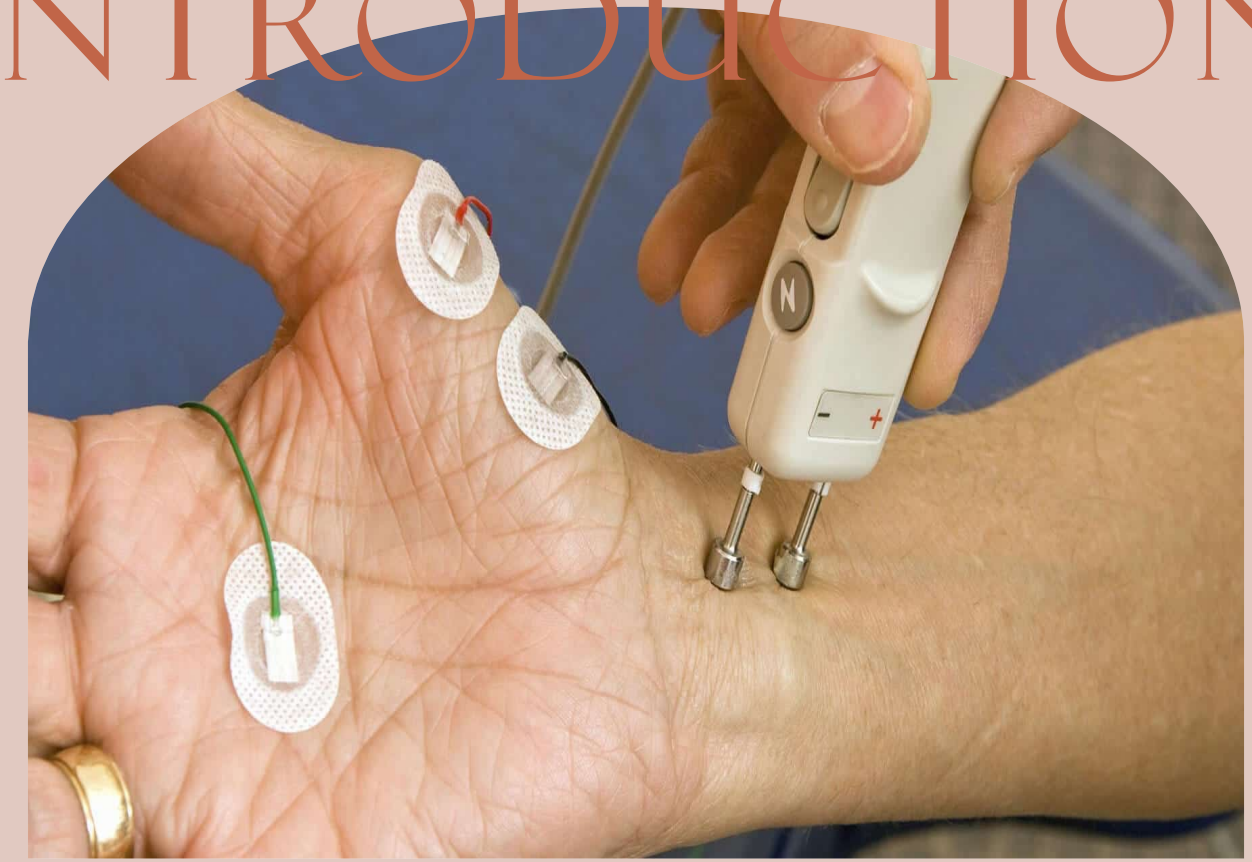
EMG SPIKE SORTING

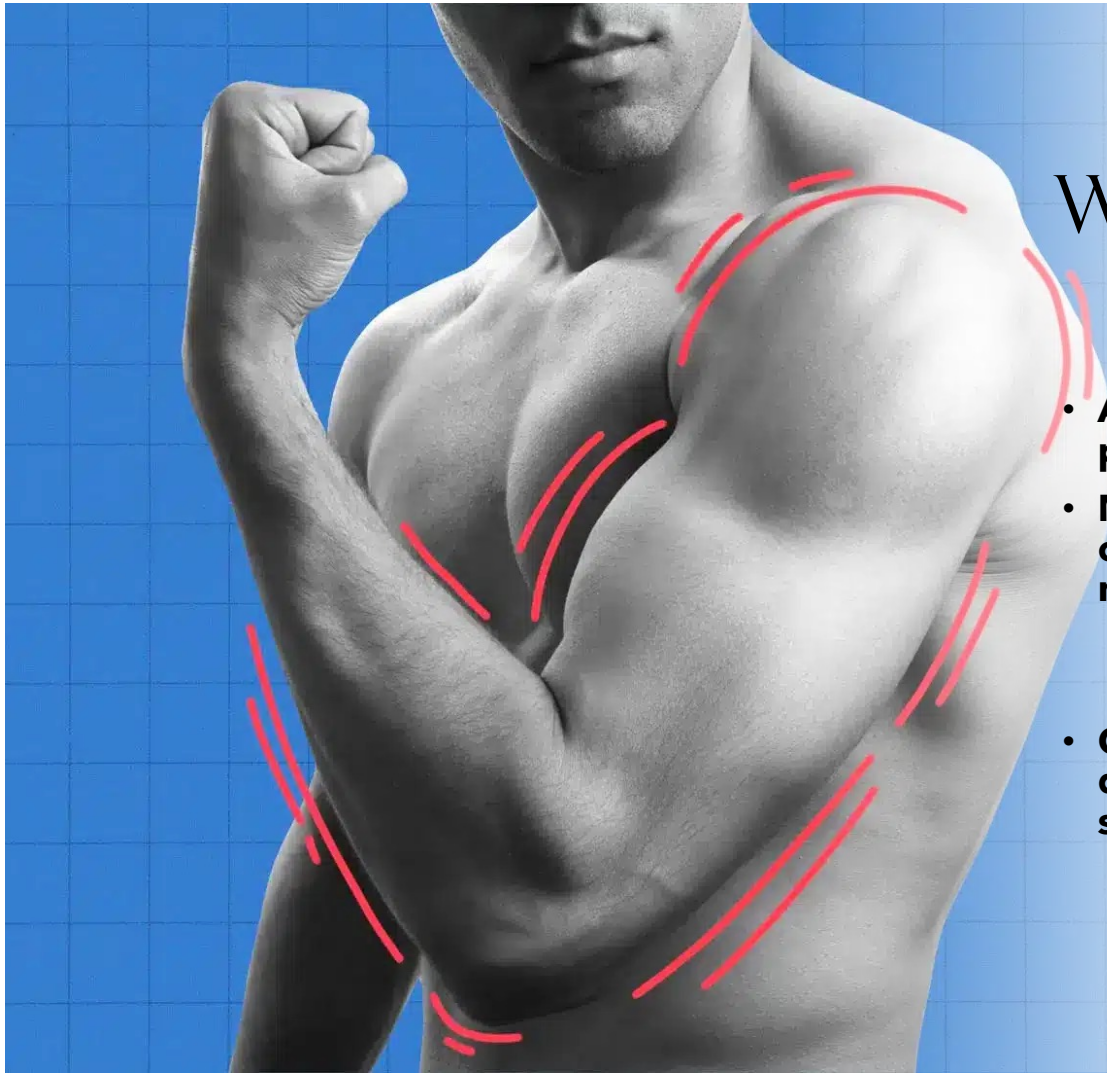
Martin Bourdev

December 10, 2023

OUTLINE

INTRODUCTION





WHAT IS EMG?

- **Aggregation of motor unit action potentials (MUAPs)**
- **MUAPs: sums of action potentials of individual muscle fibers of one motor unit**
  - Known as muscle fiber action potential (MFAP)
- **Goal: Classify MUAPs into different motor units based on sizes and shapes.**

IN PRACTICE

Data Collection

- Electrodes collect data from many motor neurons
- Most MUAPs are too small to distinguish
- Smaller electrodes provide more resolution

Clinical Use

- Calculate frequencies of each motor neuron
- Decode patient actions from frequencies
- Build prosthetics controlled by muscles

METHODS (CHANNEL 2)

DATA PREPROCESSING

Loading Data

- Use Python numpy library
- Sampling rate: 2000 Hz
- 40001 data points
- Two channels

