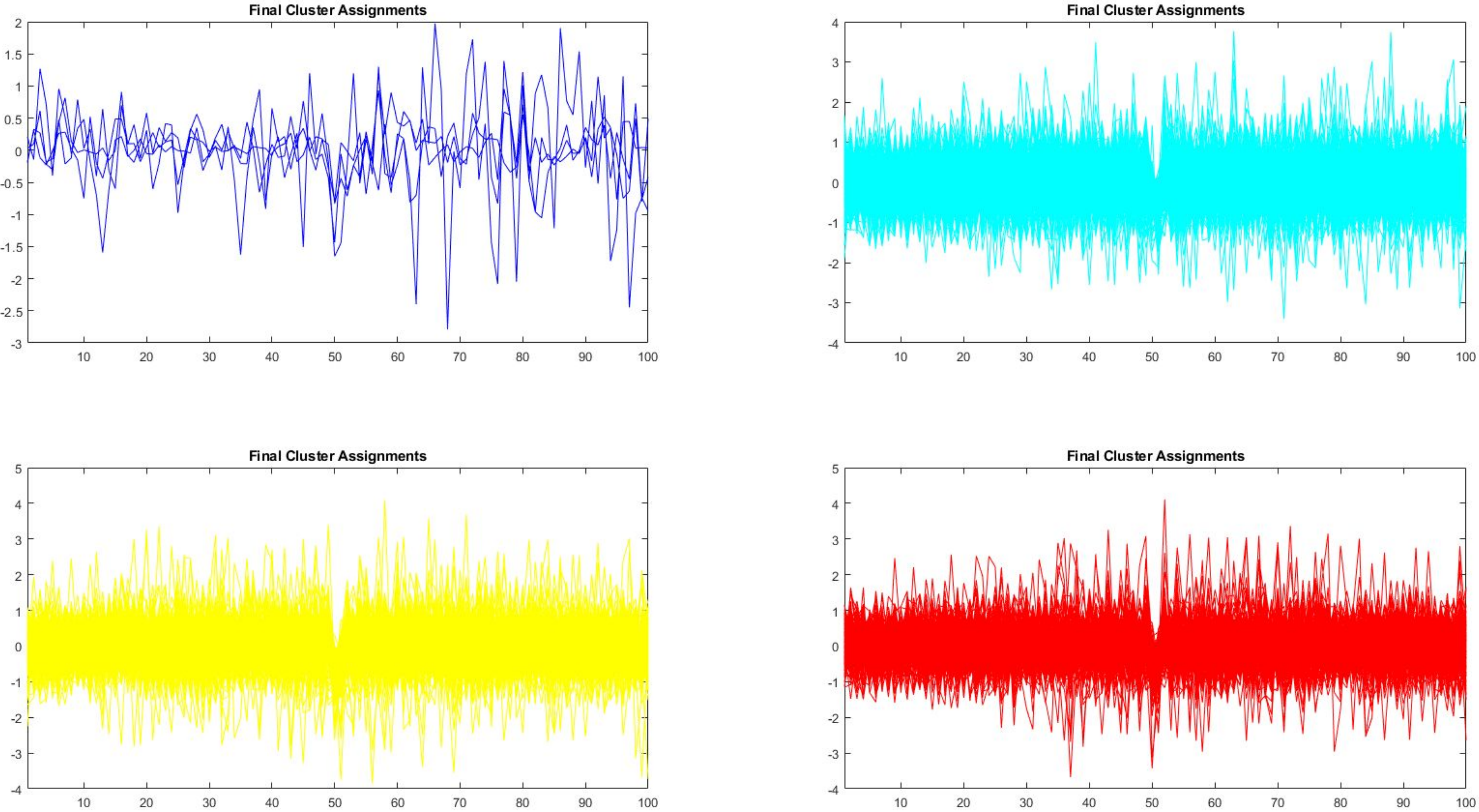

Centroids of initial k-Means
Cluster 1 centroid
Cluster 2 centroid
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
10
20
30
40
50
60
70
80
90
100

Classify Spikes & Analysis

- We then classify the spikes in the larger first cluster
- We use the *elbow* method of determining the right number of clusters
  - Cannot use t-SNE here since t-SNE cannot separate or shatter this space that well
  - Elbow method works similarly to checking whether or not the percent variance explained is greater than some cut off
  - The percent variance increases as the number of clusters increases
  - By doing so we can identify individual spike shapes
  - To increase the robustness of spike identification due to the low sampling rate, we also employ cubic spline interpolation, which gives rise to the smooth spike shapes

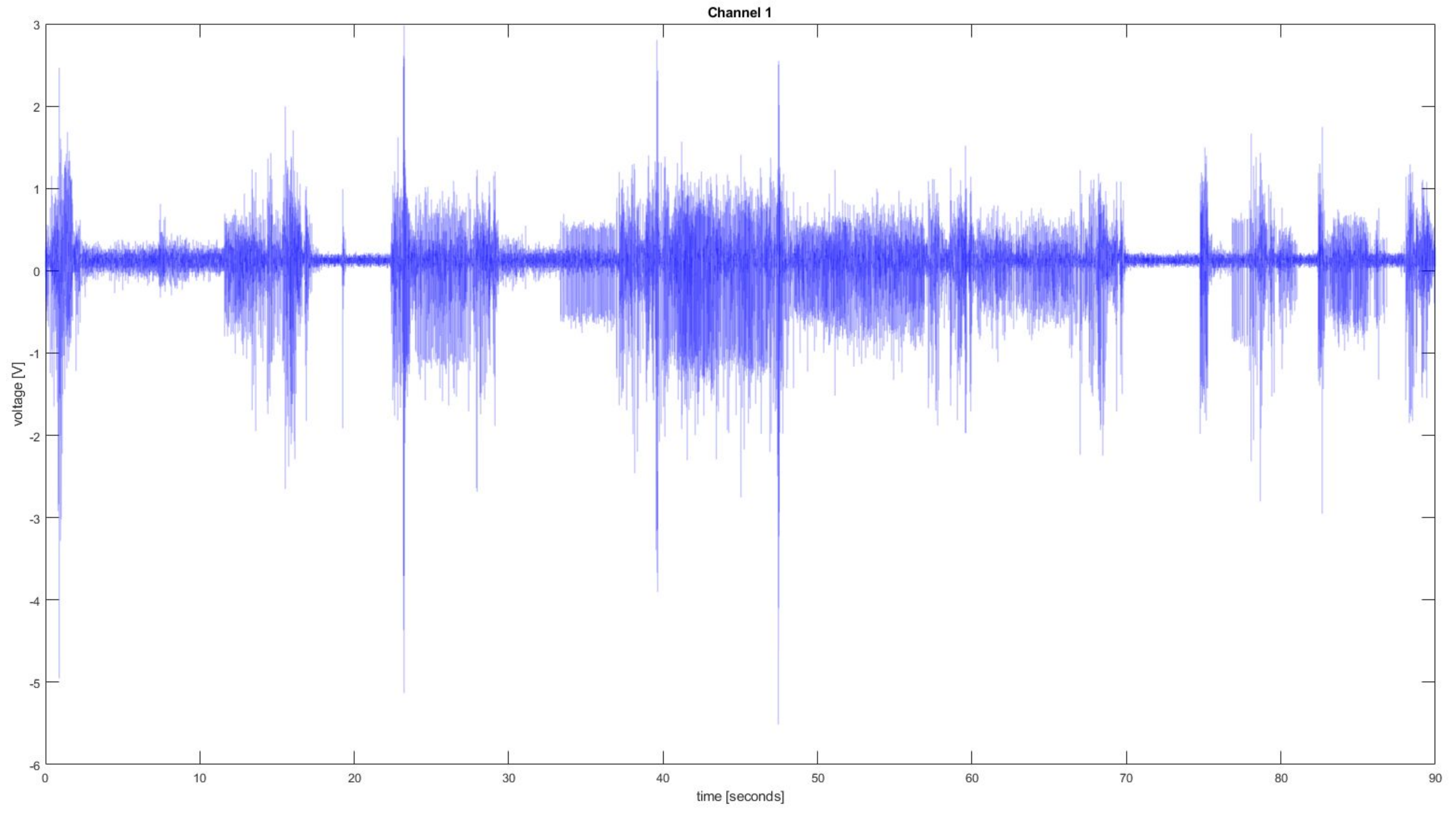

Sorted Spikes in Cluster 1
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
0
20
40
60
80
100
120


