

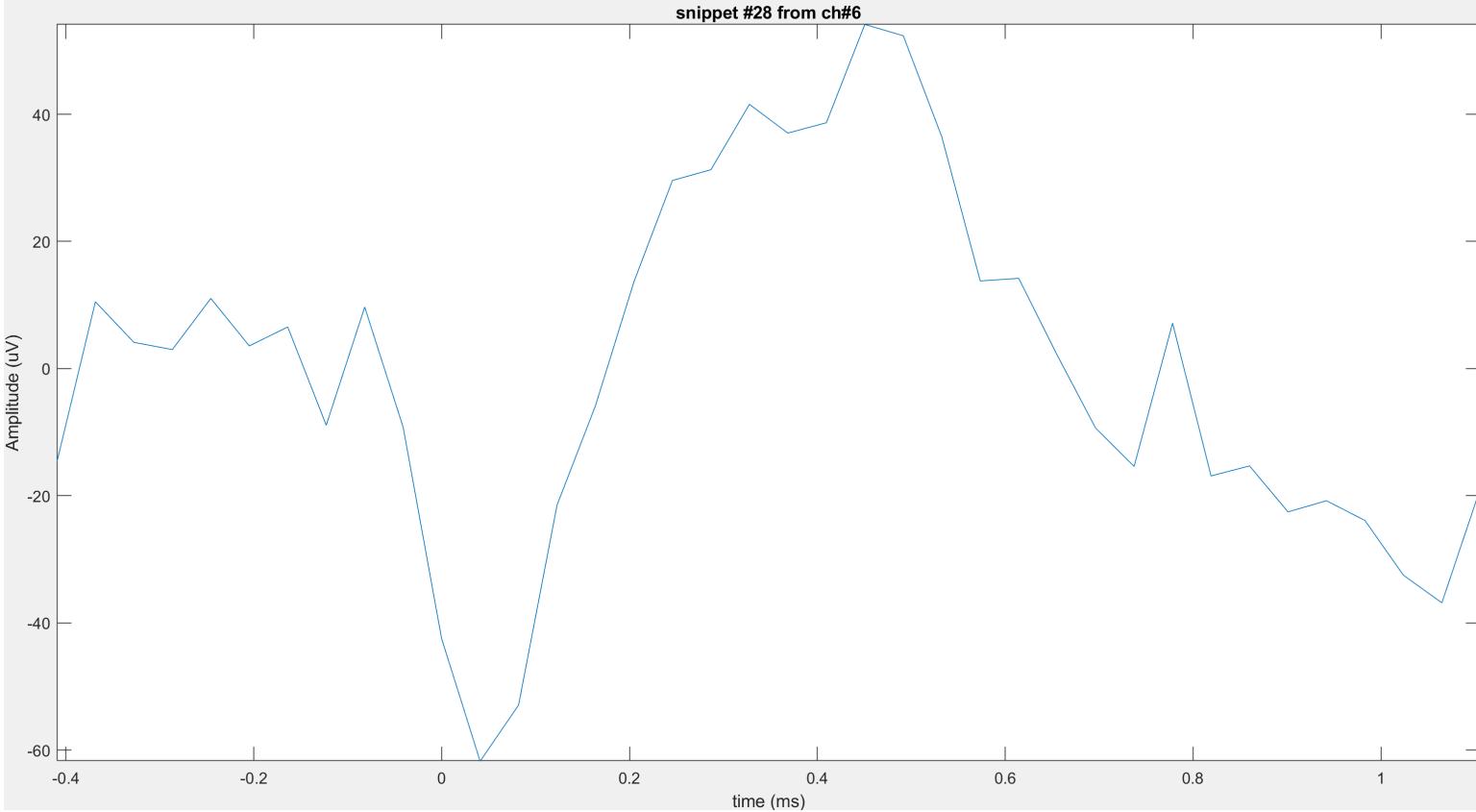
BIOE2615 | HW5

Arjan Singh Puniani

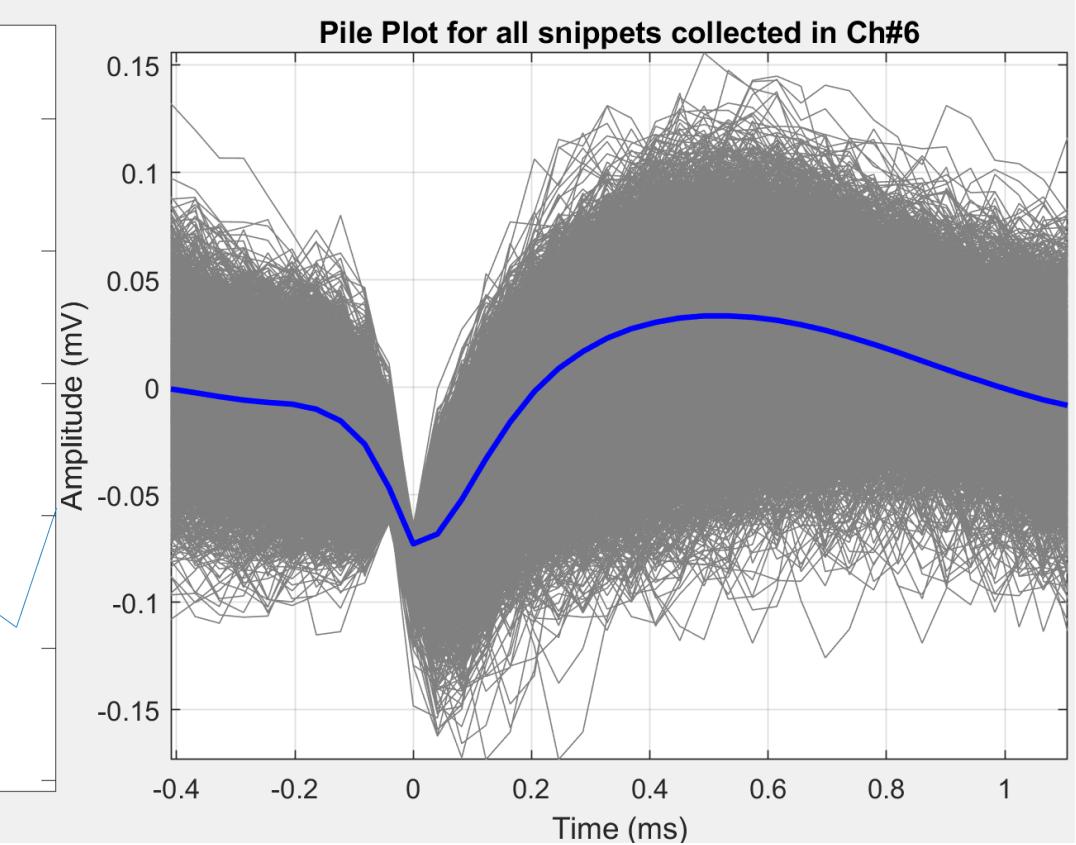
4537572 | arp150

Figure 1: in question 10, we looked at a snippet from our channel 6 data filtered with a 4th order high-pass Butterworth at a 300 Hz cutoff frequency. A high-pass filter is a way to parse our raw neural data into high-frequency action potentials (which we kept) and discard low-frequency potentials. We see a characteristic negative depolarization in *a*) and *b*) since the data was recorded extracellularly. In *a*), a single snippet in microvolts was plotted from -0.4 ms to +1.1 ms. A pile plot was made to show the average superimposed over all of channel 6's snippets.

a)



b)



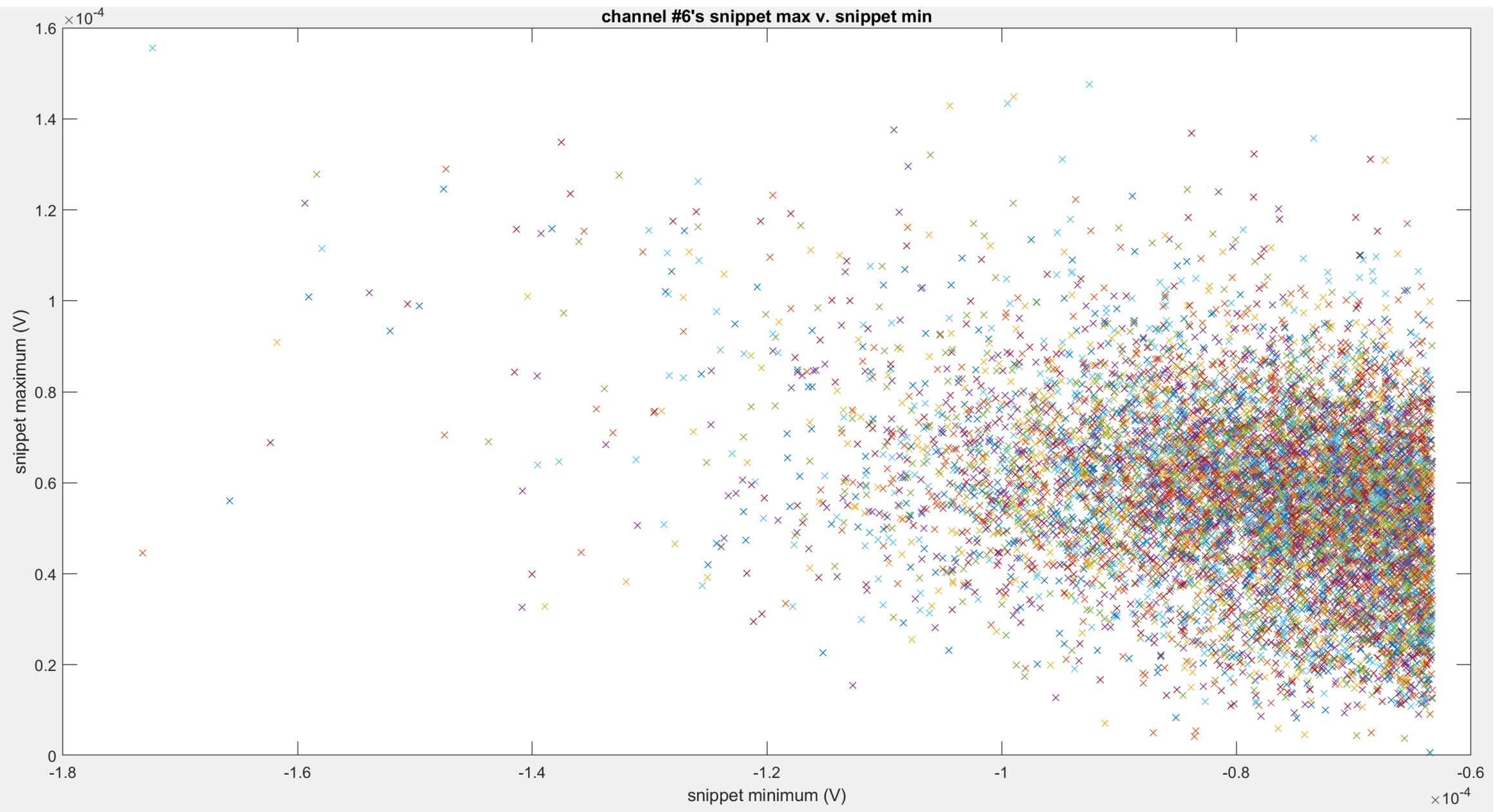
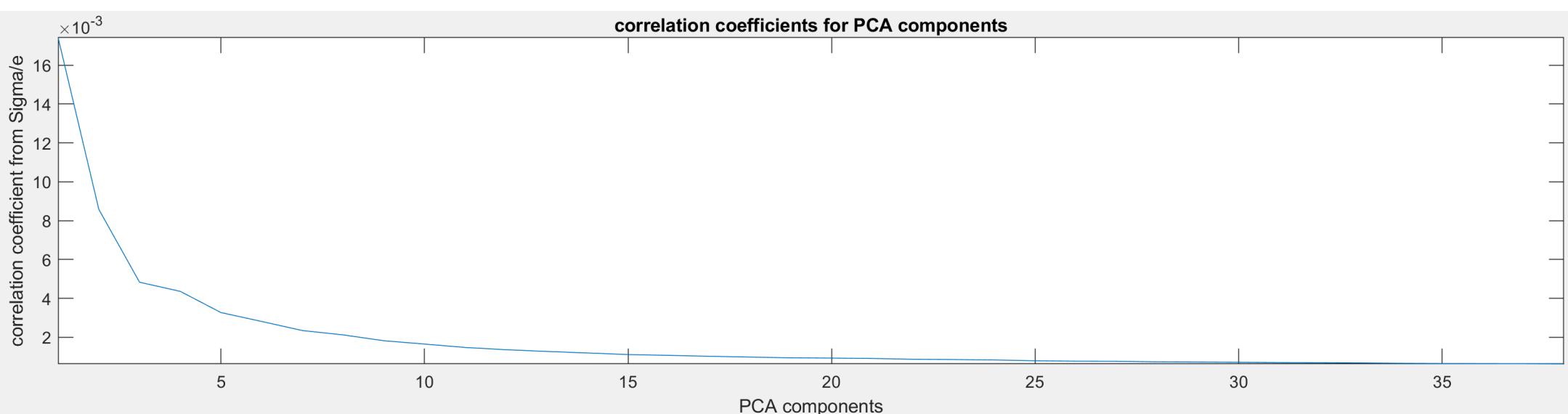


Figure 2: in question 11, we looked at the maxima and minima of each snippet from channel 6's data. There were no obvious clusters, so it appears as if the average snippet from channel 6 likes to hover between extrema.

a)



b)

