Spike Sorting in Matlab

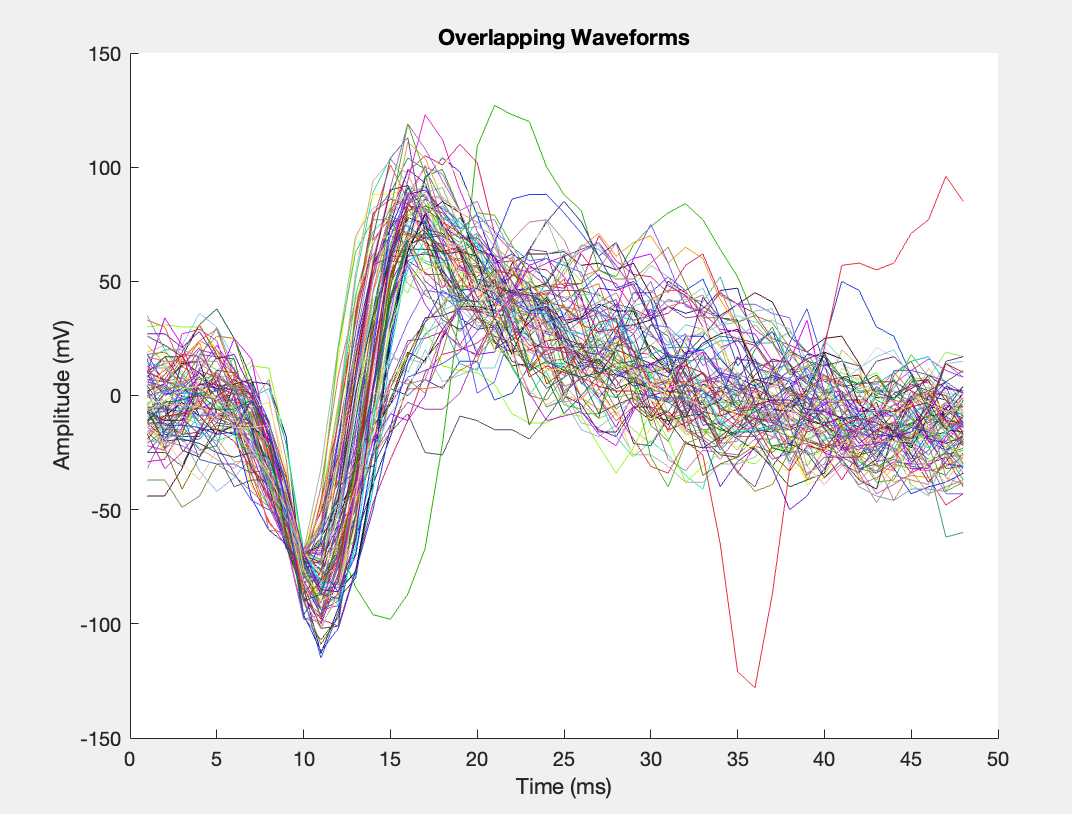
Neural Data Analysis

Singh, Jaspreet

Feb 2024

**Step 1.**

**A**

**Visual Inspection of Spikes in 100 different waveforms**

A screenshot of a computer screen

Description automatically generated

Original 100 Waveforms

**B**

**Figure 1.** *Data was collected from one electrode during experimentation in an animal model. Figure 1 shows the snippets of neural activity from around the time of a threshold crossing. Each waveform is displayed over a period of 50 milliseconds (x-axis represents time). The snippets are saved from .4 ms before the threshold crossing to 1.6 ms after it. Additionally, the snippets are measured display amplitude, measured in millivolts (y-axis). The visualization of these 100 waveforms collected during this experiment allows for sorting by eye. Spike sorting by eye can help differentiate between two different neurons which are present during experimentation. In this experiment 80,397 such samples were collected.*

*A. The top portion of figure A shows all of the waveforms overlapping one another. We can see two distinct waveforms in the image.*

*B. 100 of these waveforms were randomly selected to visually inspect to try to ensure lack of bias and representation of neuron/neuron populations in the sample. Upon visual inspection of these 100 waveforms, there appear to be around three distinct waveforms. Distinct shapes can be seen in this sample, for example: waveform 1 (row 1, column 1) has a spike and then faster hyperpolarization than waveform 33296 (row 4, column 3) whose hyperpolarization is must slower, and waveforms like waveform 67403 (row 8, column 5) appear to have less sharpness or structure in their spike pattern and have voltages closer to baseline of 0.*

**Step 2** A graph of a graph with a red and yellow triangle

Description automatically generated

**Figure 2.** *Data was collected from one electrode during experimentation in an animal model. Figure 2 shows the 3-dimensional principal component analysis (PCA) of the collected neural data from this experiment. PCA is a machine learning technique that can help reduce noise, visualize data, and reduce the number of variables or the dimensions in a dataset. The snippets in this dataset were sampled at 48 points. Each waveform snippet thus becomes a point in a 48-D space. So, neurons that are more similar to one another would be found more densely in space.*

*When applied to this data set, using the surf function in MATLAB, there appear to be two dense gaussian style waveforms in the data. These two spikes appear to show that there are two distinct types of neurons in this dataset. One, shown on the left, is more dense than the other, shown on the right. Upon performing this 3D PCA analysis on the neural dataset, there appear to be two different types of neurons.*