Filtering Data

- scipy Butterworth filter
- 100 Hz to 999 Hz
- Filter out low frequency noise, centers data
- Chosen from literature review

Nonlinear Energy Operator

- Emphasizes spikes
- $y(t) = x^2(t) - x(t-1)x(t+1)$
- Better for detection than filtered EMG data

Spike Detection

- Determine amplitude threshold from NEO
- Exclude points that are too close to each other to avoid double-counting
- Calculate indices of spike peaks

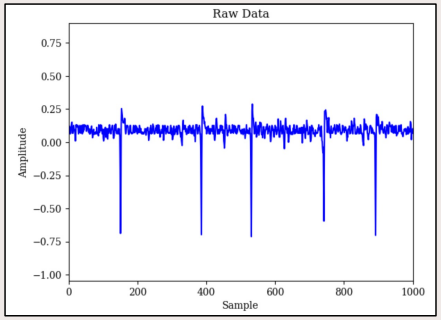
DATA PREPROCESSING

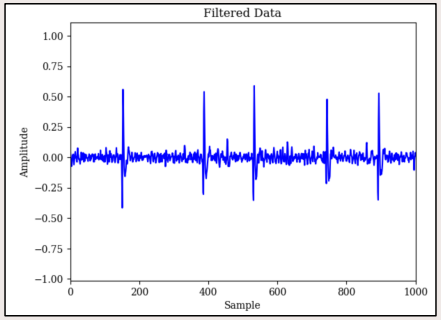
Loading Data

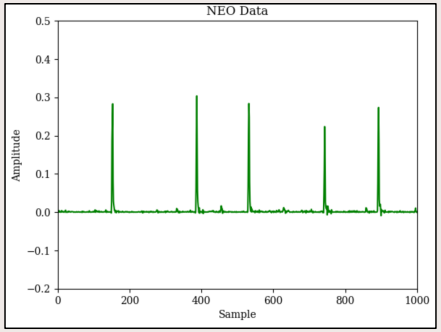
Filtering Data

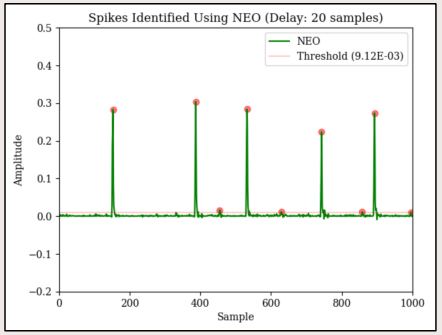
Nonlinear Energy Operator

Spike Detection









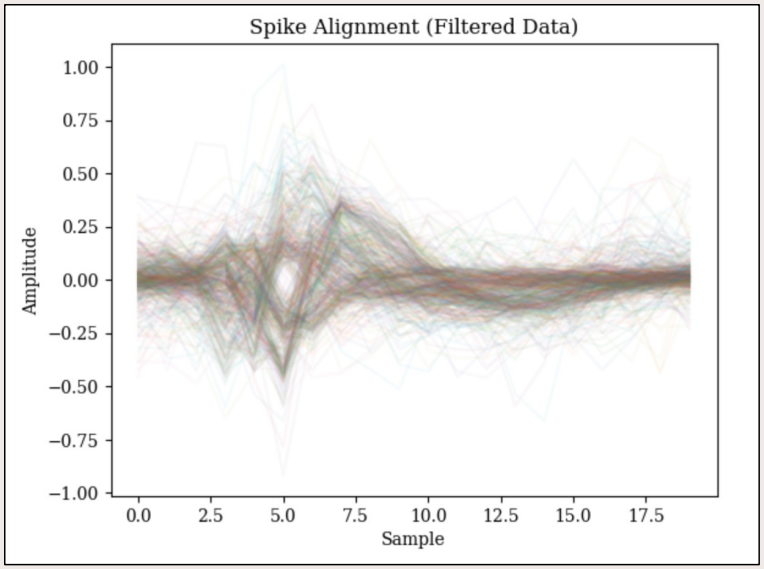
SPIKE ALIGNMENT

Methods

- Find the local maximum of identified spikes
- Align them with 5 data points to the left and 15 data points to the right
- Large potential for error due to misalignment

Usage

- Construct 20–dimensional matrix of spike data
- Compress using PCA
- Cluster and classify spikes based on waveform