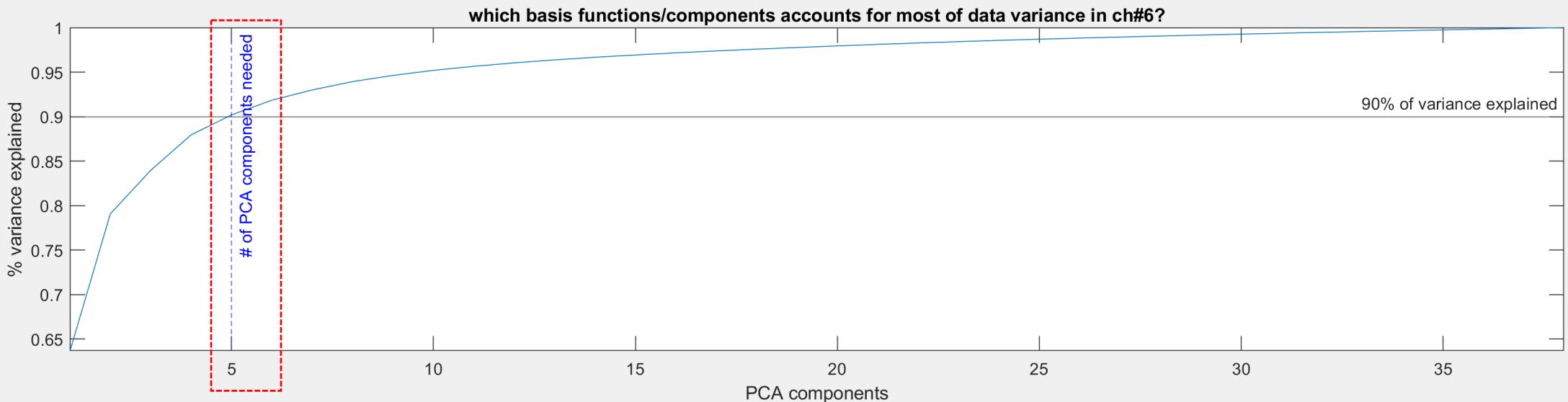


Figure 3a: in question 12, after performing singular value decomposition (svd) on the matrix of snippets for channel 6, we analyzed principal component contributions to % variance. Approximately 5 PCA components are required to explain 90% of the snippet data variance in b). Since the coefficient scaling matrix factorized out of the snippet data, e, is rank-ordered, we know this must be the first 5 diagonal elements. Every diagonal entry from e (in a)) shows decreasing subsequent contributions.

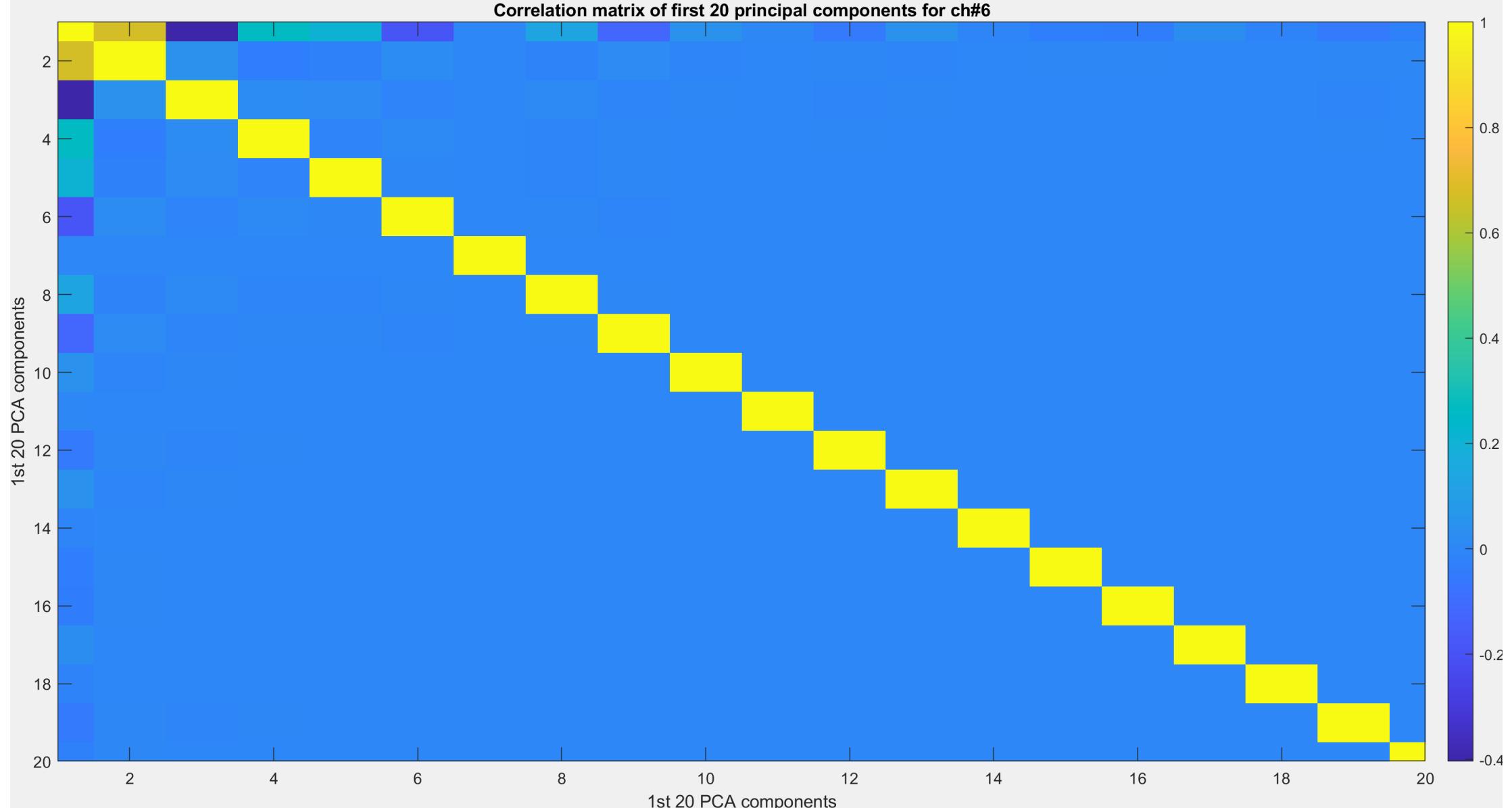
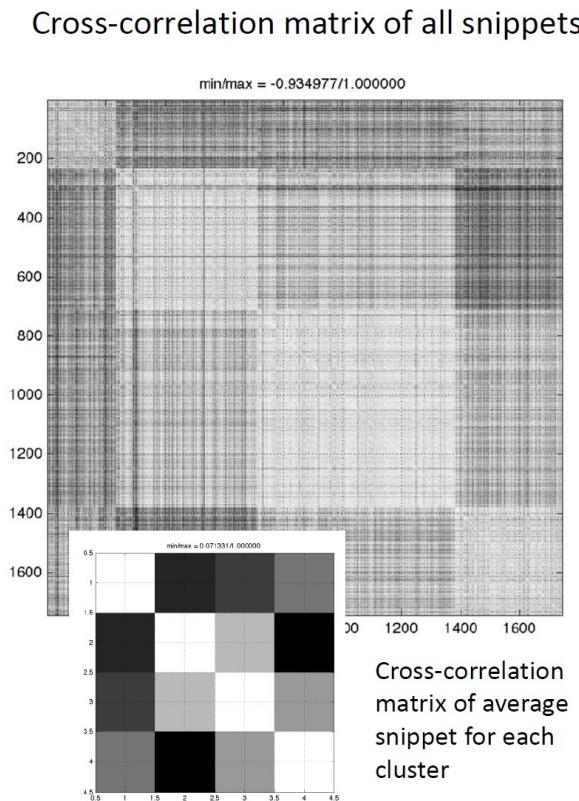


Figure 3b: Question 12 continued: Plotting the first 20 principal components in a cross-correlation matrix confirmed that most of the covariance can be explained by the first 5 PCA components. The diagonal elements, which are correlations between identical principal components, will have perfect correlations of 1.0 (yellow from the color bar), whereas any off-diagonal element represents a covariance between whatever two PCA components index that square.

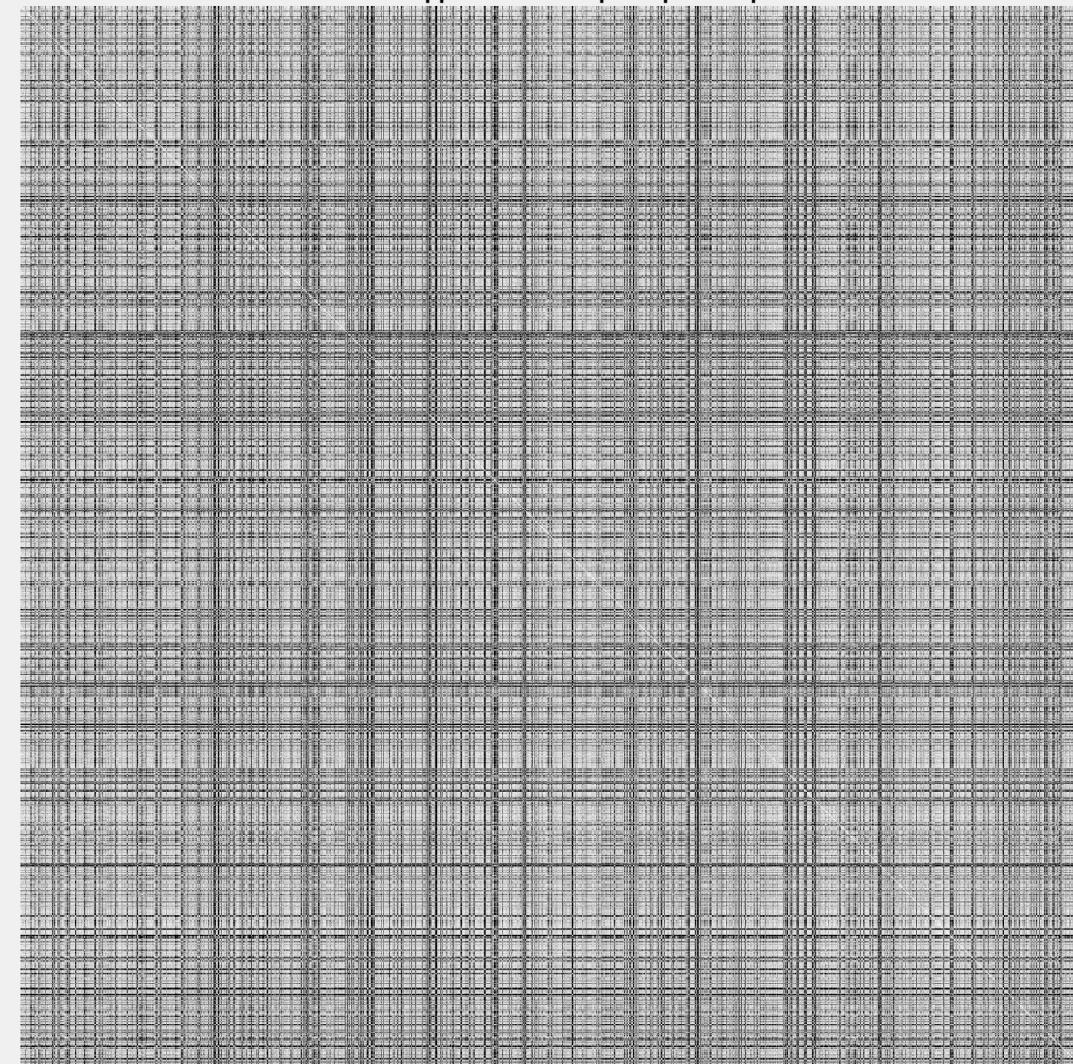
Figure 3c: For question 12, a cross-correlation matrix for all 9,307 snippets is plotted in *a*). The hope was to see distinct clusters, like we did in class (in *b*). The fact there's no clear discernible clusters in *a*) may have to do with the fact that the entries of the matrix are not normalized to the max of the snippets. The title of the cross-correlation matrix of all snippets in *b*) shows a max of 1.00, indicating possible normalization or standardization.

b)



a)

Cross-correlation of snippets matrix of principal components for ch#6



PCA components

Q13 -

- A is a rank-ordered square matrix ($m \times m$) of orthogonal waveforms (eigenvectors); in other words, a matrix containing the temporal shapes that we care about
- e (or Σ) is a diagonal matrix of scaling factors with dimensions $m \times n$
- W is a matrix of coefficients that gets scaled by e/Σ with dimensions $n \times n$
- The idea is that if we multiply A , e/Σ , and W , we will reproduce the snippet data
 - Some texts/articles present the svd equation as: $y = Ab$, where b is the product of ΣW^T , and y is our snippet data