A blue dot diagram with white text

Description automatically generated with medium confidence

**Figure 3.** *Figure three is a two-dimensional principal component analysis plot. There appear to be two distinct, dense cluster and one more sparely populated cluster. These clusters help to indicate how many types of neurons are present in our recordings. The density relates to how often these thresholds were detected. This figure indicates the possibility of up to 3 neuron types.*

**Step 3.**

A screen shot of a graph

Description automatically generated

**Figure 4.** *Figure 4 shows k-means clustering of the collected neural data during this experiment. Data was collected from one electrode during experimentation in an animal model. K-means is a helpful technique to identify clusters of data by sectioning out parts of data. The number of clusters to input into the k-means algorithm was manipulated. Ultimately, the two clusters (blue and red, shown above) were identified. These appear to be the appropriate number of k-means cluster for this dataset because the clustering represents and aligns with earlier plotting of PCA analysis. K-means is done through the manipulation of PCA 1 (x-axis) and PCA 2 (y-axis).*

**Step 4.**

**100 waveforms by cluster**

*A screenshot of a computer

Description automatically generated*A screenshot of a graph

Description automatically generated

Red – cluster 1

Blue – cluster 2

**100 waveforms by cluster**

**Time (mS)**

Amplitude (A)

**Figure 5.**  *Figure 5 (top) demonstrates 100 randomly selected waveforms from the collected neural data output of the electrode recording. Earlier figures show evidence of 2 different clusters, suggesting evidence of two different neurons were recorded. Based on the precious cluster assignments from k-means, the blue and red clustering were applied, and the waveforms were plotted above. Red waveforms appear to be similar in shape and seem to be “faster-spiking” whereas the other blue waveforms appear to have a more gradual hyperpolarization period. This indicates that clustering was appropriate, as waveforms in the red cluster appear similar to other red clustered waveforms and blue cluster waveforms appear similar to other waveforms. This is significant as it yields information about the type and number of cell types that the electrode was able to collect data from. While there is tremendous diversity in the properties of different types of neurons, more broad classifications can be made on cells types dependent on where electrode recording took place. The distance from an electrode can affect waveform shape, as well. Some examples of fast spiking neurons include Interneurons in the cortex (GABAergic interneurons in the cerebral cortex, such as parvalbumin-expressing interneuron which play a crucial role in regulating the balance of excitation and inhibition in cortical circuits, Purkinje Cells in the cerebellum (inhibitory neurons in the cerebellum which play a role in motor coordination and learning), basket cells in the hippocampus (interneuron in the hippocampus which contribute to the regulation of network activity and synaptic plasticity in the hippocampus). Some examples of slow spiking neurons include: pyramidal neurons in the cortex (principal excitatory neurons in the cerebral cortex and often exhibit slower firing rates compared to fast-spiking interneurons which play a central role in cortical information processing), thalamocortical neurons (relay sensory information from the thalamus to the cortex and contribute to the modulation of attention and consciousness), hippocampal Principal Cells (such as CA1 pyramidal neurons, generally have a slower firing rate compared to some fast-spiking interneurons in the same region and are involved in learning and memory processes). Figure 5 (bottom) shows two clusters (determined above), red and blue. This figure is another visualization of two different neuron types.*

Step 5.

Reconstructed waveform

time (mS)

Amplitude (A)

**A screenshot of a graph

Description automatically generated**

***Figure 6.*** *Figure 6 shows PCA components 1 and 2 reconstructed after rotation through space. The rotation of the waveform along these PCA components back into the original space displays information regarding the variability captured by each principal component. These reconstructions enable some comparison to the original waveforms which allows interpretation of which features contribute most strongly to a particular principal component and how the relationships among features manifest in the data.*

(see below for step 6)

**Step 7.**

**Raster Data Plot**

A comparison of a bar code

Description automatically generated with medium confidence

**Time (s)**

**Neural unit**

***Figure 7.*** *Figure 7 shows a raster plot of the data over a smaller time period. A small window of time was selected (10 seconds) and the amount of activity of the two neurons identified earlier (red and blue) was identified. Each raster line corresponds to 1 spike or action potential from a neuron. One cell type, shown to be red in the plot above, fires at a much higher rate than the other cell type, shown to be blue.*

**Step 6**

A graph with a dotted line

Description automatically generated

**Figure 8.** *Figure 8 shows the variance of the data. The original data has 48 dimensions. However, when variance is plotted, we can see that 13 dimensions account for 90% of the dimensions which can be visualized by the elbow of the graph. The elbow in a variance plot is indicative of appropriate clustering, as our elbow occurs above 90% it suggests that the clustering was done appropriately and represents over 90% of the data.*

**Step 8.**

**Interspike Interval for Dataset**

A graph of a number of different types of numbers

Description automatically generated with medium confidence

**Figure 9.** *Figure displays histograms of interspike intervals (ISIs) for each clustered neuron – on the left we see neuron one and the right shows us neuron 2. The x-axis displays the ISI duration (in seconds) which is plotted against the y-axis that displats the count of occurrences. This provides insight into the firing patterns of neurons within each cluster. We can see that there are refractory periods where there is a minimum in the spiking patterns, showcasing that these are distinct neurons which need time to recover prior to more spiking. Additionally, both plots show ISI histograms with a distinct peaks showing that they are firing regularly and are distinct neurons. The presence of clear refractory periods in the ISI histograms suggests well-isolated single neurons, indicating successful spike sorting and isolation of individual neuronal units.*