Final Cluster Assignments
5
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100

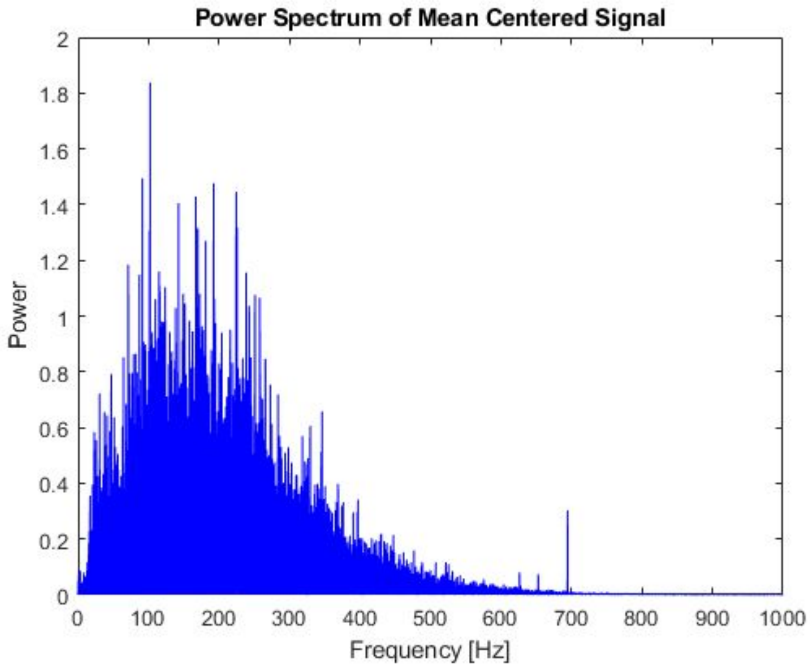


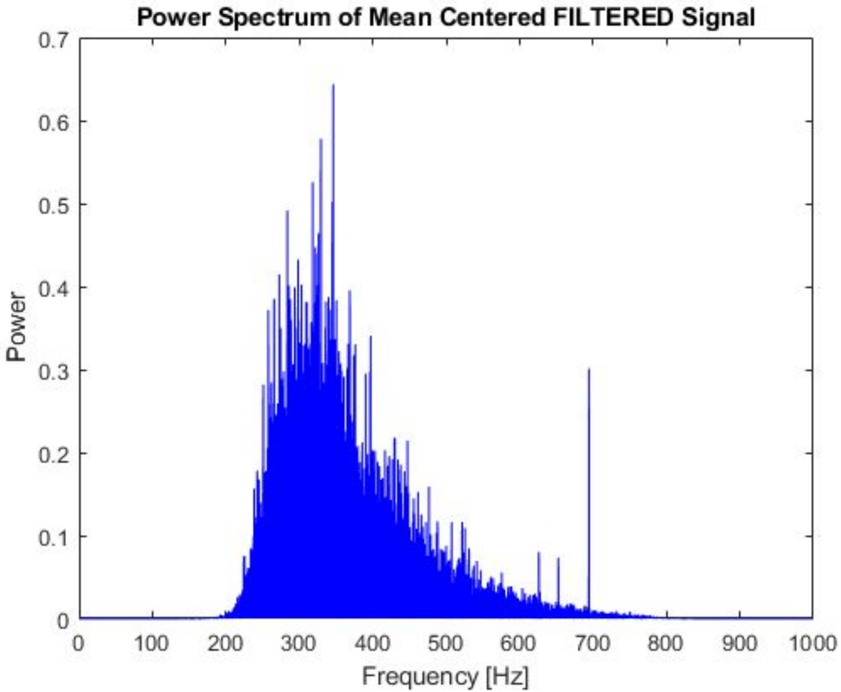
Final Cluster Assignments
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
10
20
30
40
50
60
70
80
90
100
Final Cluster Assignments
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100
Final Cluster Assignments
5
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100
Final Cluster Assignments
5
4
3
2
1
0
-1
-2
-3
-4
10
20
30
40
50
60
70
80
90
100

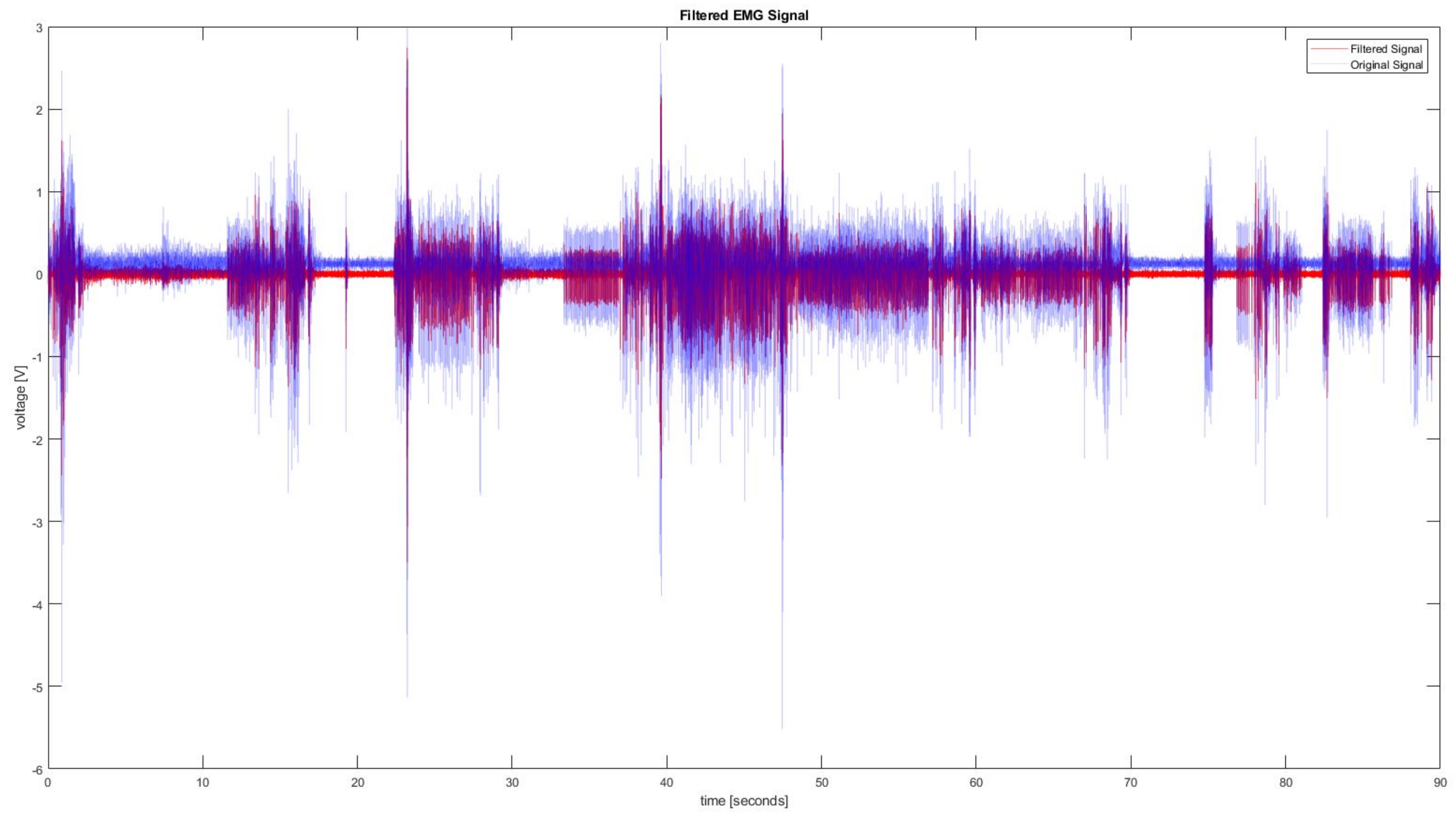
Results on `EMG_example_1_90s_fs_2k.csv`