PCA CLUSTERING

K-MEANS

Style	Determining K	Process
• Top-Down • Divide clusters using straight lines	• Elbow method • Inertia metric	• Randomly assign k centroids • Assign data to clusters • Update centroids, repeat

AGGLOMERATIVE

Style	Determining K	Process
• Bottom-Up • No clear cluster boundaries	• View dendrogram • Track distance between cluster steps	• Assign each point to its own cluster • Combine two closest clusters • Repeat until one cluster

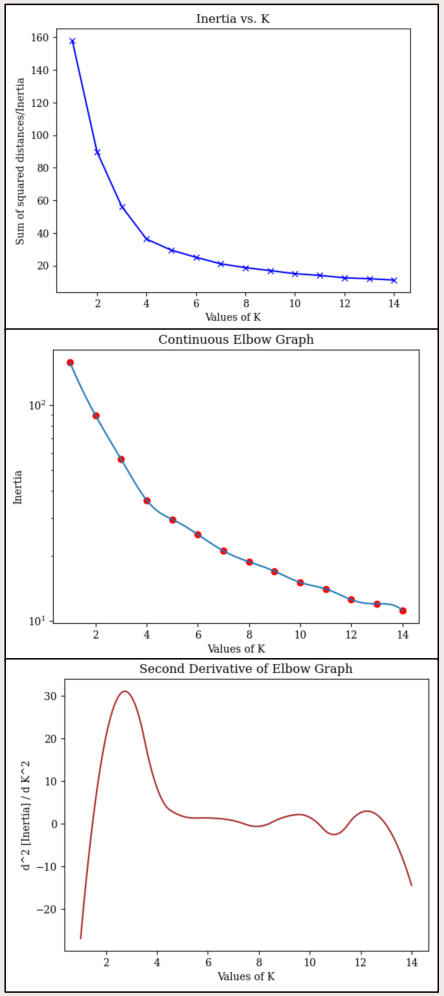
DETERMINING K (K-MEANS)

What is inertia?

- Inertia: average squared distance between points to centroid
- Will always decrease with number of clusters
- Must balance with number of clusters

Determining k

- Plot inertia vs. k (elbow graph)
- Make continuous, take second derivative
- Find peak of the second derivative, indicating inflection point (best tradeoff between inertia and k)
- In this case, **k = 4**.



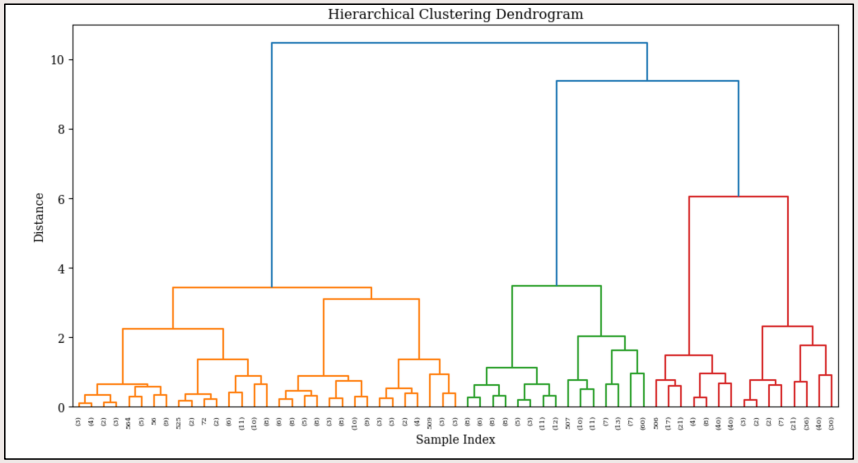
DETERMINING K (AGGLOM.)

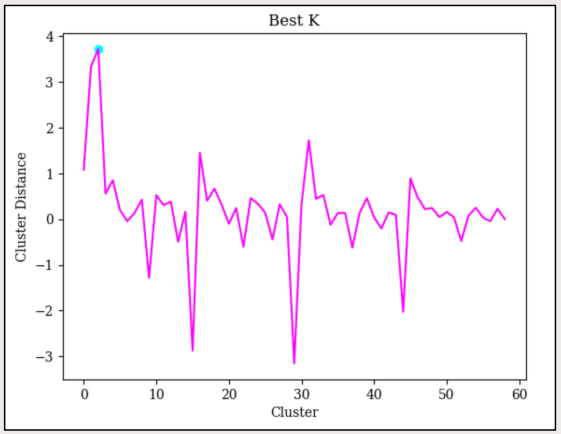
What is a dendrogram?