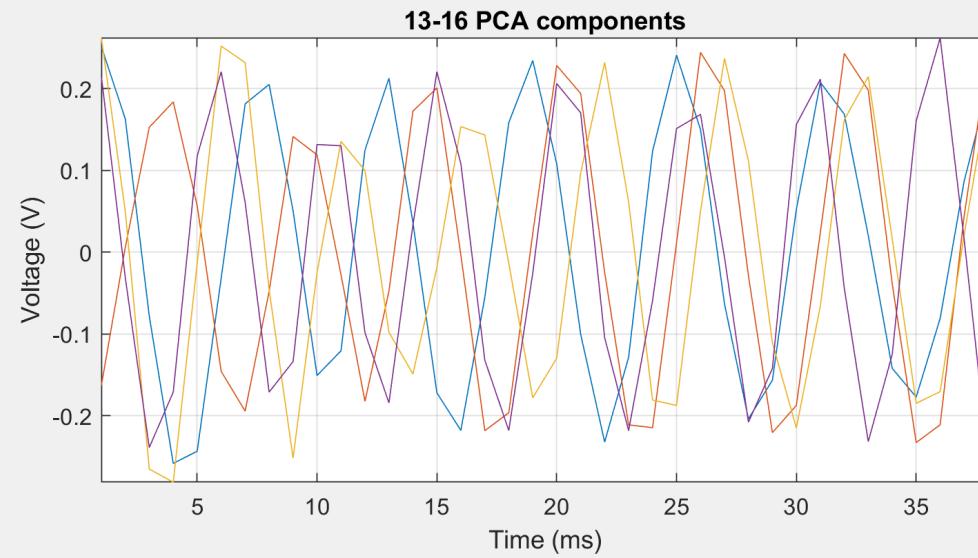
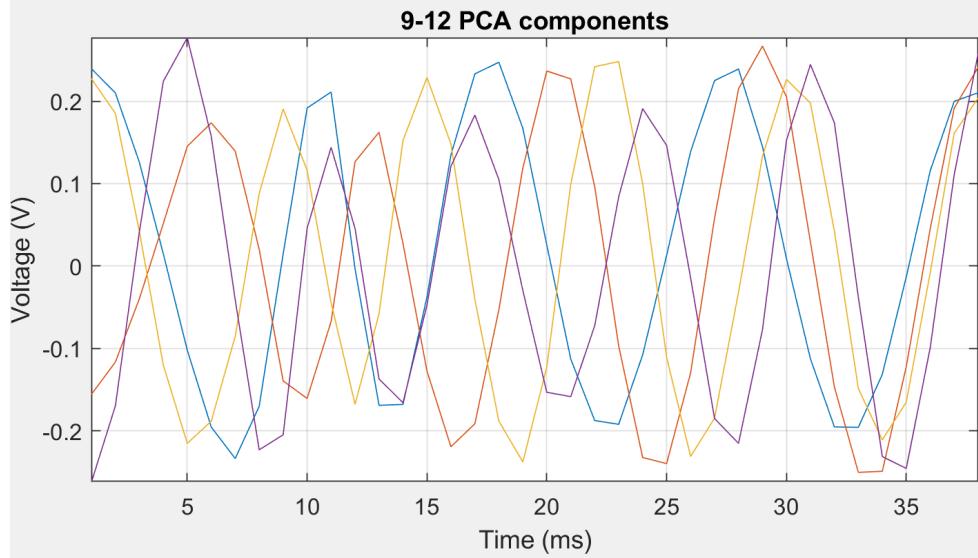
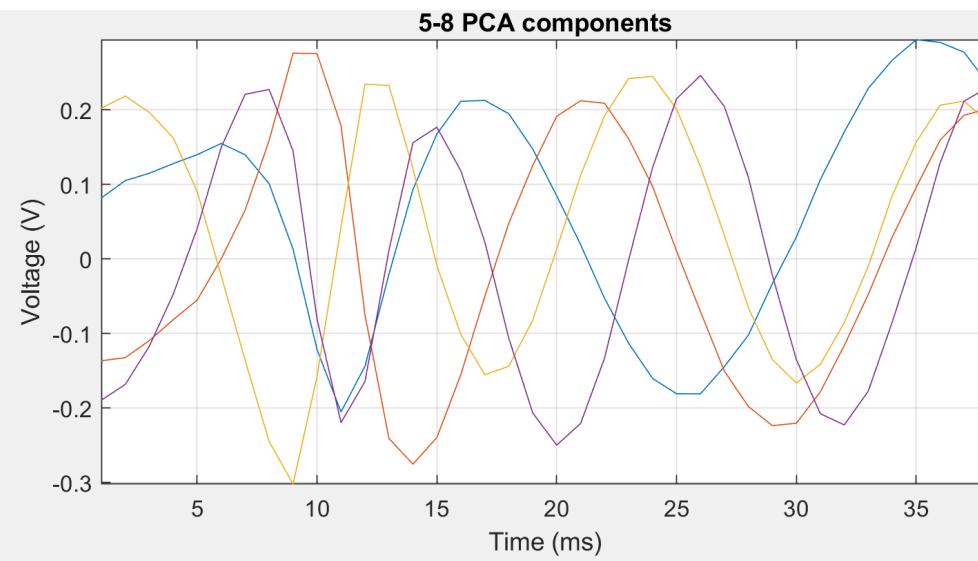
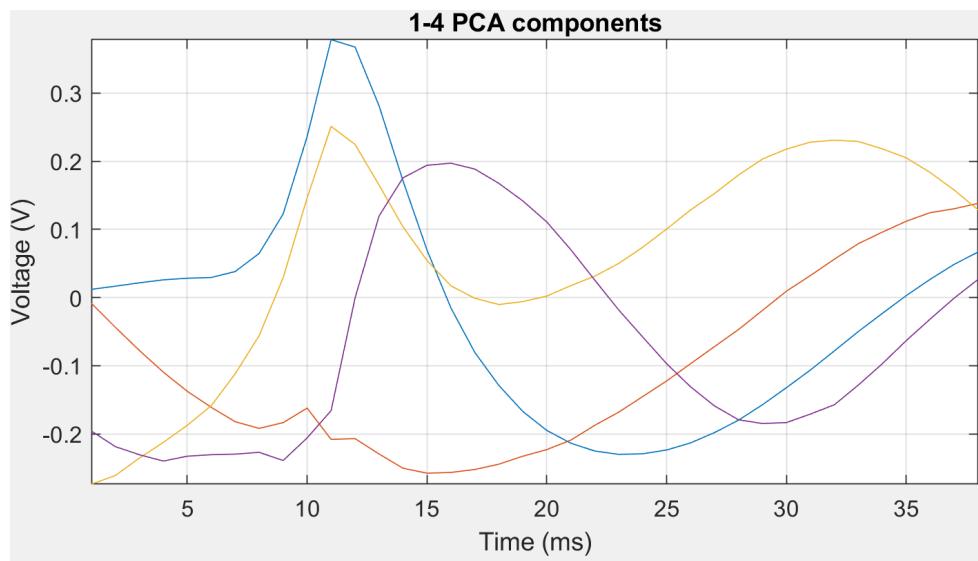


Figure 4a: in question 14, we plotted eigenvectors in groups of 4 from our post-svd “shapes” matrix, A. Eigenvectors from columns 5-16 appeared like noise, since there was not much distinct information being communicated. It was more difficult to tell if we would need eigenvectors 5-8 to adequately reproduce the snippet data. The first four principal components appeared dissimilar enough to explain the snippet data variability and will be the focus of the subsequent scatter plots.

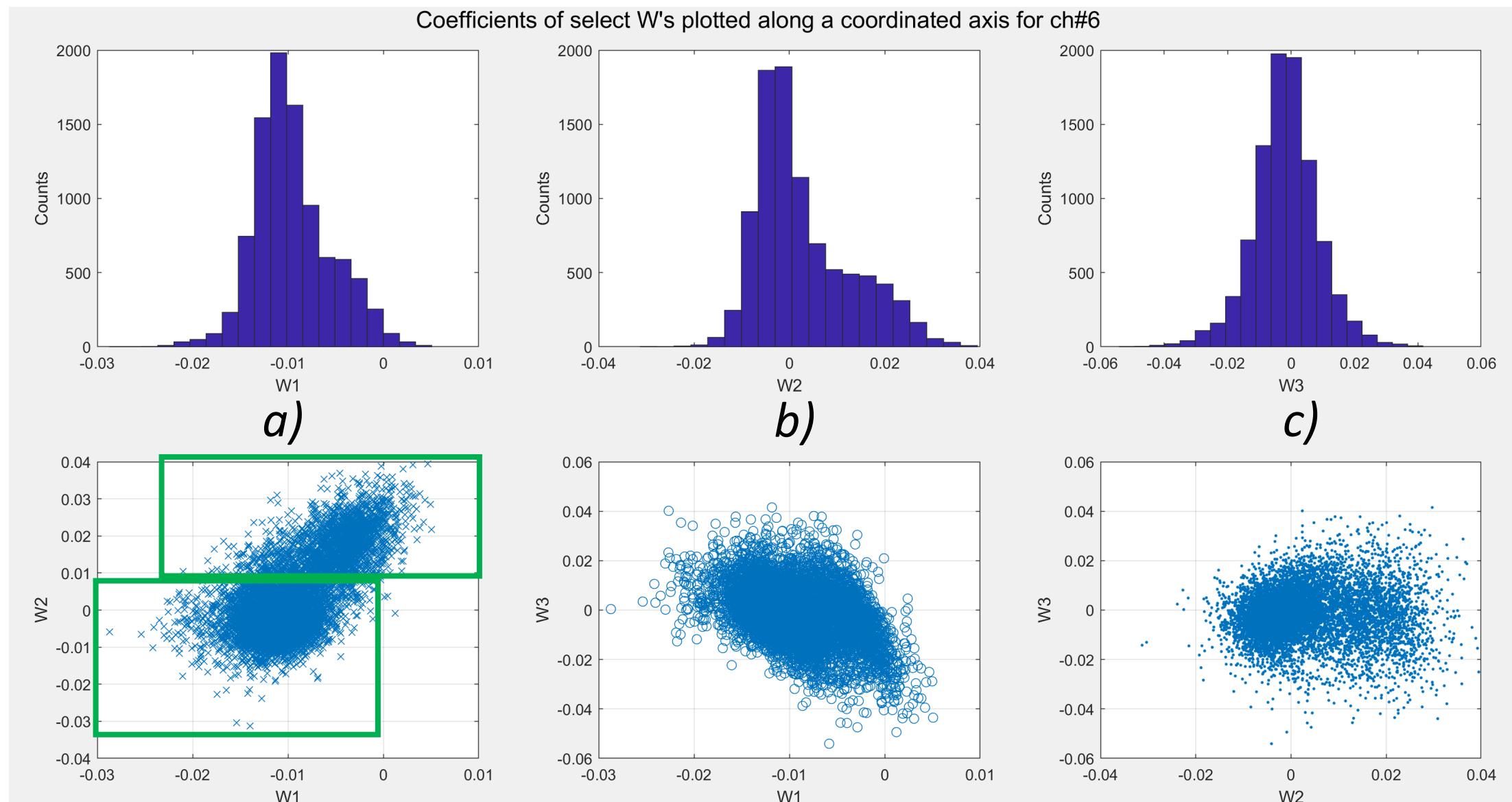
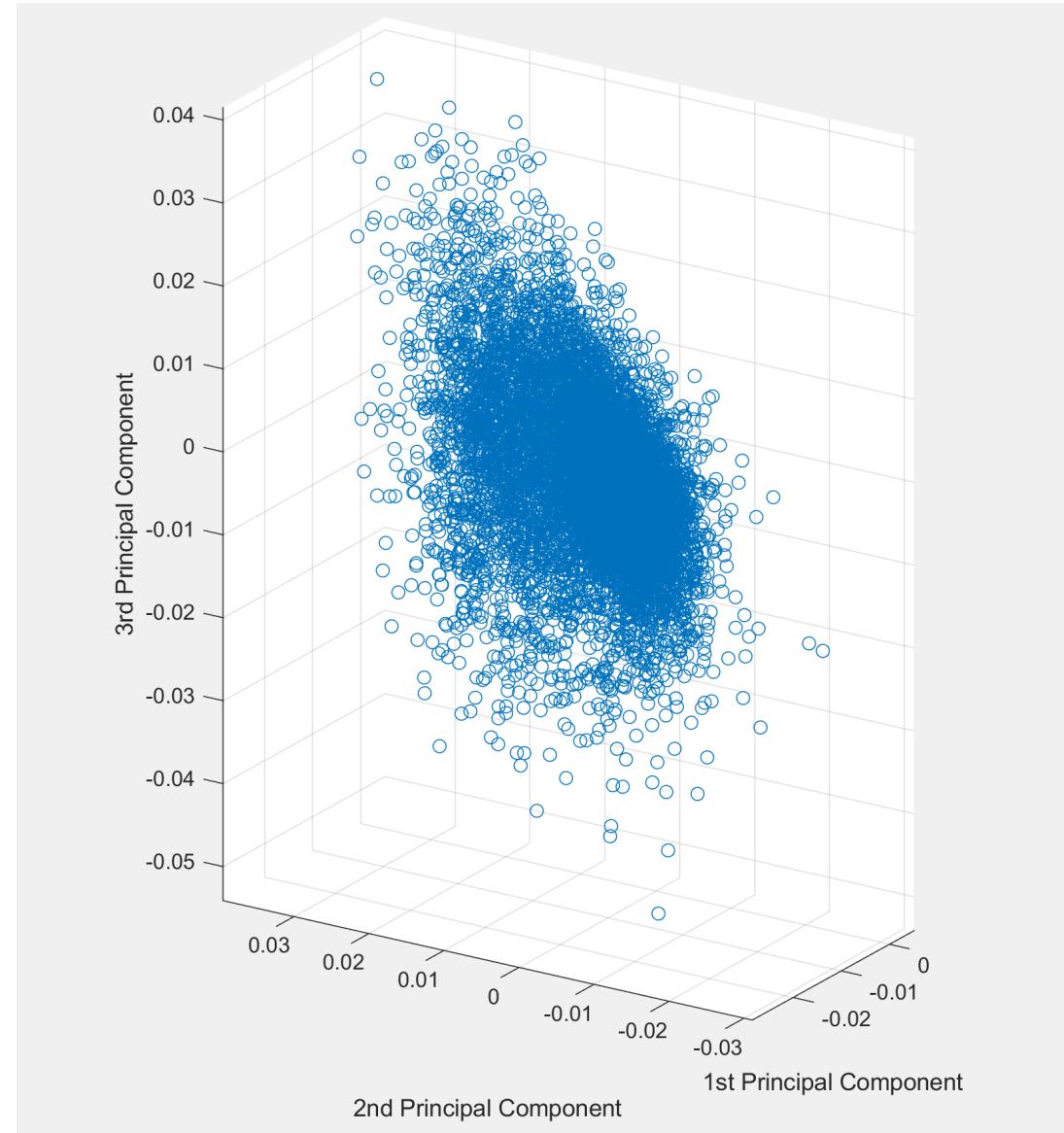


Figure 5a: in question 15, the amplitudes of the first 3 PCA eigenvectors were scatter plotted against each other pair-wise to determine how to cluster the data. Clustering occurs in our model's scaling/coefficient space, so we used the matrix W from the singular value decomposition (svd)-PCA step. There appear to be two distinguishable clusters in a), but the proximity between clusters might lead to misclassification errors. Even c) looks like there are two candidate clusters, indicating interesting, distinct activity.