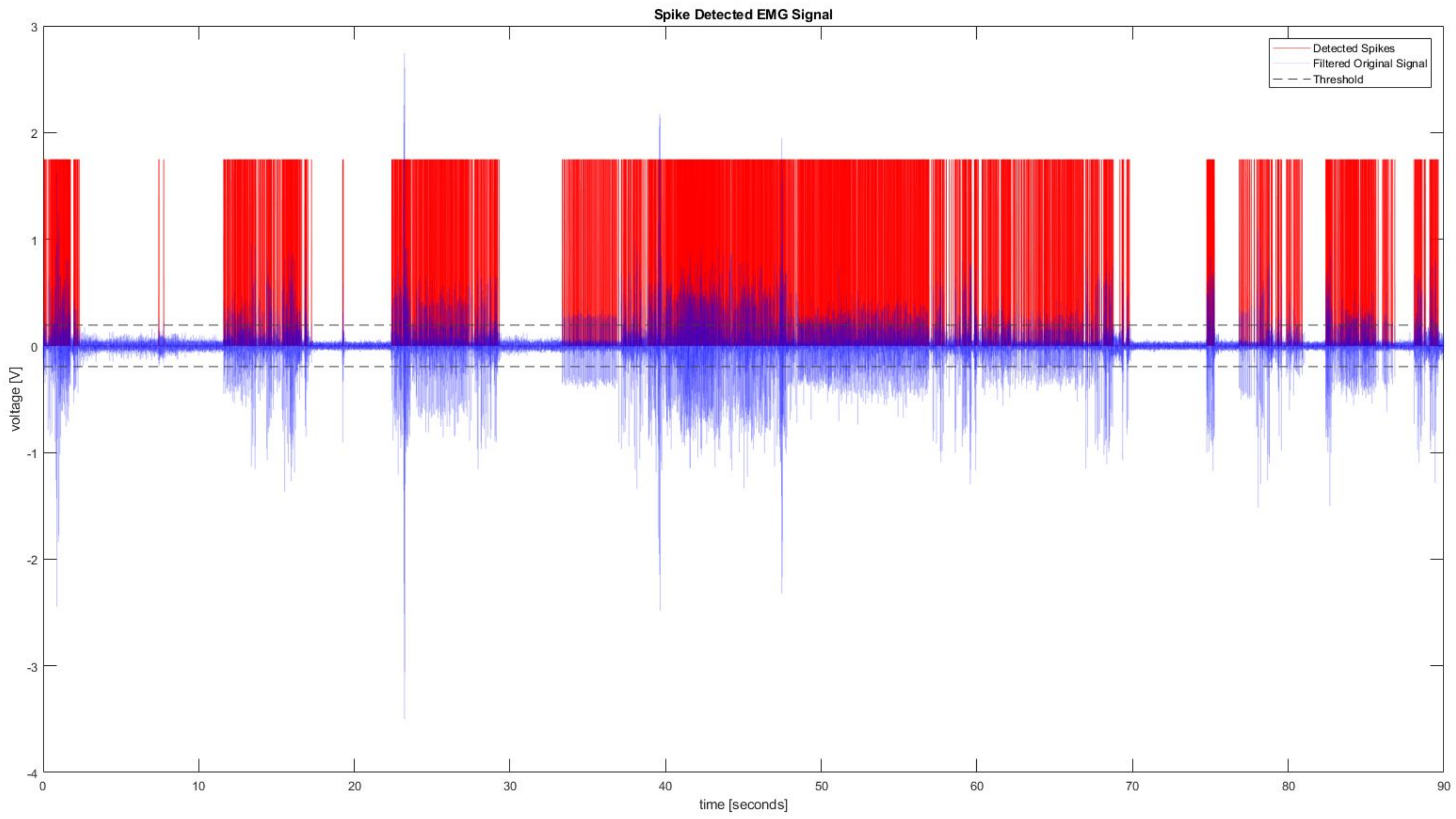
Channel 1
voltage [V]
time [seconds]
3
2
1
0
-1
-2
-3
-4
-5
-6
0
10
20
30
40
50
60
70
80
90


Power Spectrum of Mean Centered Signal
Power
Frequency [Hz]
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
100
200
300
400
500
600
700
800
900
1000

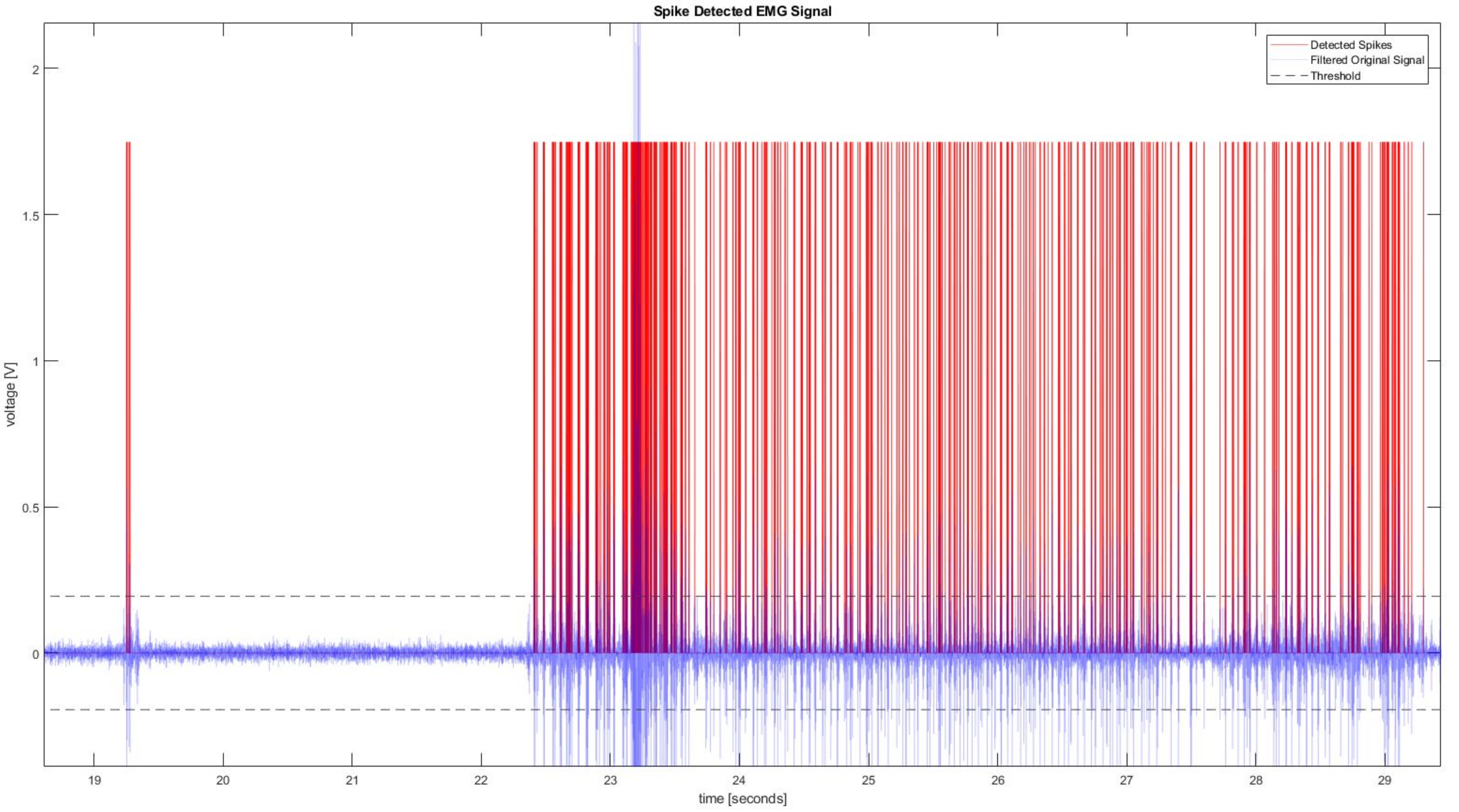

Power Spectrum of Mean Centered FILTERED Signal
Power
Frequency [Hz]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
100
200
300
400
500
600
700
800
900
1000