- Plots the distance between each cluster when they are unified
- Logically, clusters that are more real will have larger distances between each other, indicated by the y axis of the dendrogram

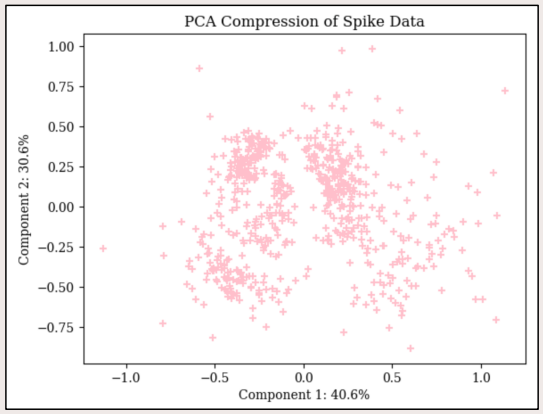
Determining k

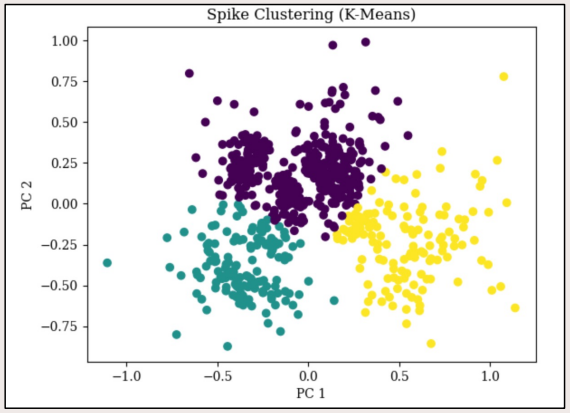
- Create dendrogram, identify distances between each cluster
- Plot the distances, find the max
- Find the largest value of k close to the max (bias toward larger k)
- In this case, **k = 4**.