Figure 5b (Question 15): the three most important principal components were visualized in a 3D scatter plot for channel 6. There was no obvious clustering seen.



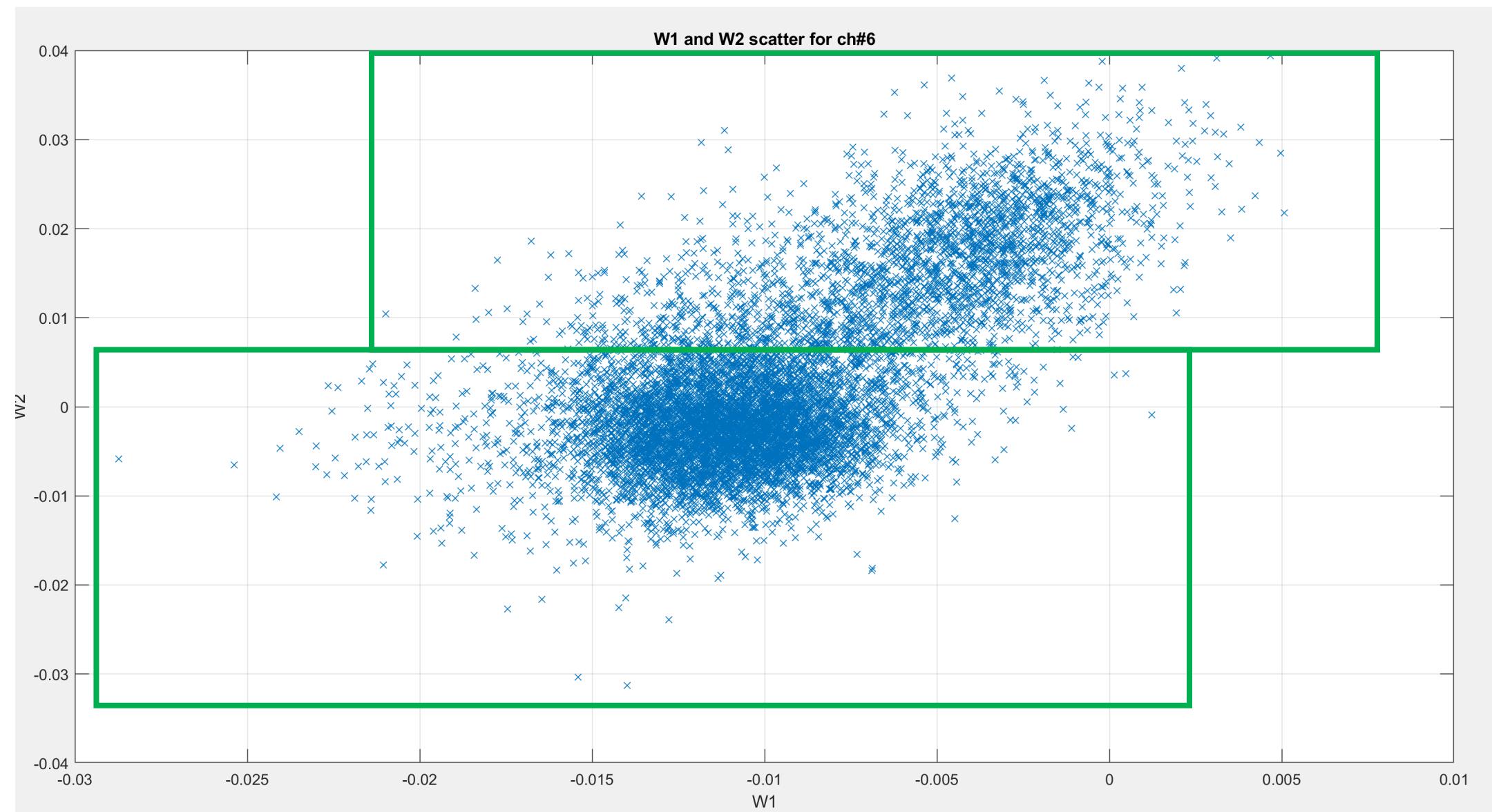


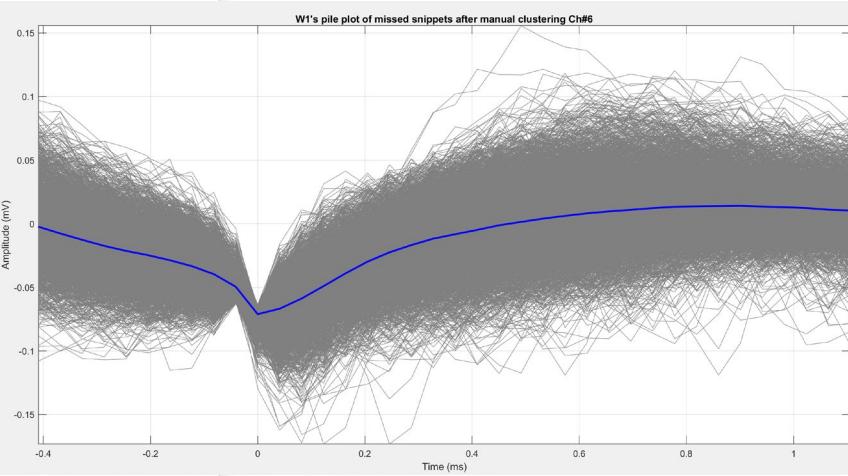
Figure 5c: From a visual inspection of the scatter plot between the coefficients of the first two principal components, I estimated:

Cluster 1: $\epsilon_{11} = -0.025 < W(:,1) < \epsilon_{12} = 0.025 \text{ && } -0.035 < W(:,2) < \epsilon_{22} = 0.005$

Cluster 2: $\epsilon_{21} = -0.022 < W(:,1) < \epsilon_{12} = 0.0055 \text{ && } 0.008 < W(:,2) < \epsilon_{22} = 0.04$

- Figure 6: (in question 16) from Figure 5c's estimated epsilon parameters, a new logical conditional in the MATLAB code indexed cluster space points from the principal components outside of the hand-placed rectangular clustering. That accidentally allowed the separated waveforms to be plotted.

a)



b)

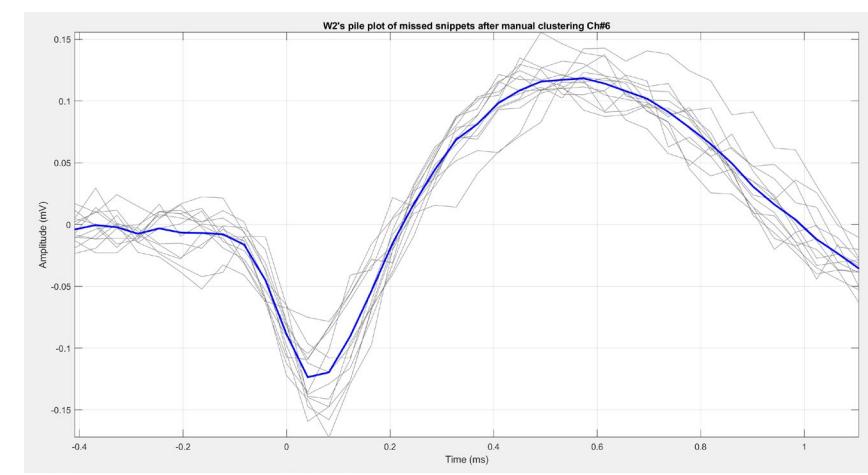
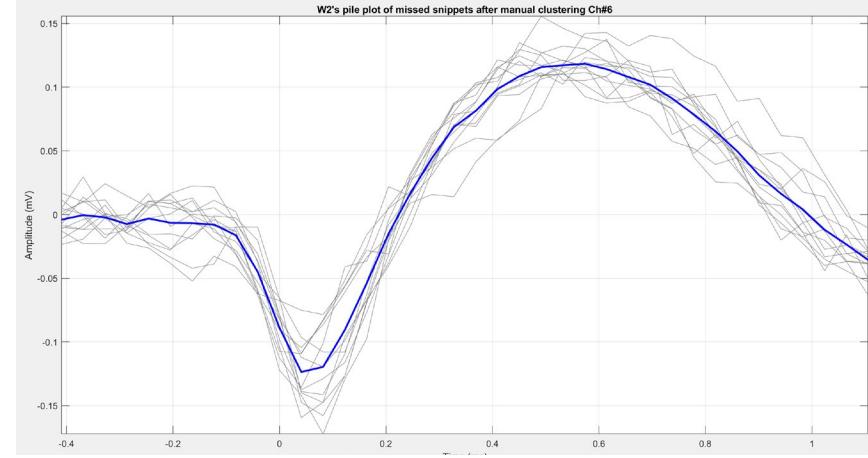
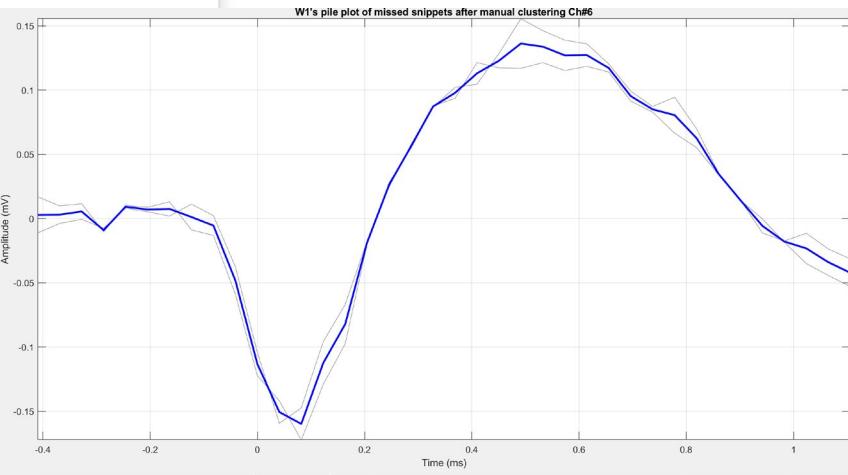
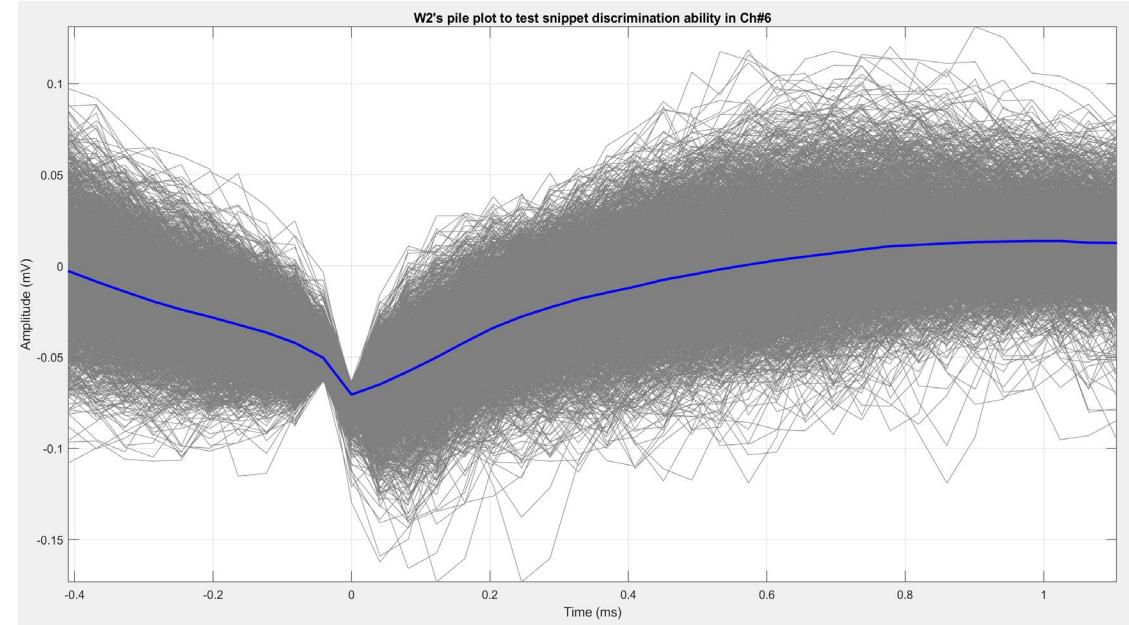
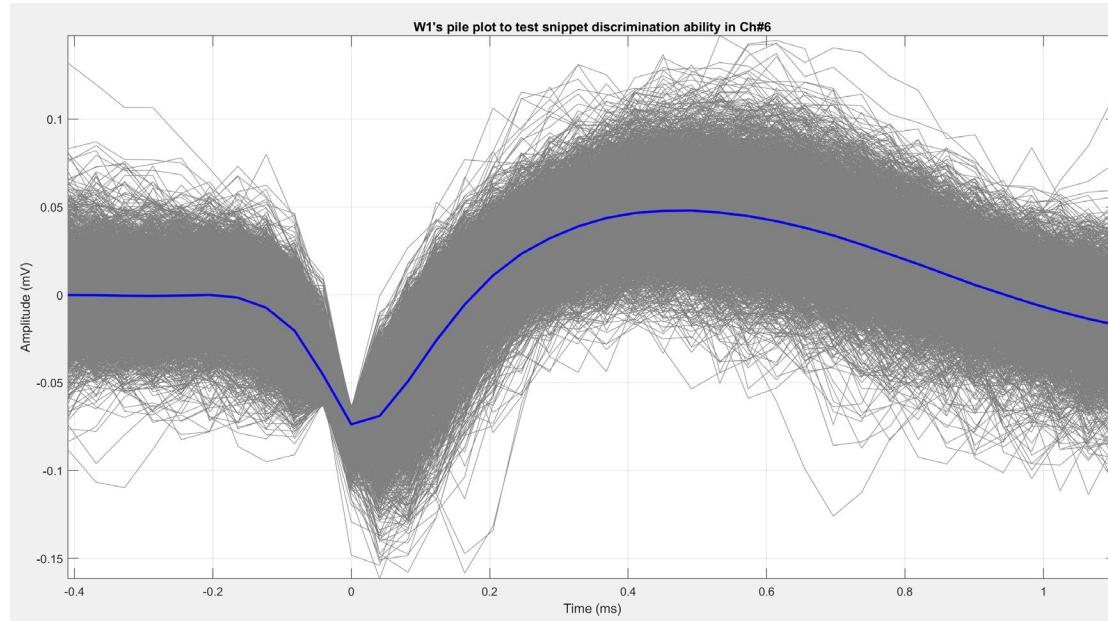


Figure 7: in question 17, using the hand drawn clustering windows determined from Figure 5c, two pile plots were made. Based on the gray snippets with negative depolarization peaks just before 0.2 ms, this method appears to be clustering distinct waveforms.