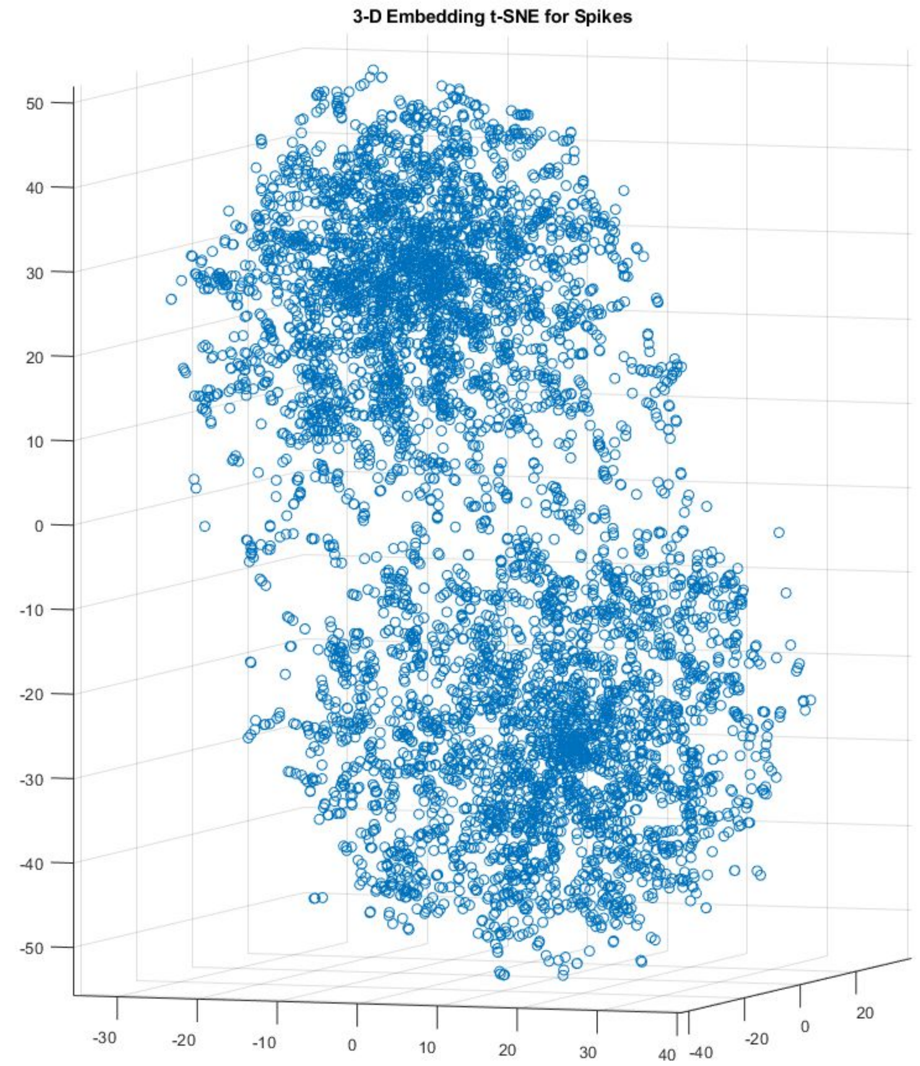
Filtered EMG Signal
Filtered Signal
Original Signal
voltage [V]
time [seconds]
3
2
1
0
-1
-2
-3
-4
-5
-6
0
10
20
30
40
50
60
70
80
90