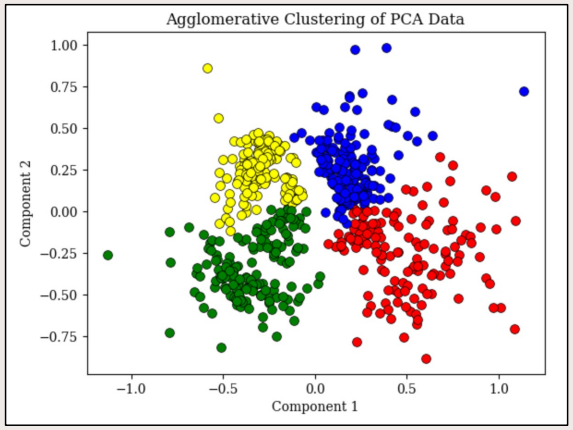




K-MEANS VS. AGGLOMERATIVE

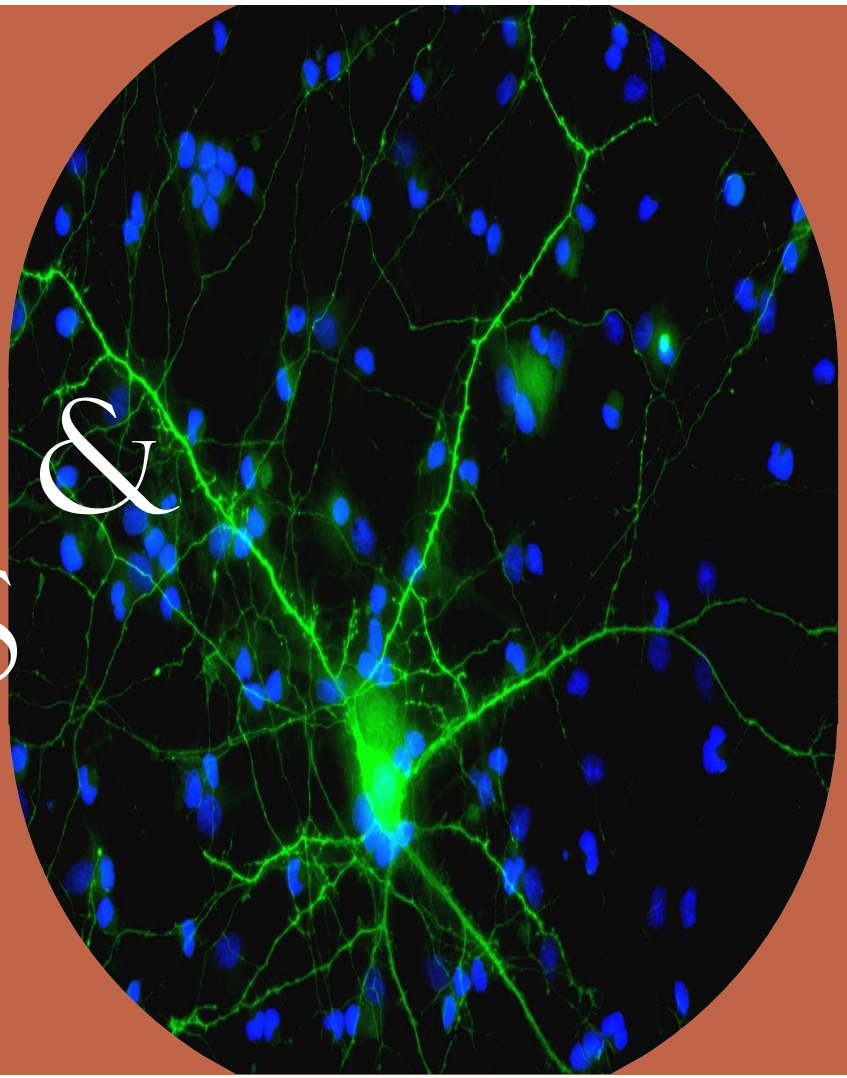


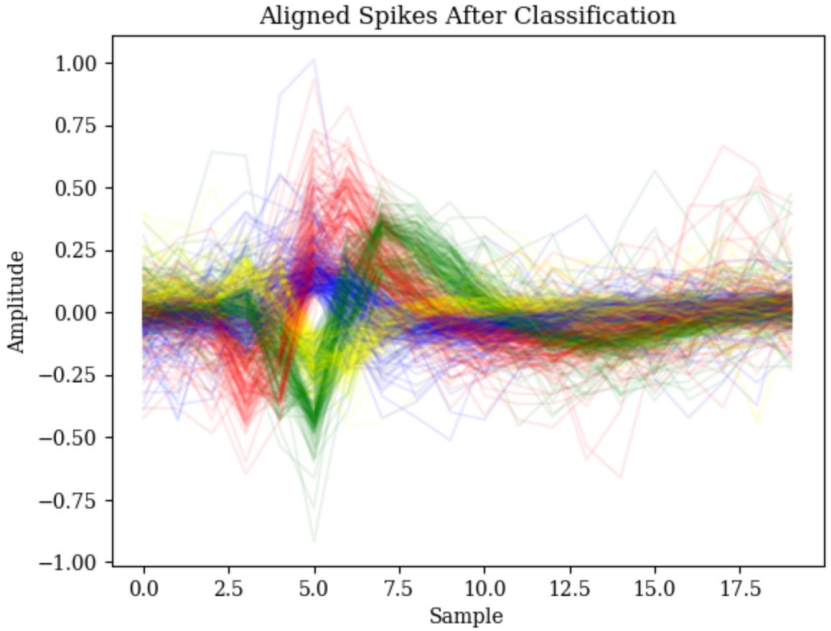


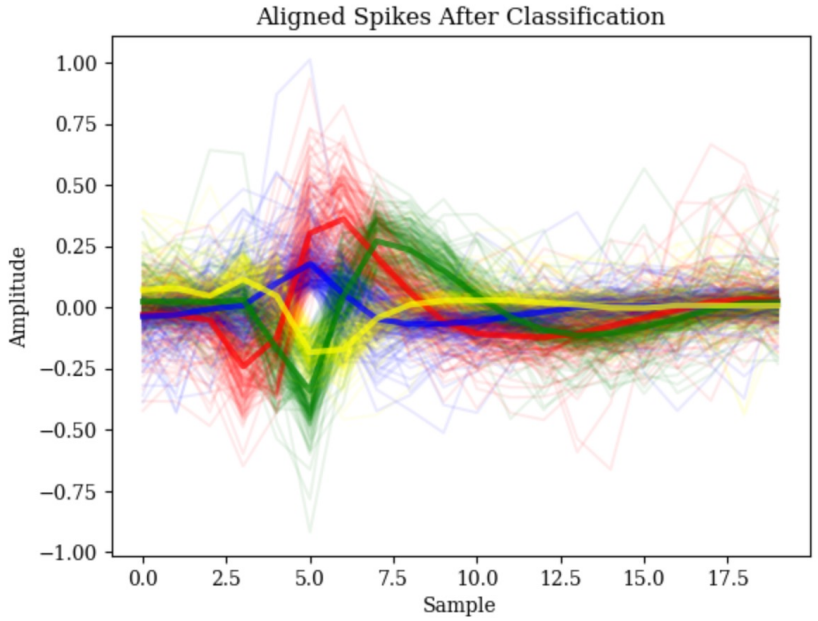


VERDICT: AGGLOMERATIVE

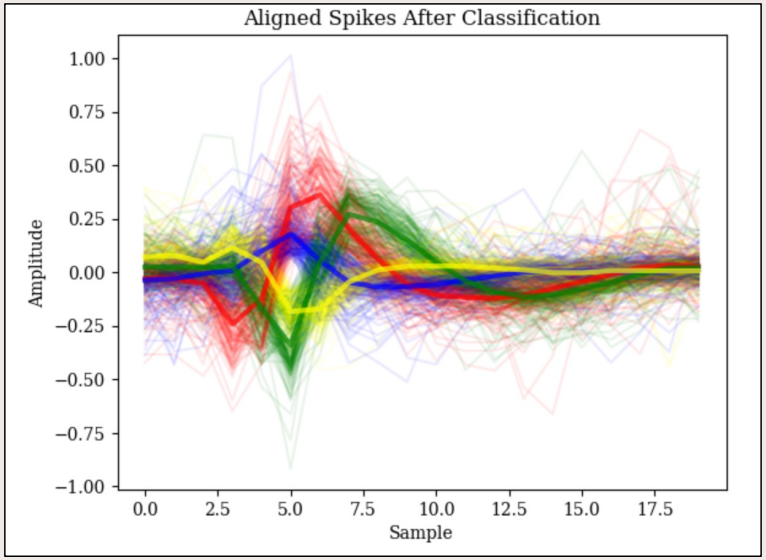
RESULTS & ANALYSIS







SPIKE CLASSIFICATION



ANALYSIS

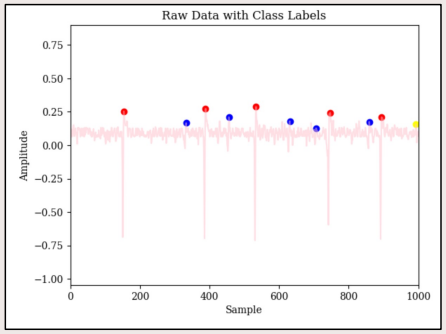
Spike Alignment

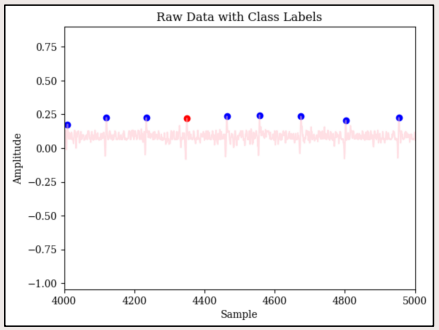
- Classification does not appear too broad (many different waveforms of the same color)
- Classification does not appear too narrow (many colors following the same waveform)
- Red and green are probably the same waveform that were not aligned correctly
- For decoding, it is better to classify one MUAP as two instead of classifying two different MUAPs as one