Although two distinct waveforms appear to be present by these pile plots, it's possible that the manual clustering algorithm based on the epsilon bounds I defined permitted too many non-similar waveforms into my cluster. Therefore, it may be prudent to plot more coefficients.

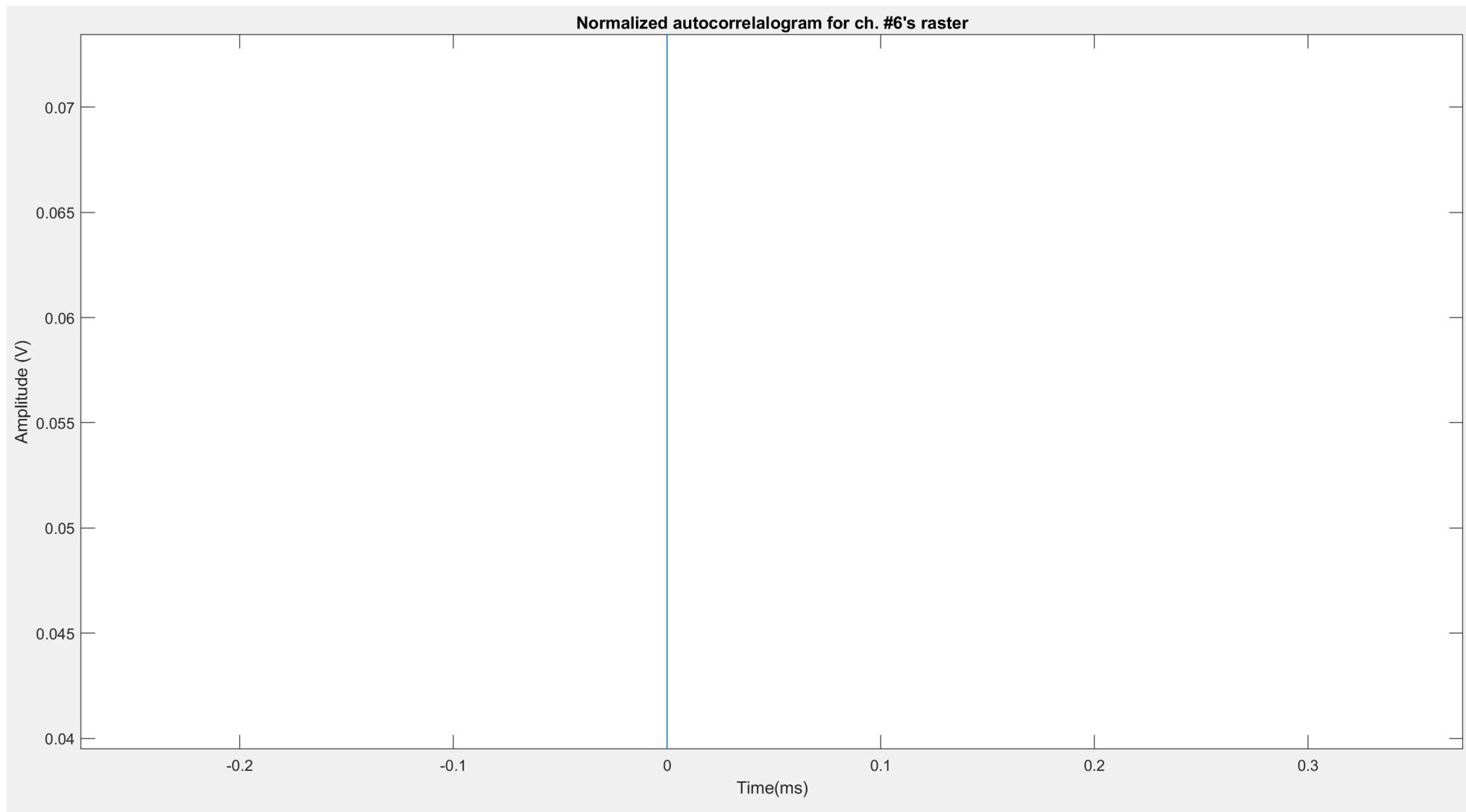


Figure 8: The autocorrelalogram is a great way to determine if discriminated spikes correspond to a single neuron or not. With the plot above, there is a sharp peak around 0, which is what is expected as single neuron activity. If there were multiple units, I would expect more ringing on the non-lag times that were non-physiological (>100 Hz activity).

Autocorrelalogram of the discriminated/clustered spikes for ch. #6

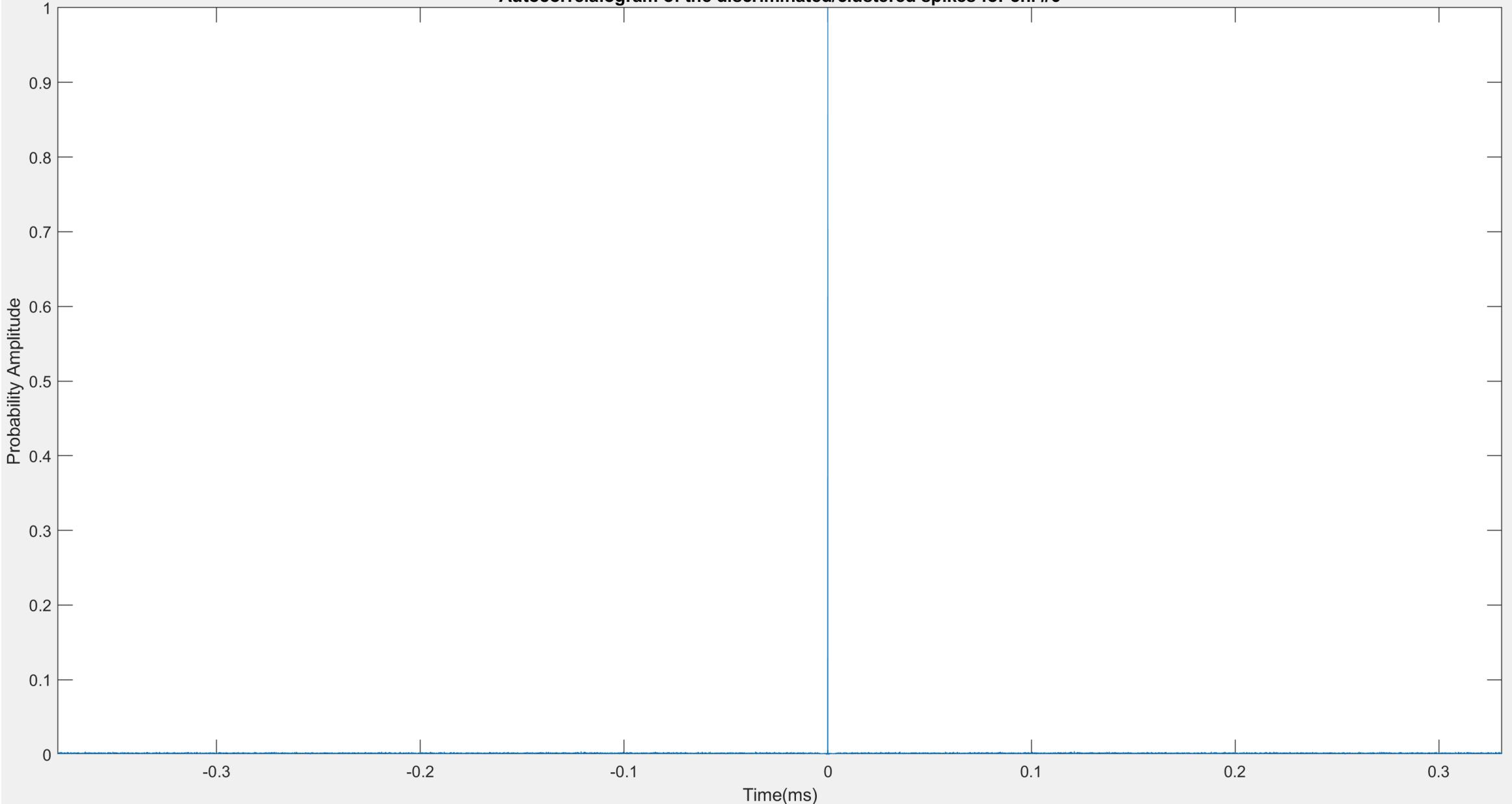
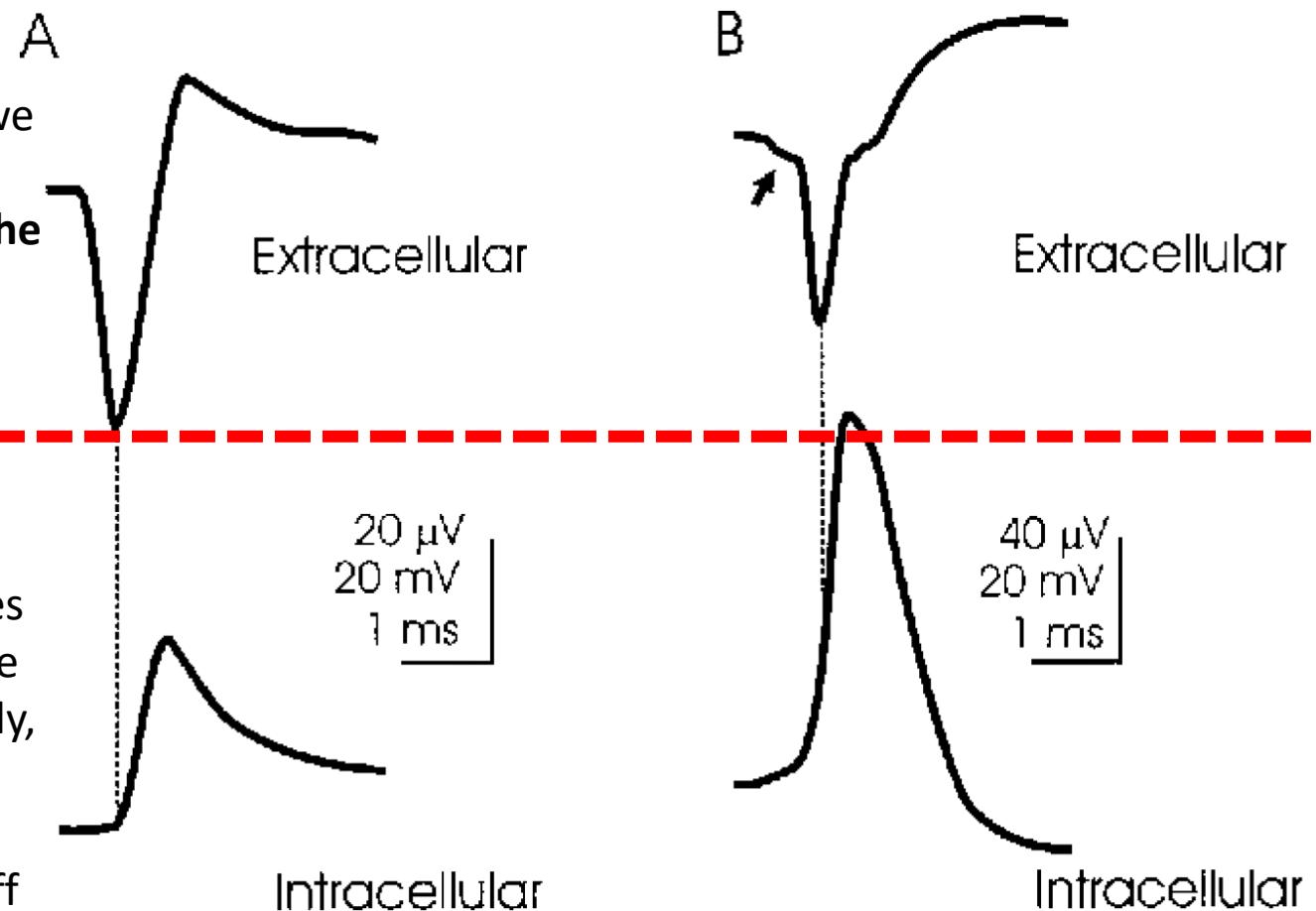


Figure 8 (question 19): Just as in the prior graph, we saw a sharp peak that dominates over the terms going away on either side of the lag time of 0; again, there is a sharp peak around 0, which is what is expected as single neuron activity.