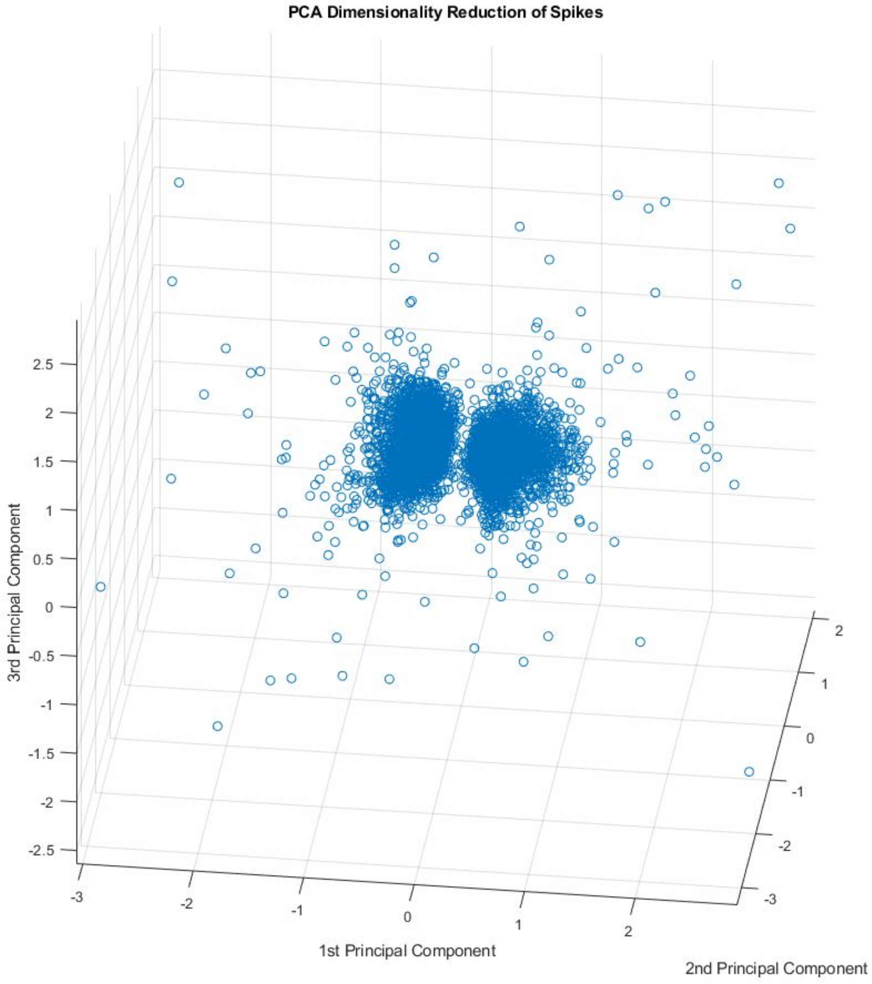


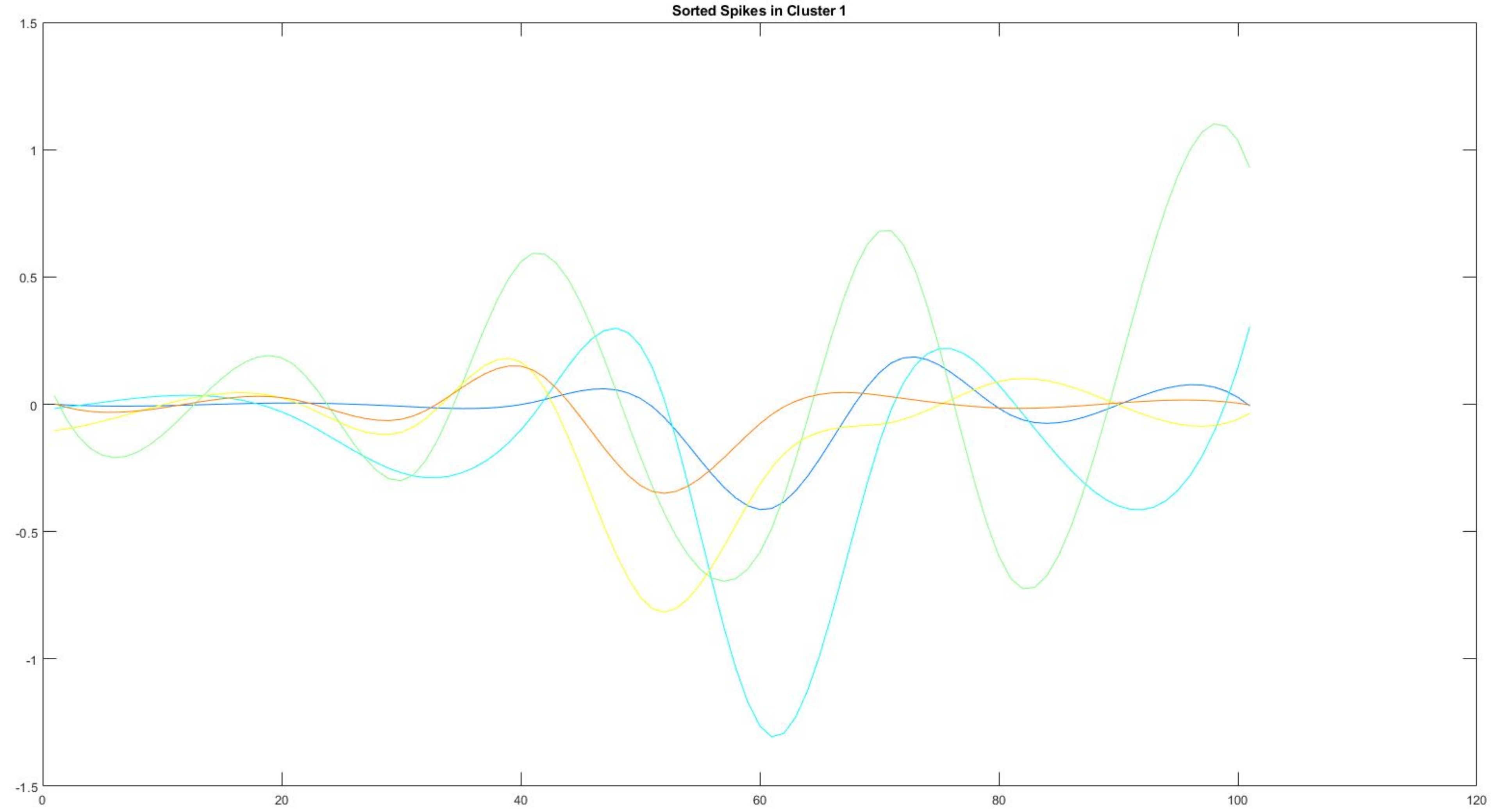
Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
time [seconds]
3
2
1
0
-1
-2
-3
-4
0
10
20
30
40
50
60
70
80
90



Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
2
1.5
1
0.5
0
19
20
21
22
23
24
25
26
27
28
29
time [seconds]

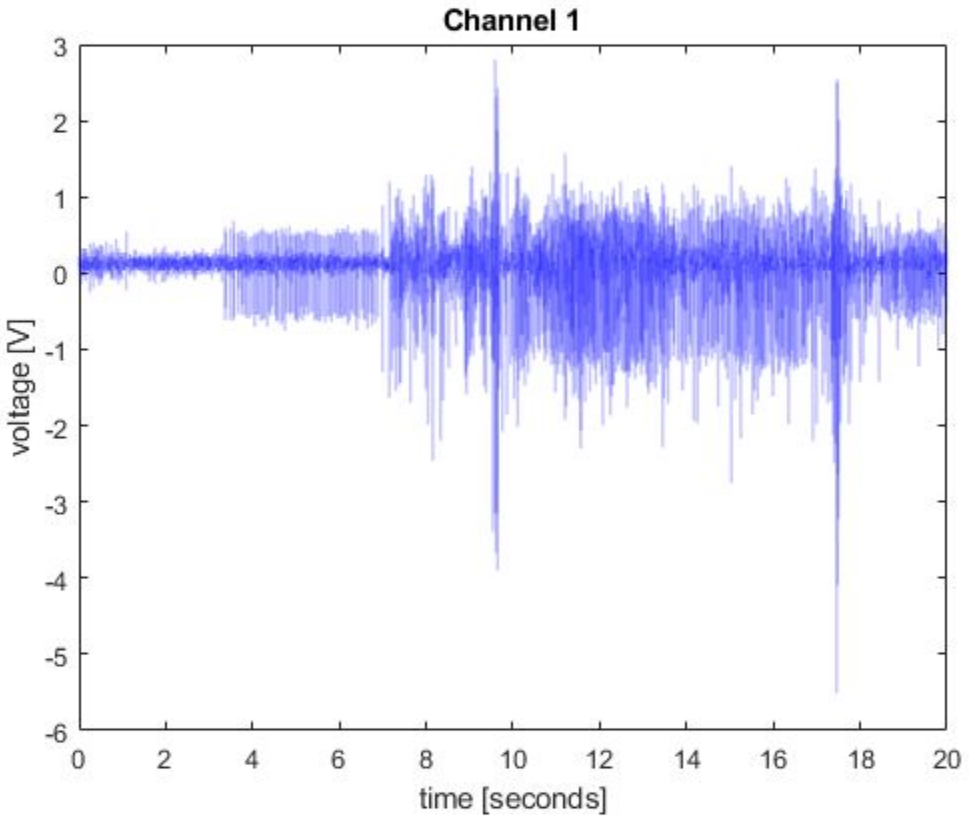

3-D Embedding t-SNE for Spikes

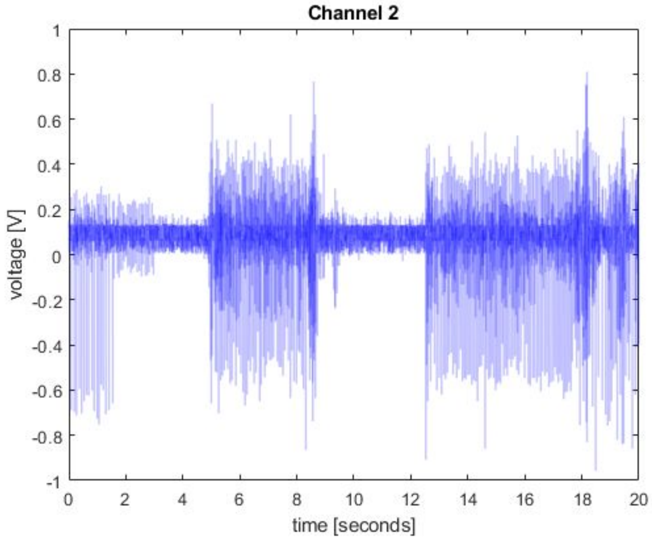

PCA Dimensionality Reduction of Spikes
3rd Principal Component
1st Principal Component
2nd Principal Component