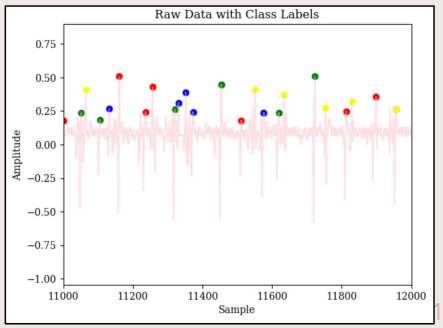
ANALYSIS

Raw Data

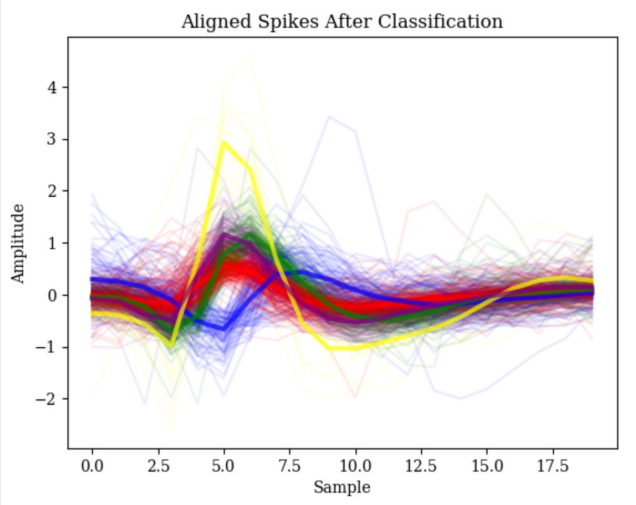
- Top: large red spikes are not double-counted, which has previously been a problem
- Middle: Blue spike likely erroneously as red spike
- Bottom: Blurred distinction between red and green spikes, as well as unusual blue spikes.
- Hard to visually distinguish small spikes from noise in raw data



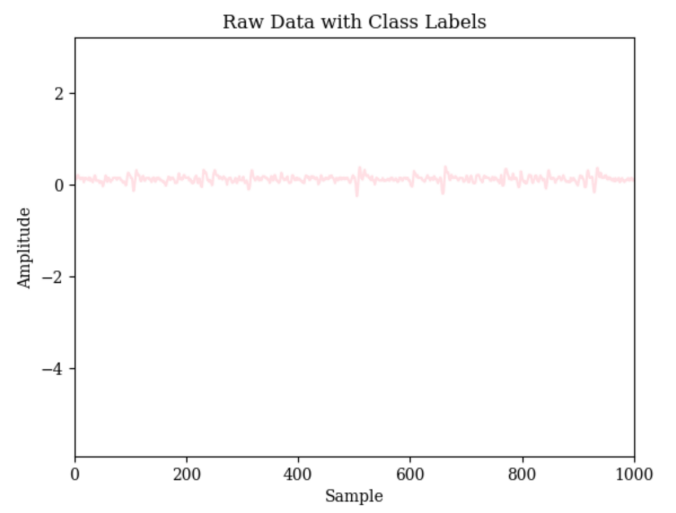




ANALYSIS (CHANNEL 1)



- K is too large (purposeful bias)
- Red, green, and purple spikes are the same waveform
- Yellow is a small number of outliers



- Visible spikes that are not classified
- Amplitude threshold was likely too high to include them
- Difficult to find amplitude threshold that works for both files

DISCUSSION

Strengths

Clear and decipherable classification, as opposed to neural networks

Weaknesses

Sensitivity to k, amplitude threshold, spike length/delay

Poor alignment in (green and red, channel 2)

Opportunities

Smarter NEO amplitude threshold calculation mechanism

Better alignment to avoid separating one waveform into multiple

Threats

Transferability from one channel/electrode to another

OPPORTUNITIES

- Build more flexible algorithm to calculate accurate spike detection thresholds for many different electrode recordings
- Use Fourier transform or spectral power instead of raw data to remove sensitivity to precise spike alignment
- Instead of hard-coding a delay in between identified spikes to avoid double-counting, calculate correlations between spike classes to identify double counts
- Provide opportunity for algorithm to combine classes that turn out to refer to the same waveform

THANK YOU

Martin Bourdev
UID: 205-491-049
UCLA
Bioengineering M260
December 10, 2023