Q20

In the biophysics of electrophysiology lecture, we learned that action potentials look different depending on the location of the electrode. **If the electrode is extracellularly located, then the depolarization is negative.**

The “stereotypical” action potentials that we are used to in textbooks and introductory bio lectures have a positively orientated depolarization. These are from electrodes recording within the cell body, so intracellularly. In general, the cytoplasm is negatively charged at rest and the extracellular environment is positive. Depolarizations shoot off in the opposite direction to the starting state, e.g., if extracellular is positive, depolarization is negative, which is seen in the red box above



Henze et al *J Neurophys* 2000

Q21 How can you help isolate spikes prior to data acquisition? How would you isolate individual spikes from the snippet data already acquired?

- Prior to data acquisition...ensure that we tailor our setup to maximize spike isolation from noise. This means that we choose an electrode that maximizes capturing single neuron activity from a cell body or axon. Our choice of electrode is important, since degradation will impact spike recording fidelity. Thus, choosing the right conductor and insulator is important, as well as if it is laminar/linear or a tetrode. Multi-channel recording along our electrode means we can gain shared information that we can build into our clustering model. In a sense, each channel may provide more evidence that our cluster modeling is correct.
- Post-data acquisition...extract ~1.2ms waveform snippets based on threshold crossings from the spike data stream. The 1.2 ms is from physiological constraints of action potential firing rates. The threshold should be selected to avoid admitting false positives as much as possible. Isolating single units can be accomplished by centering the data around the mean/median ($\text{data}-\text{mean}(\text{data})$) and performing singular value decomposition (svd) to get the basis eigenvectors that describe as much snippet data variance as possible.

Q22 How could you use PCA analysis to determine whether there are individual neurons that only respond to the visual stimulus?

- Align raster data to the onset of a visual stimulus, e.g. `time(0) = presentation of stimulus.` (use `xline(0,'—','stimulus onset')` to have a persistent vertical line)
- Compute the spike density per time bin, which is plotted as a peri-stimulus time **histogram**. If you see consistent neural activity after the stimulus presentation time, then you can be reasonably confident that there is a causal relationship between the stimulus and neural activity.
- To confirm if certain neurons in your raw neural data correspond to visually-sensitive ones, you can color-code the neurons you clustered and apply it to the raster to tease them apart

Q23 How would you exclude snippets in a cluster that are weakly associated with the cluster?

I would exclude snippets that are outside the tolerance range for the centroid (the mean or “the center of gravity”) of my cluster candidates.

If you take the square of the distances between cdata points and your candidate centroids, then just eliminate the ones at ridiculously large squared distances

There are several distance candidates we can use, including the squared Euclidean distance, the cosine (since the dot product of the two vectors gives an idea of collinearity). For binary matrices, the hamming distance may be suitable, since the hamming metric looks for changes in bit proportions

Q24 How does the autocorrelogram help us discriminate single units? What criteria is usually used for this purpose?

Auto-correlation = correlation of signal with itself; helps determine spike properties and distributions.

X-axis = time lags, and y-axis is the autocorrelation amplitudes. Normalizable.

Informs if the spike occurrence is physiologically realistic or not, e.g., if you see spiking activity within 5 ms of lag = 0, you probably need to re-cluster your raster

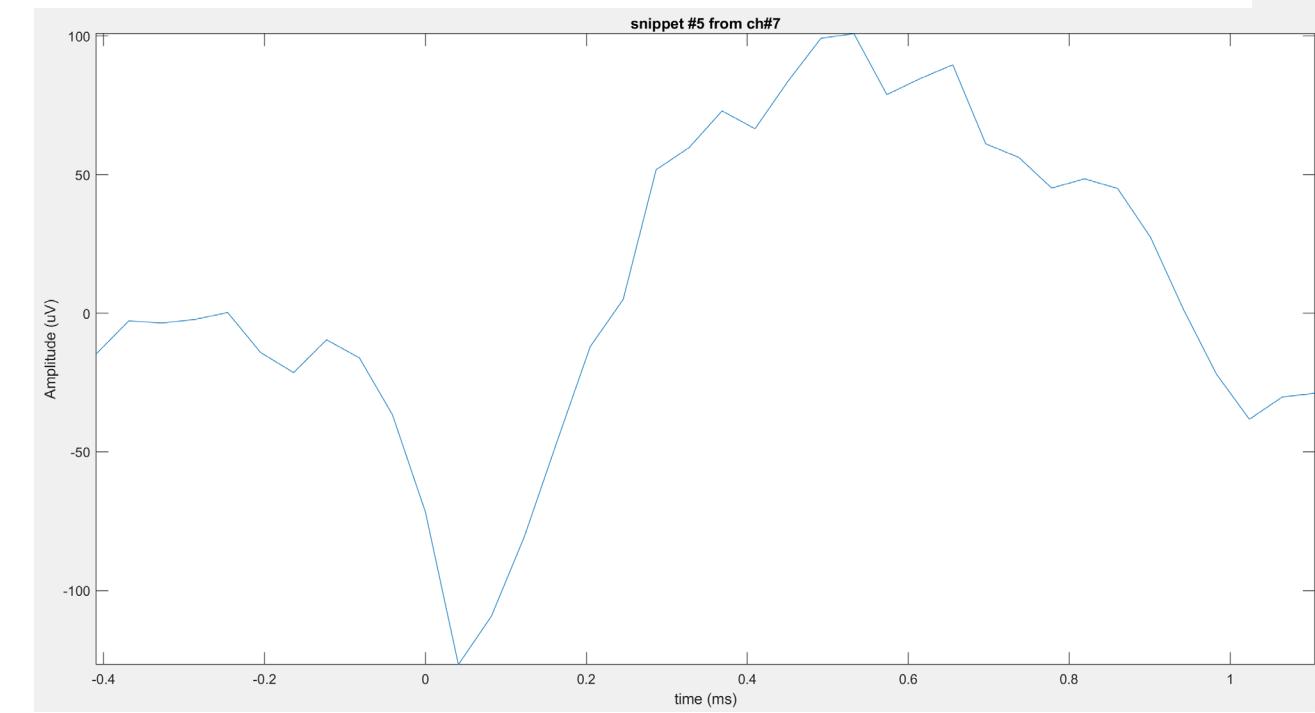
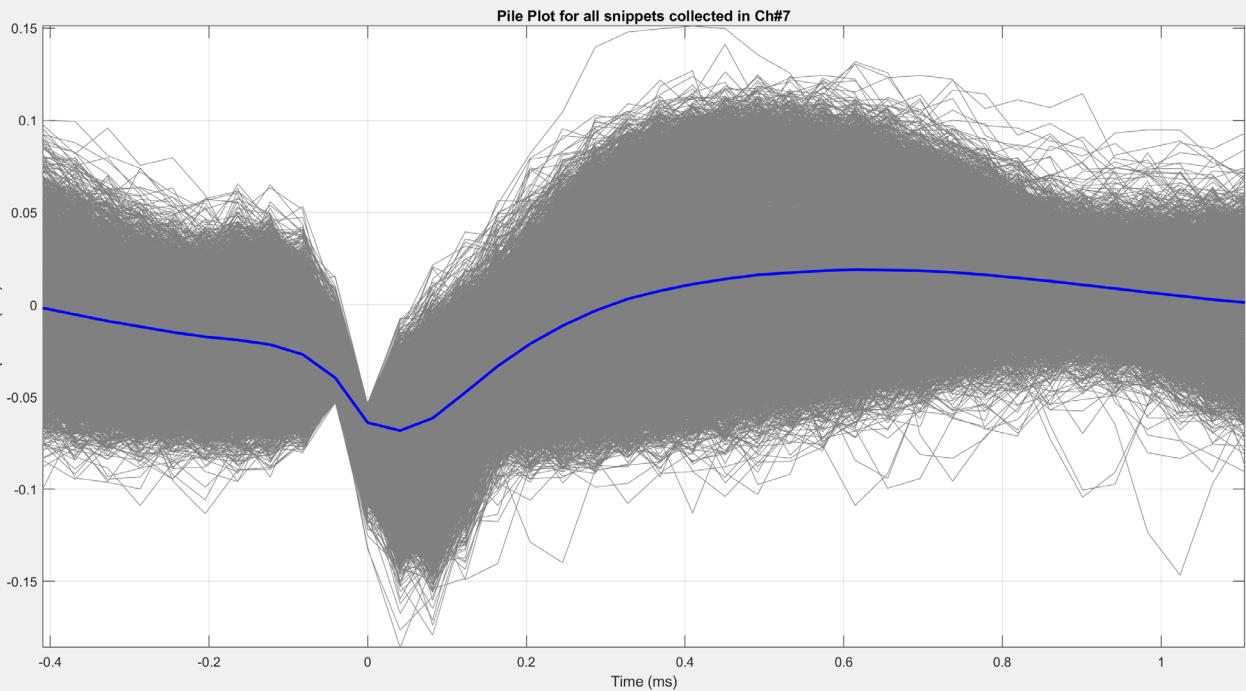
If you truly isolated a single neuron: perfect correlation at lag time 0 is expected, whereas there's little correlation when time shifted at any other point

If you see high correlations at lag times outside of 0, then you may be dealing with multiple neurons and not truly distinct waveforms from a single unit

Comparing a signal with itself with a delay, which helps us tell if the signal is discrete (single- or multi-unit)

Bonus: channel 7

Channel 7



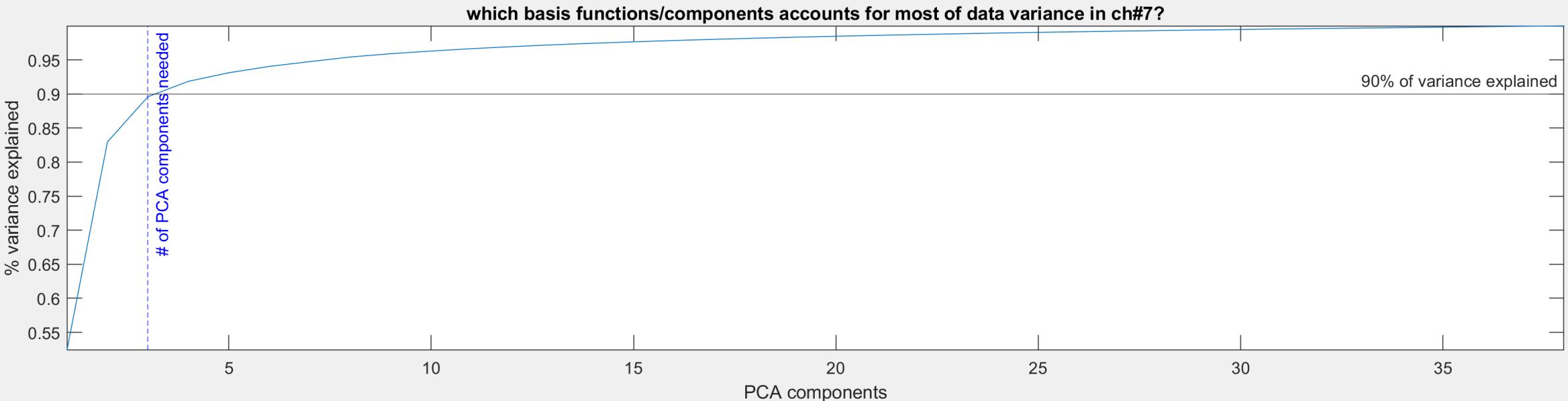
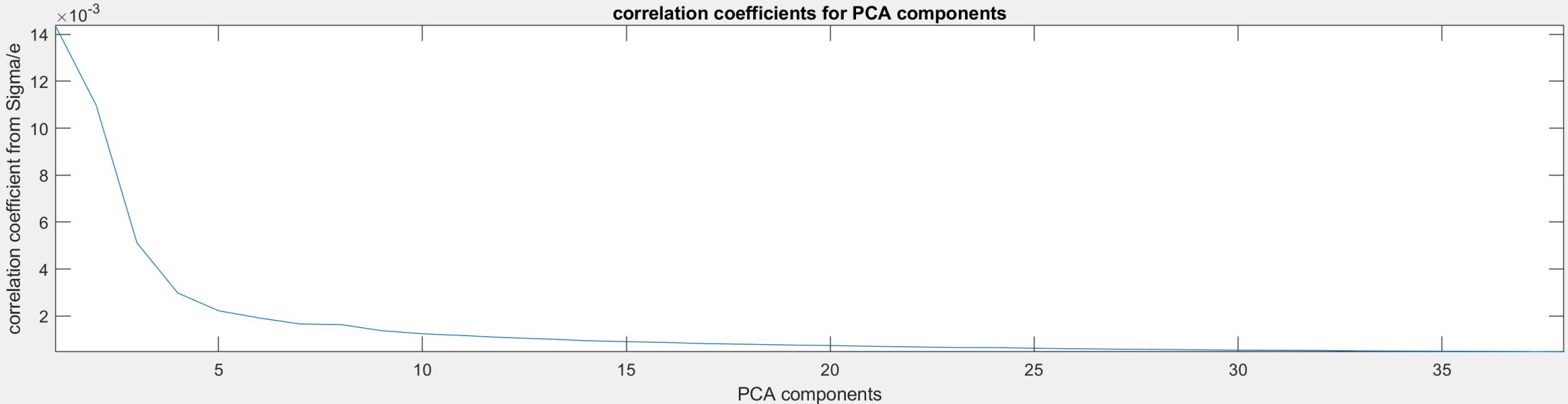


Figure: Repeated Question 12 for channel #7 and now only need 3 principal components to explain 90% of the variance