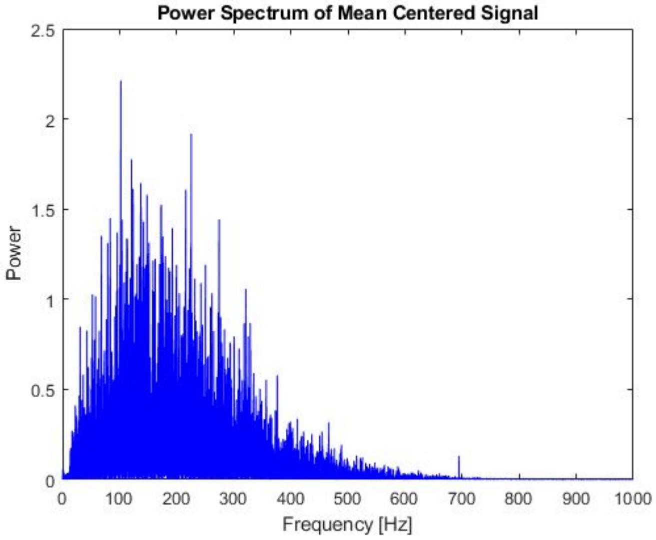


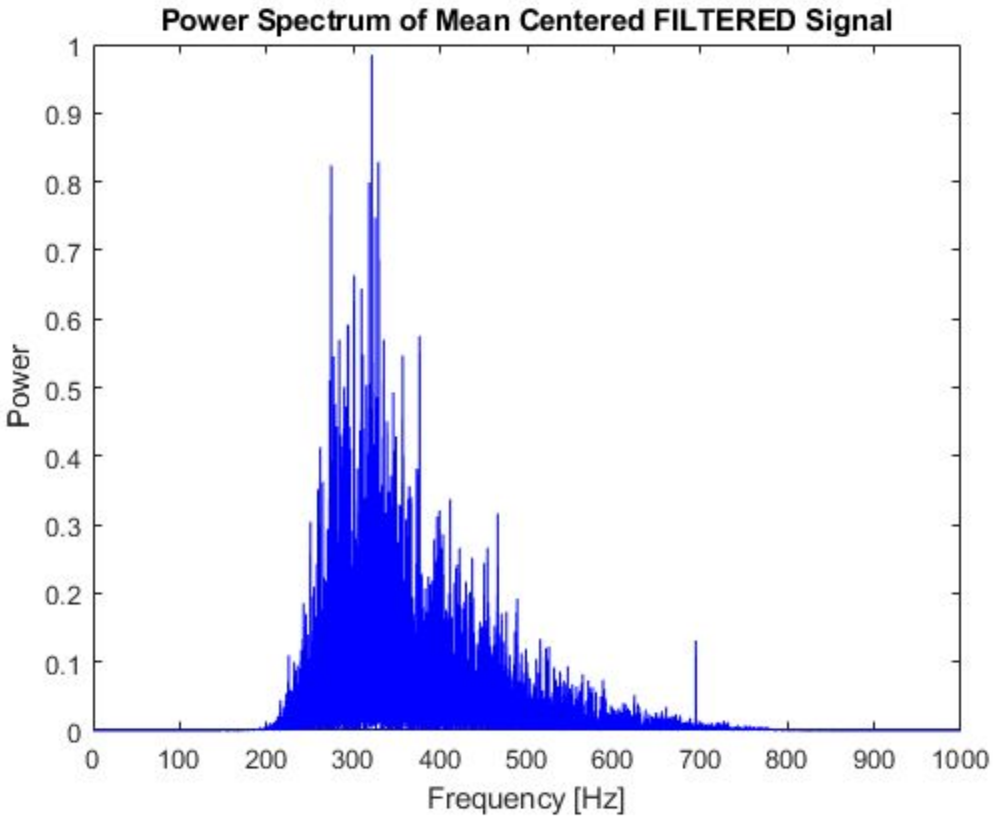
Sorted Spikes in Cluster 1
1.5
1
0.5
0
-0.5
-1
-1.5
0
20
40
60
80
100
120

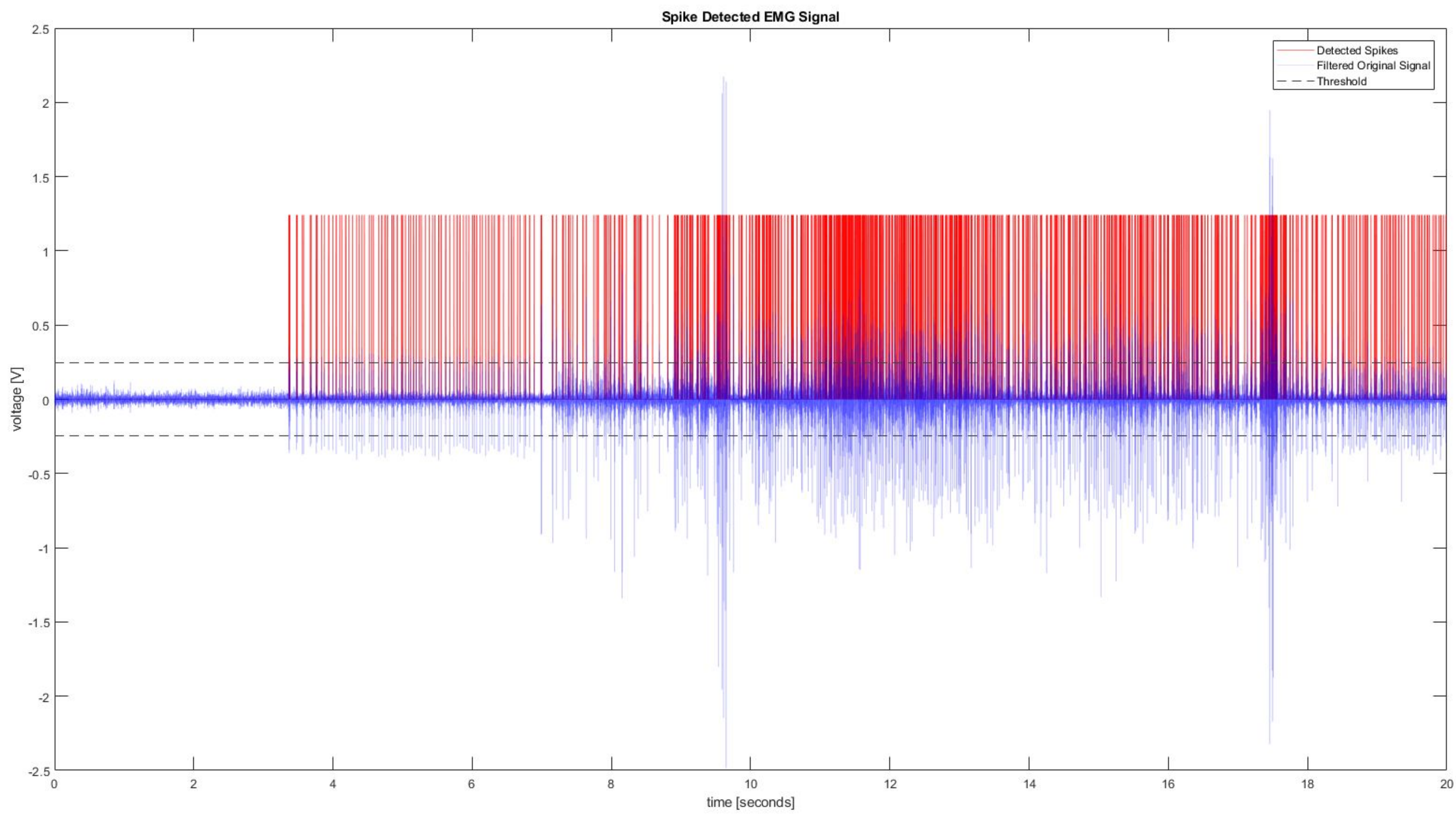
Results on `EMG_example_20s_2000Hz-2016.csv`


Channel 1
voltage [V]
time [seconds]
3
2
1
0
-1
-2
-3
-4
-5
-6
0
2
4
6
8
10
12
14
16
18
20