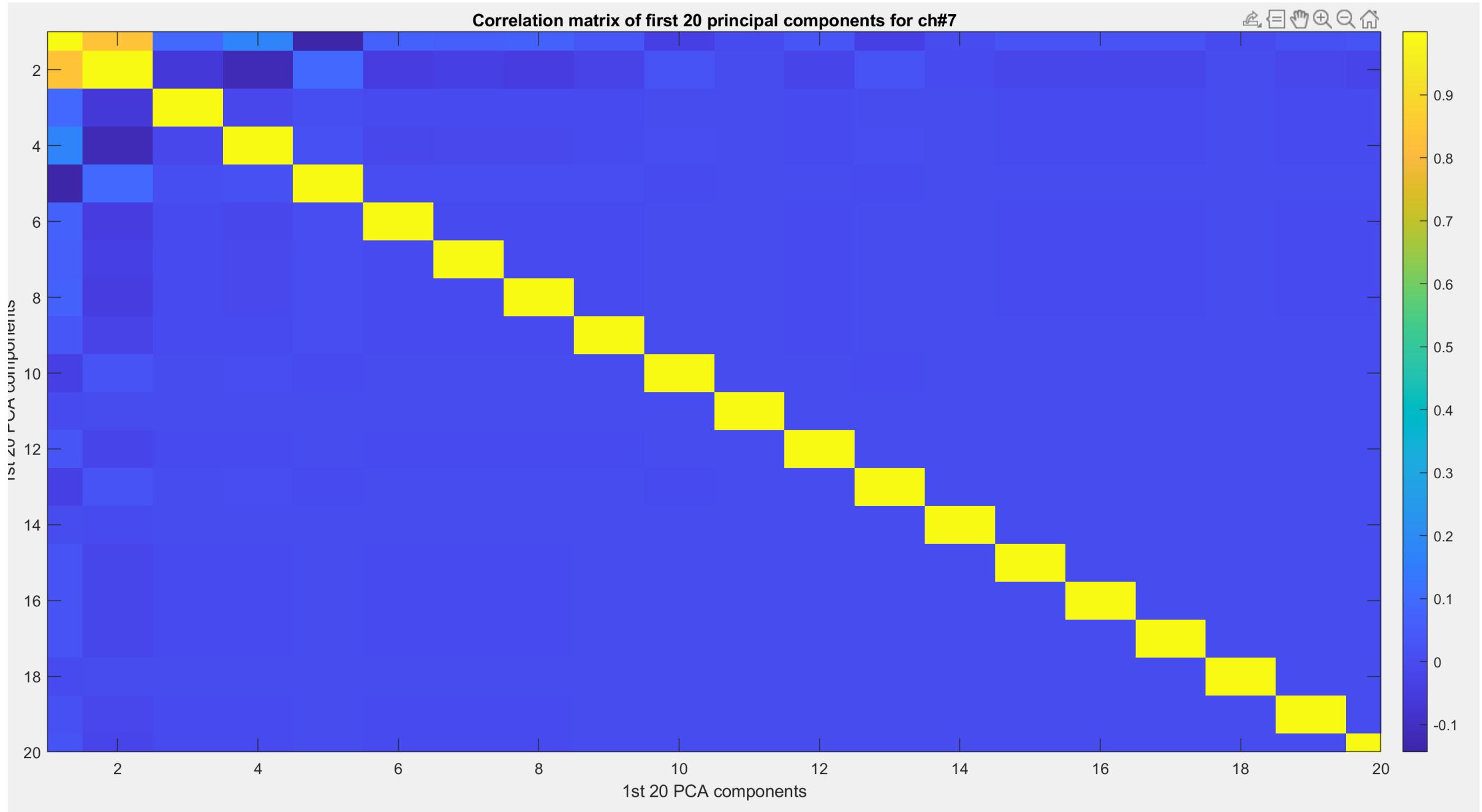


Figure: Repeated Question 12 for channel #7

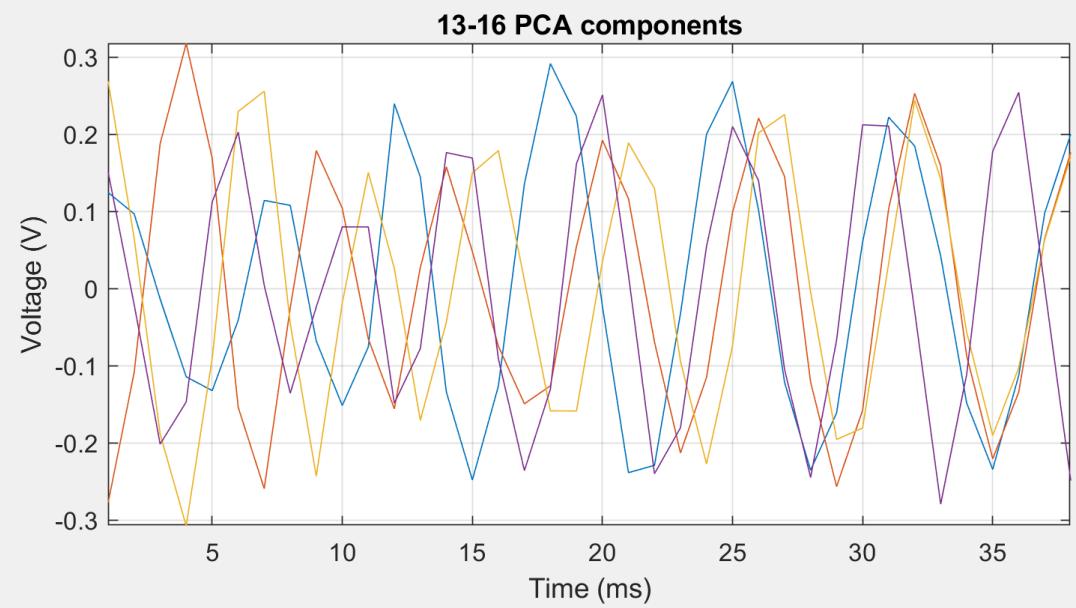
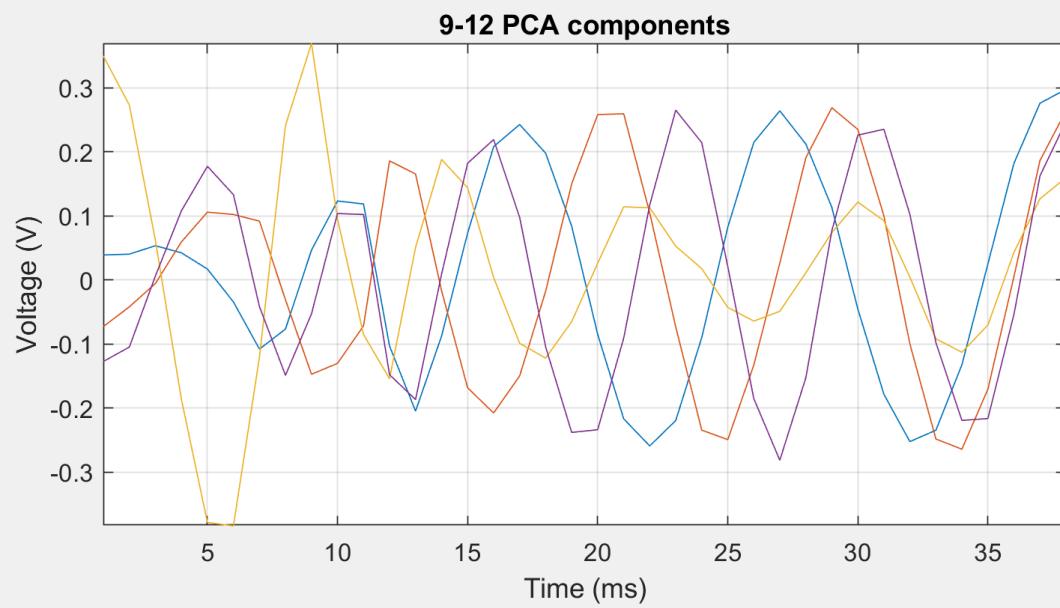
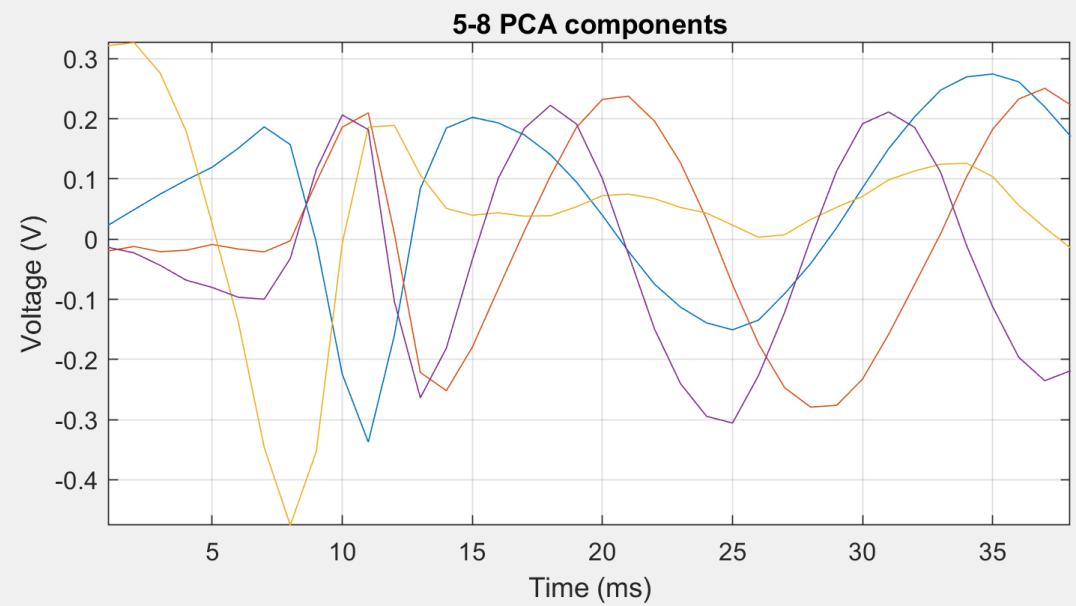
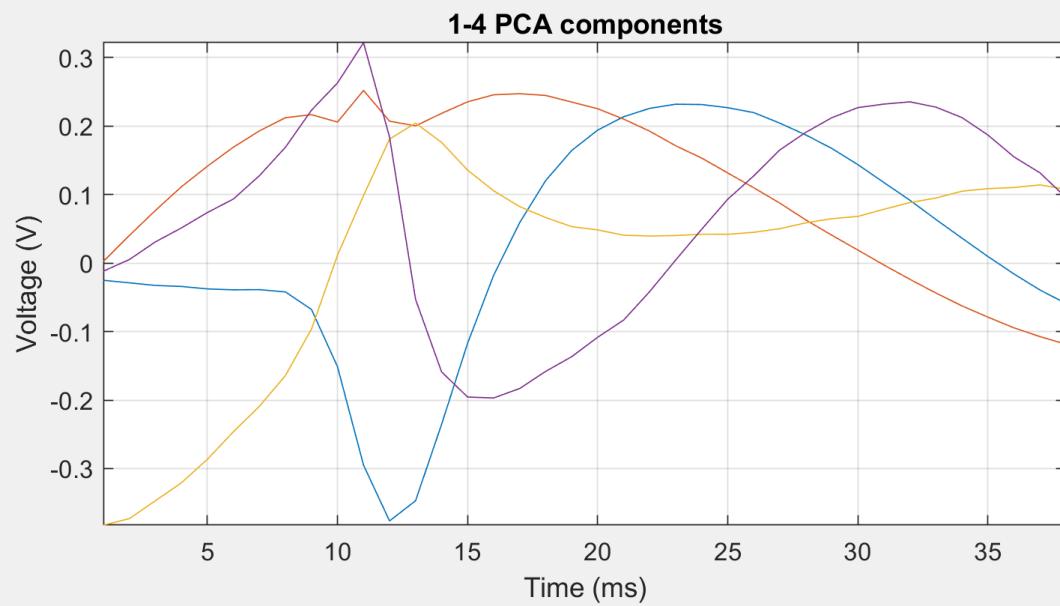