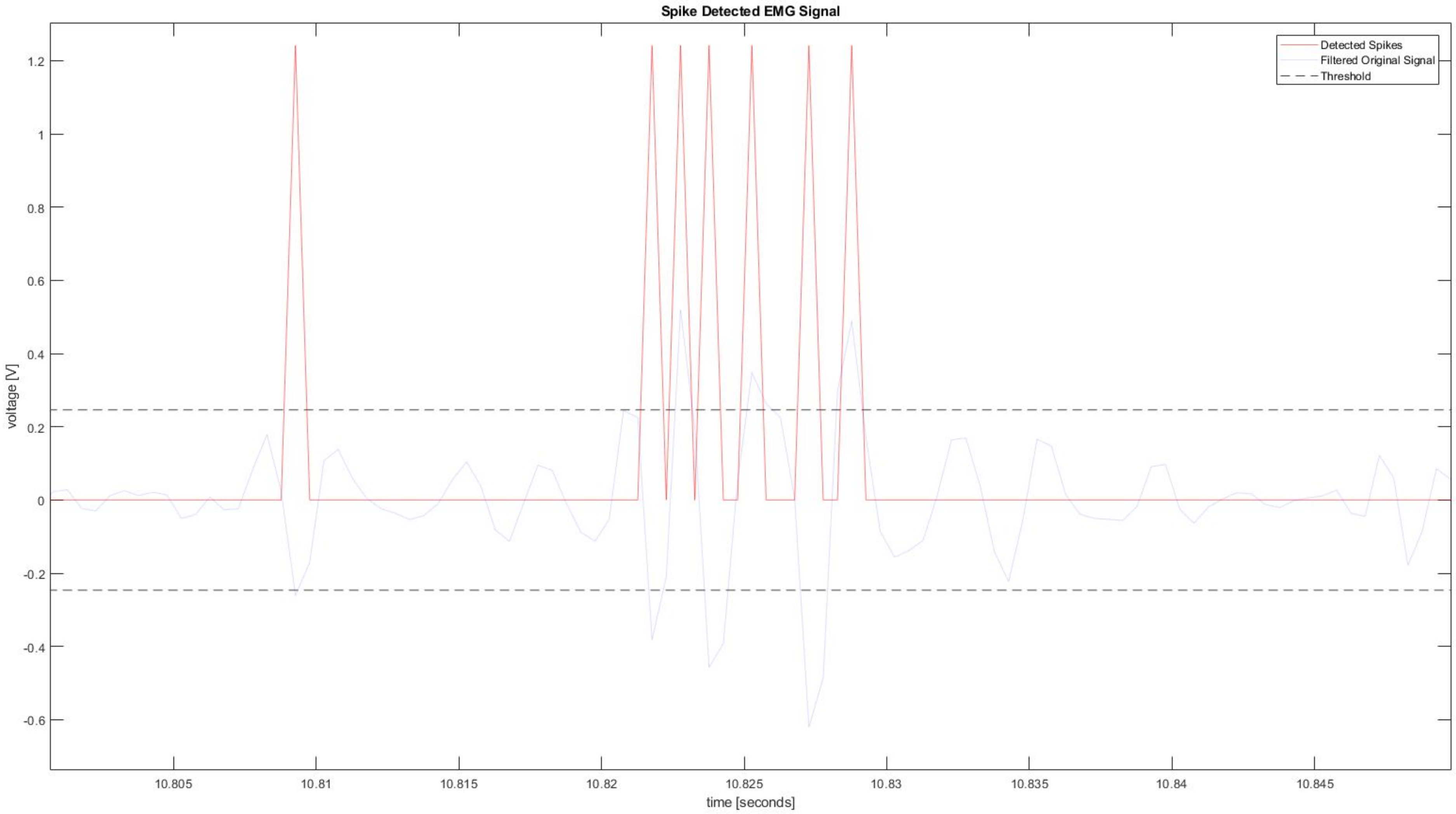

Channel 2
voltage [V]
time [seconds]
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
2
4
6
8
10
12
14
16
18
20


Power Spectrum of Mean Centered Signal
Power
Frequency [Hz]
0
0.5
1
1.5
2
2.5
0
100
200
300
400
500
600
700
800
900
1000

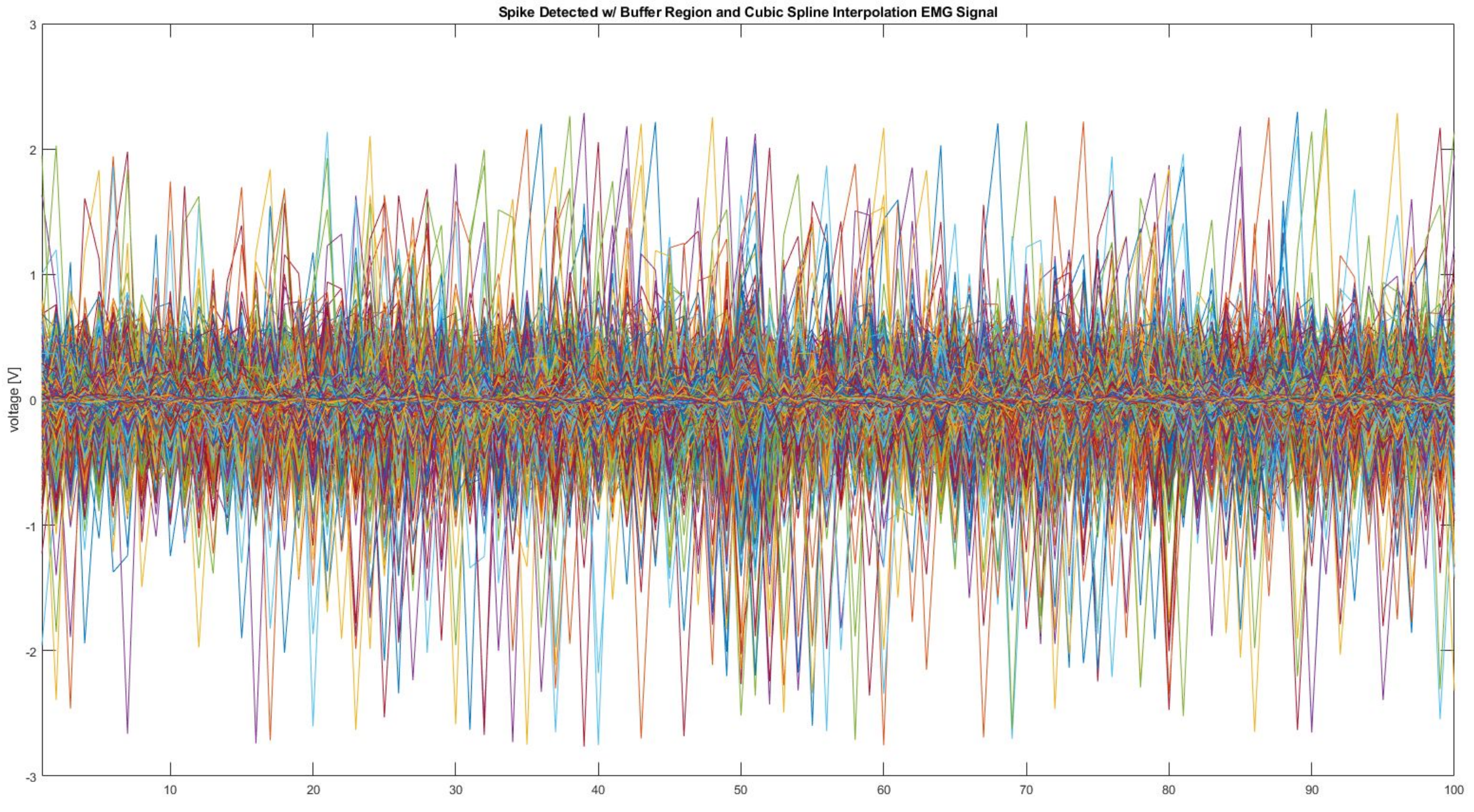

Power Spectrum of Mean Centered FILTERED Signal
Power
Frequency [Hz]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000

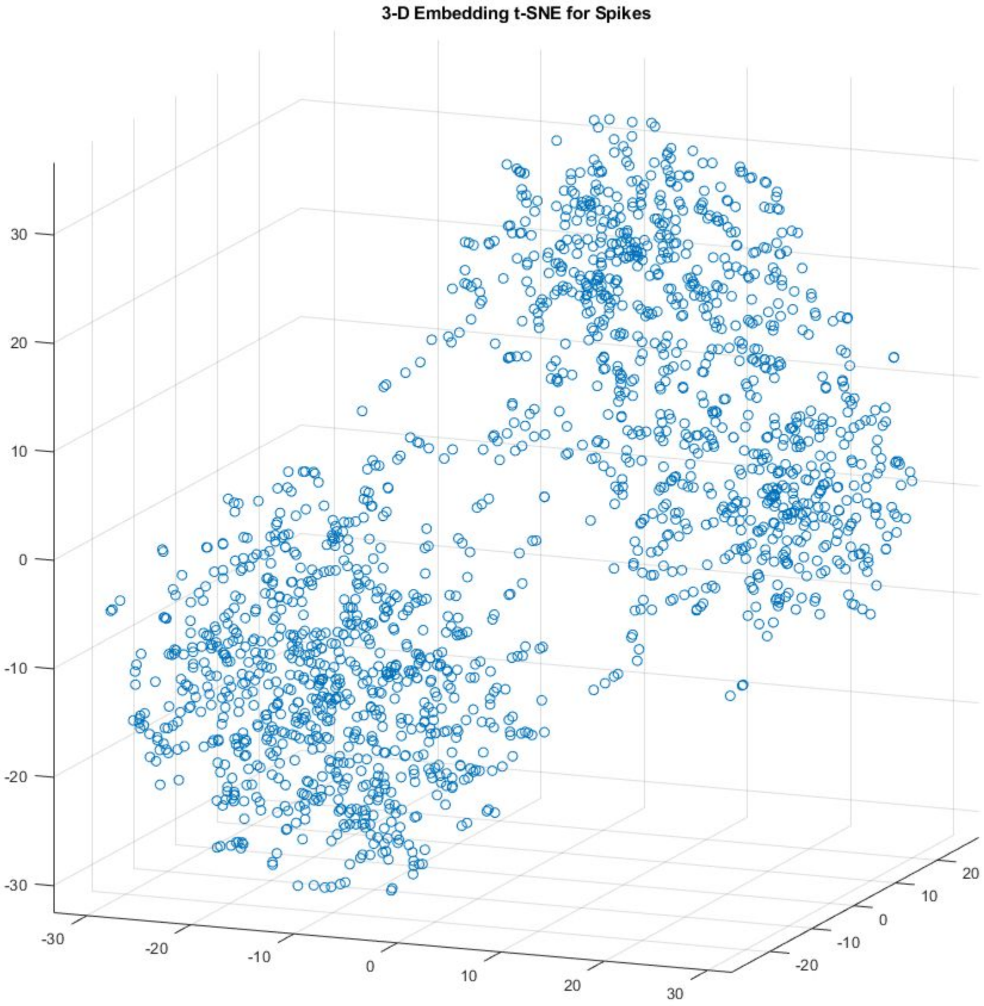


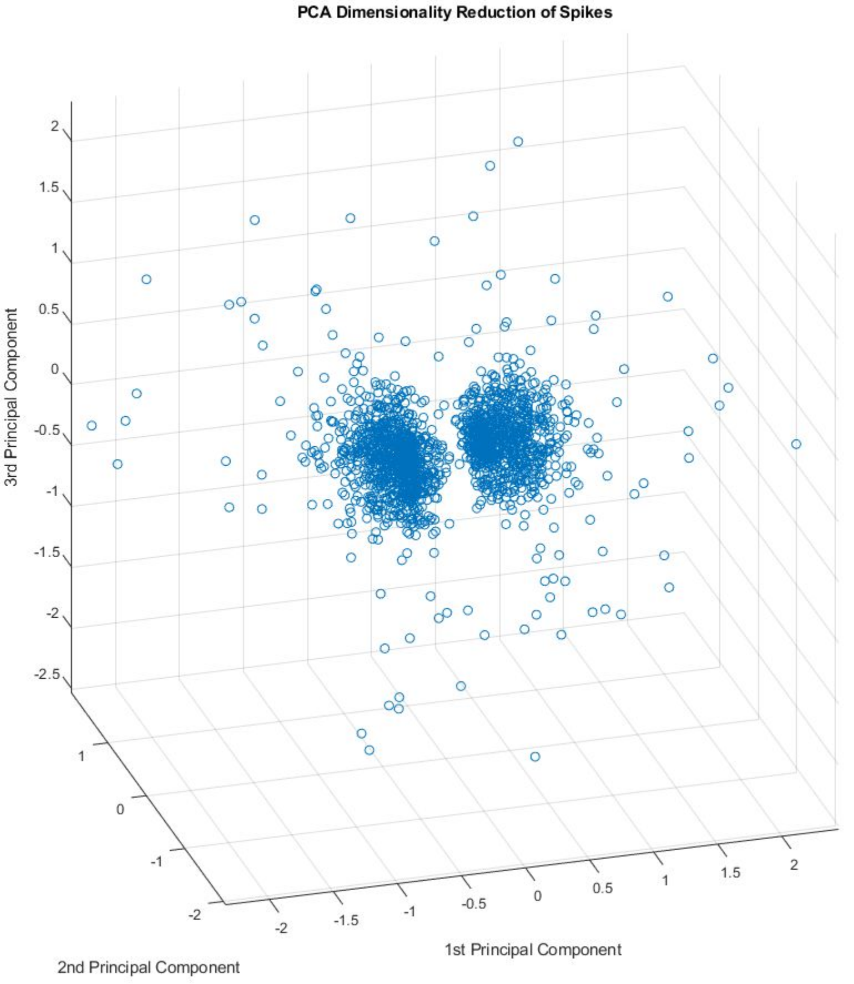
Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
time [seconds]
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
0
2
4
6
8
10
12
14
16
18
20

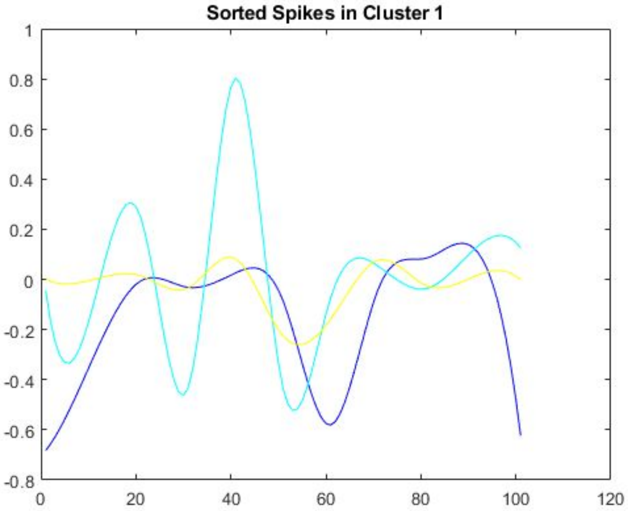


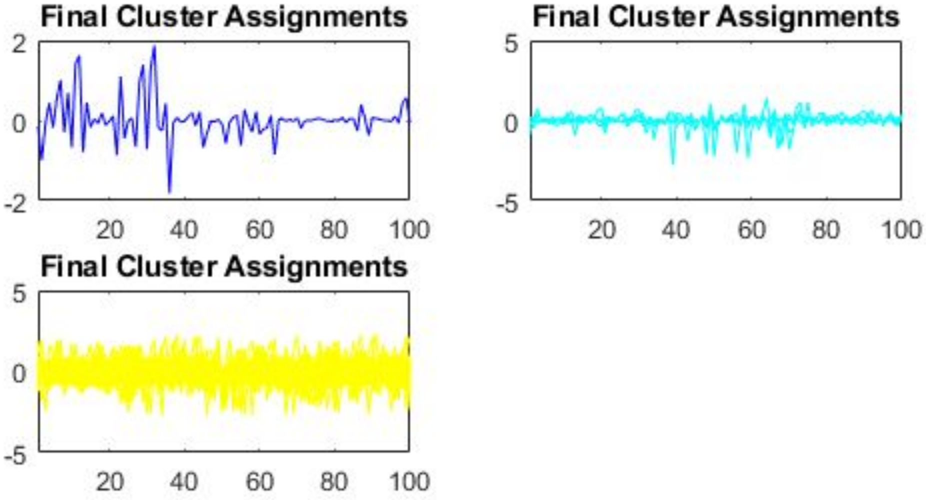
Spike Detected EMG Signal
Detected Spikes
Filtered Original Signal
Threshold
voltage [V]
time [seconds]
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
10.805
10.81
10.815
10.82
10.825
10.83
10.835
10.84
10.845


Spike Detected w/ Buffer Region and Cubic Spline Interpolation EMG Signal
voltage [V]
3
2
1
0
-1
-2
-3
10
20
30
40
50
60
70
80
90
100


3-D Embedding t-SNE for Spikes


PCA Dimensionality Reduction of Spikes
3rd Principal Component
2nd Principal Component
1st Principal Component


Sorted Spikes in Cluster 1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
20
40
60
80
100
120


Final Cluster Assignments
2
0
-2
20
40
60
80
100
Final Cluster Assignments
5
0
-5
20
40
60
80
100
Final Cluster Assignments
5
0
-5
20
40
60
80
100