Figure: Repeated Question 14 for channel #7

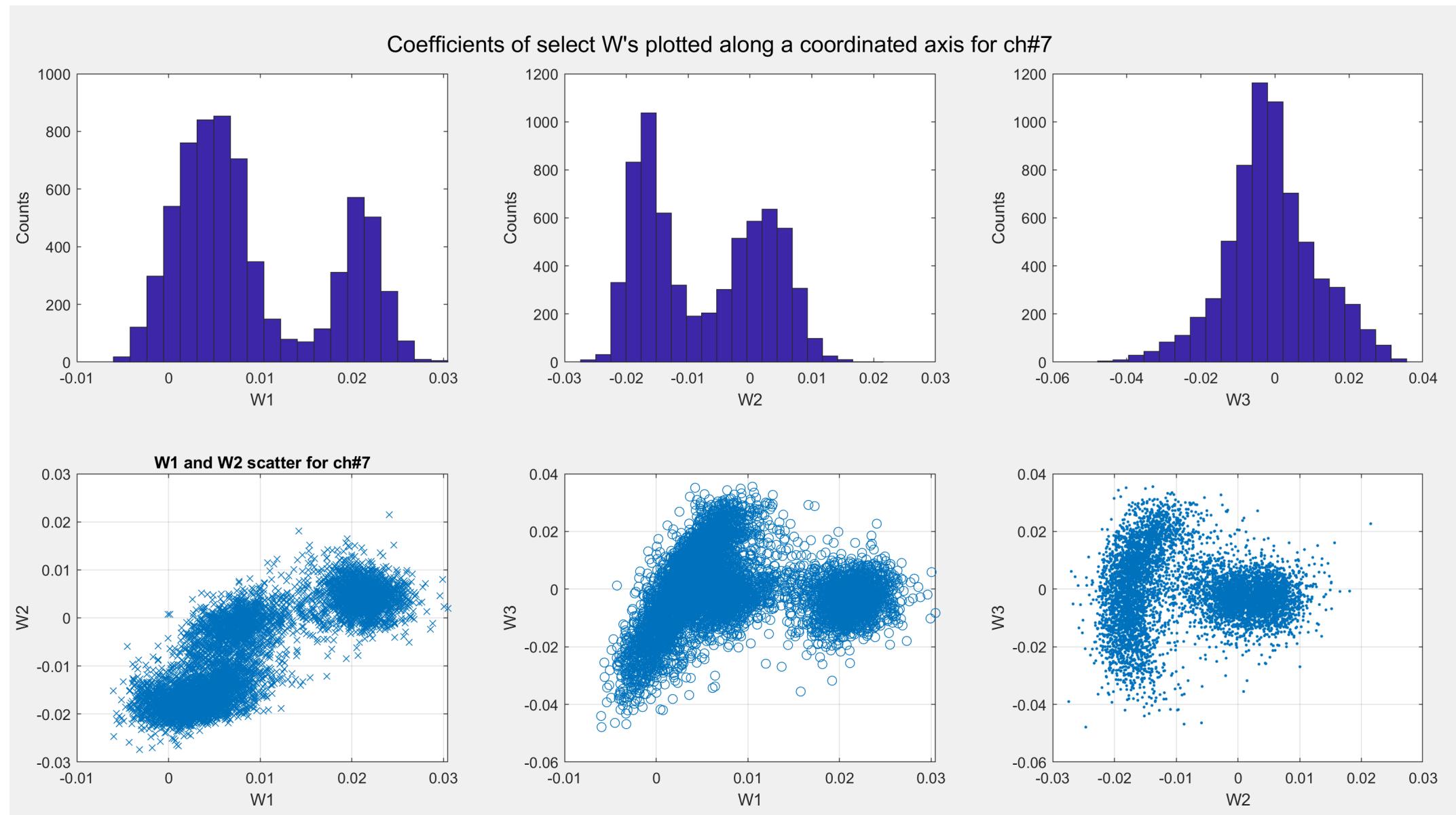
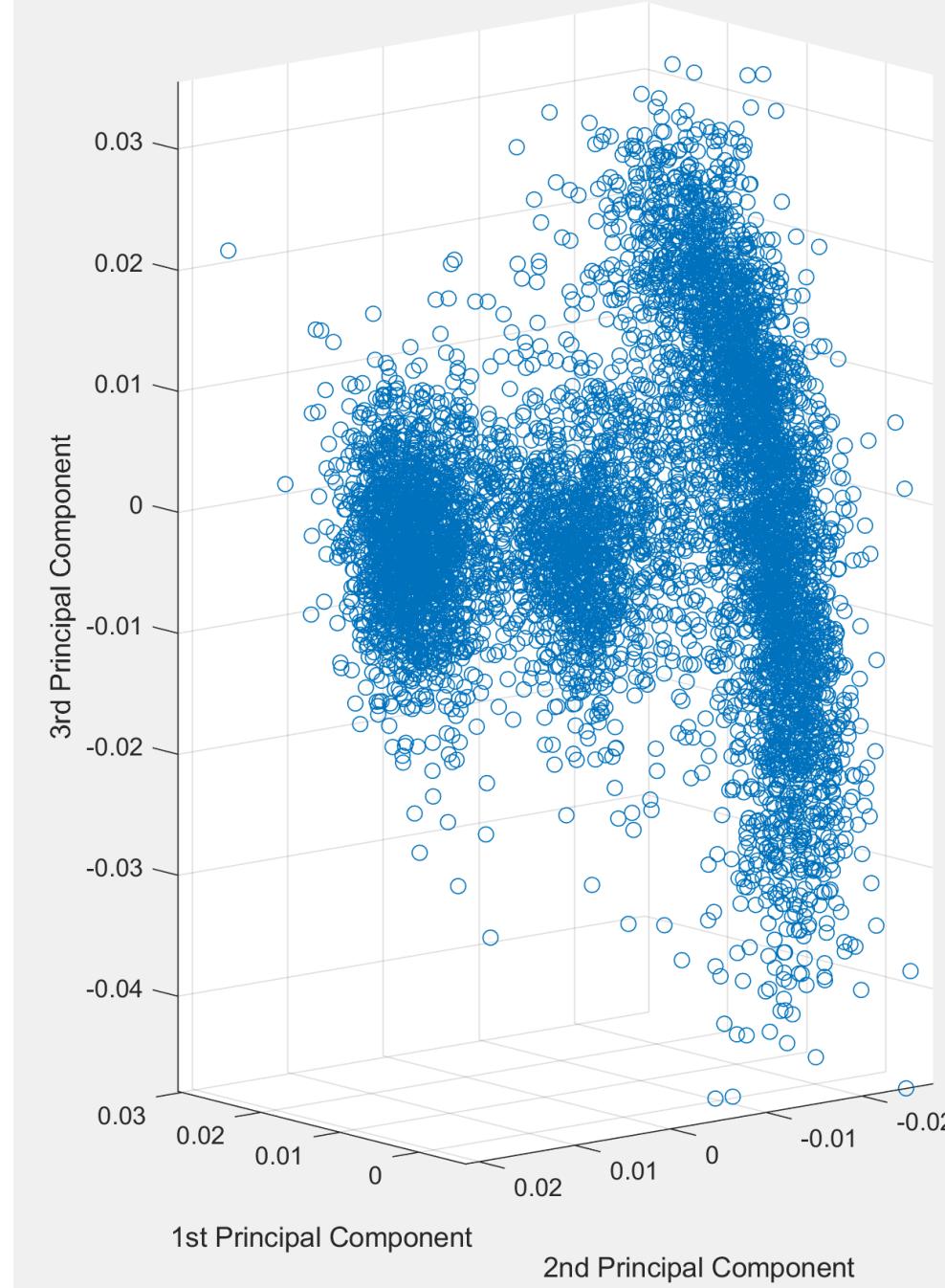
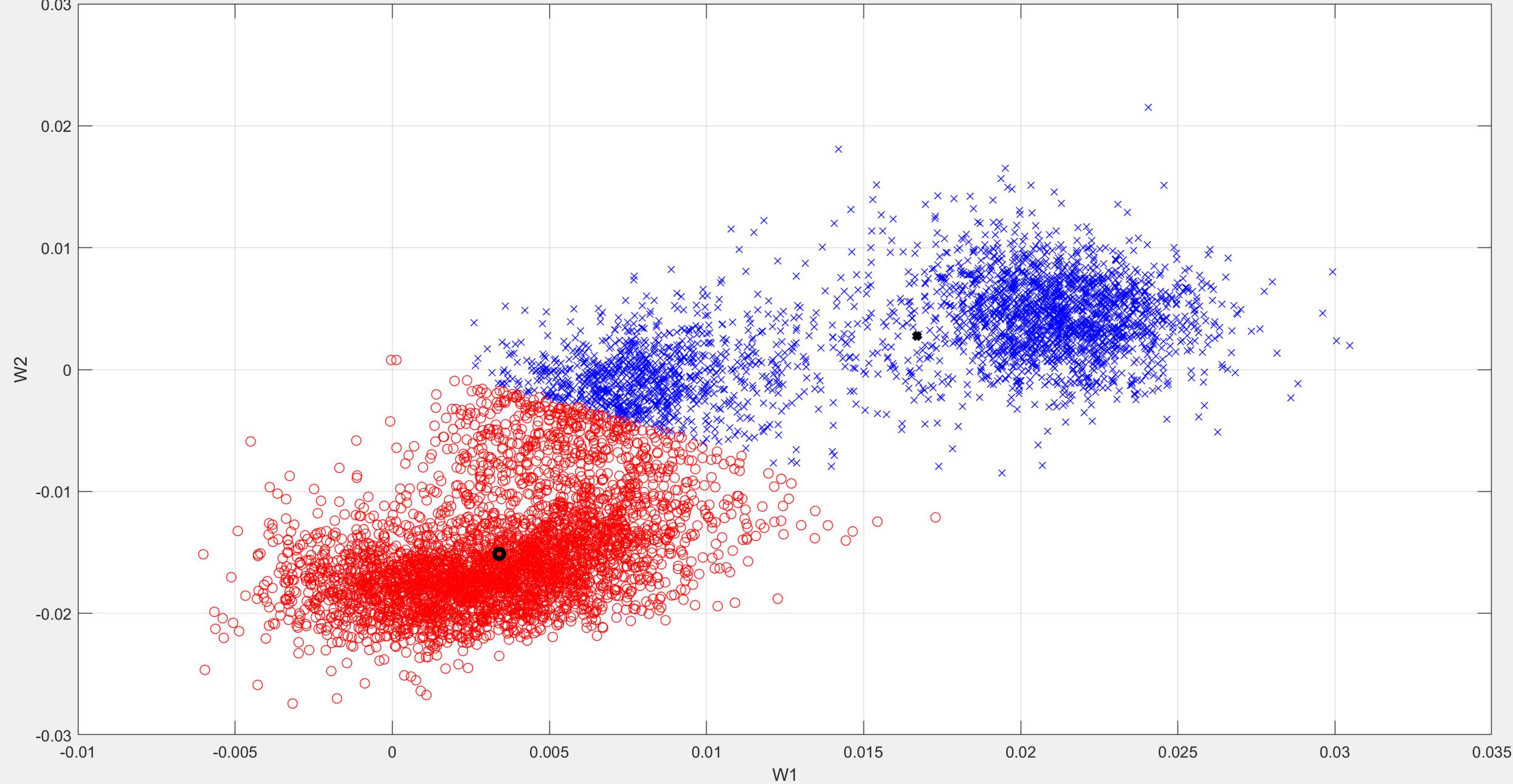


Figure: Repeated Question 15 for channel #7 and LOOK at this clustering!

Figure: Repeated Question
15 for channel #7 in a 3D
scatterplot



K-means Clustered PCA W1 and W2



Bonus: channel 8

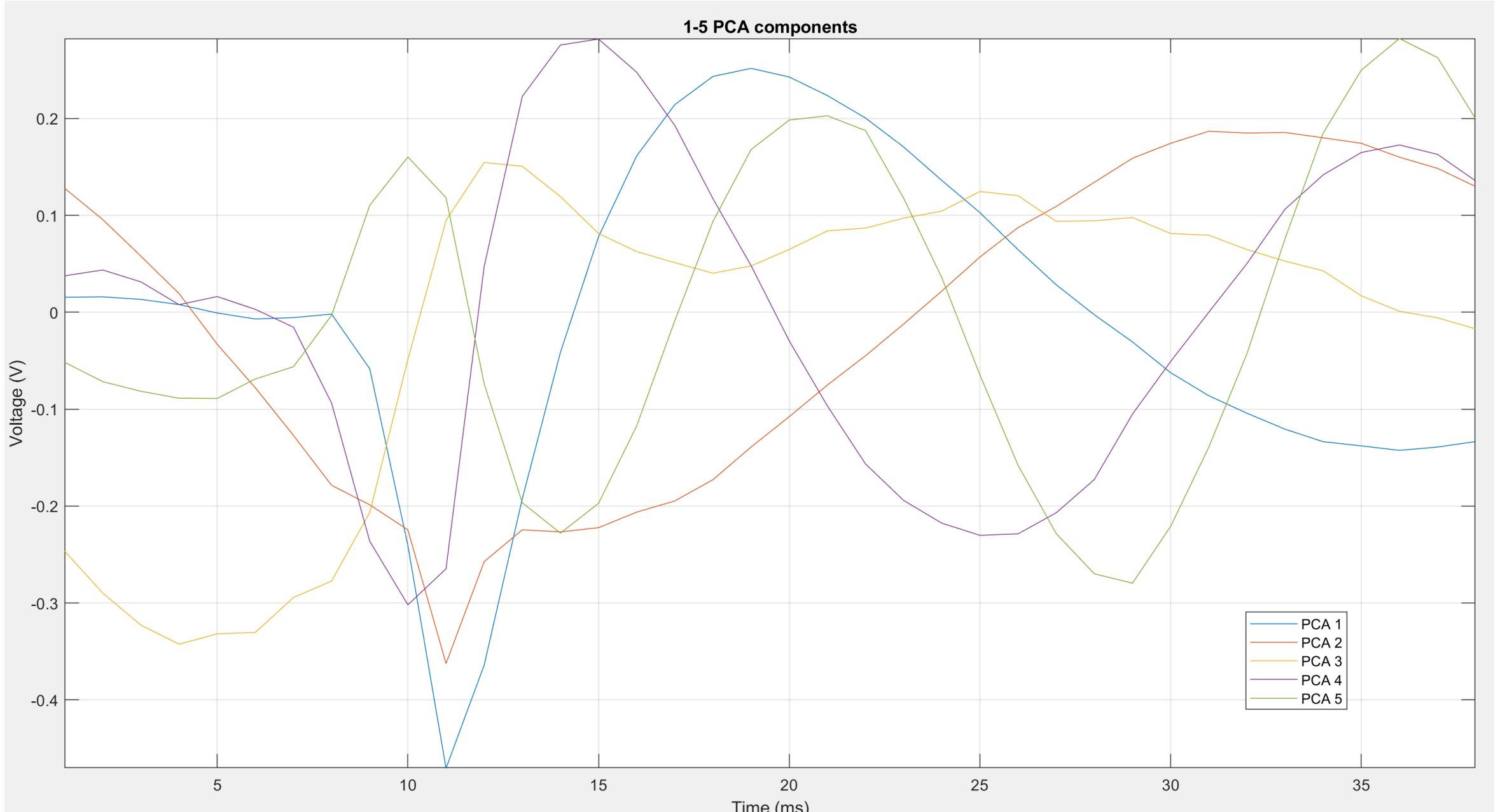


Figure 4b: First 5 eigenvectors from our “shapes” matrix, A, from channel 8. These appear to be distinct waveforms since the peaks and valleys do not align. Any coincident valleys or troughs, like with PCA components 1 and 2, have non-coinciding peaks.

1-6 PCA components

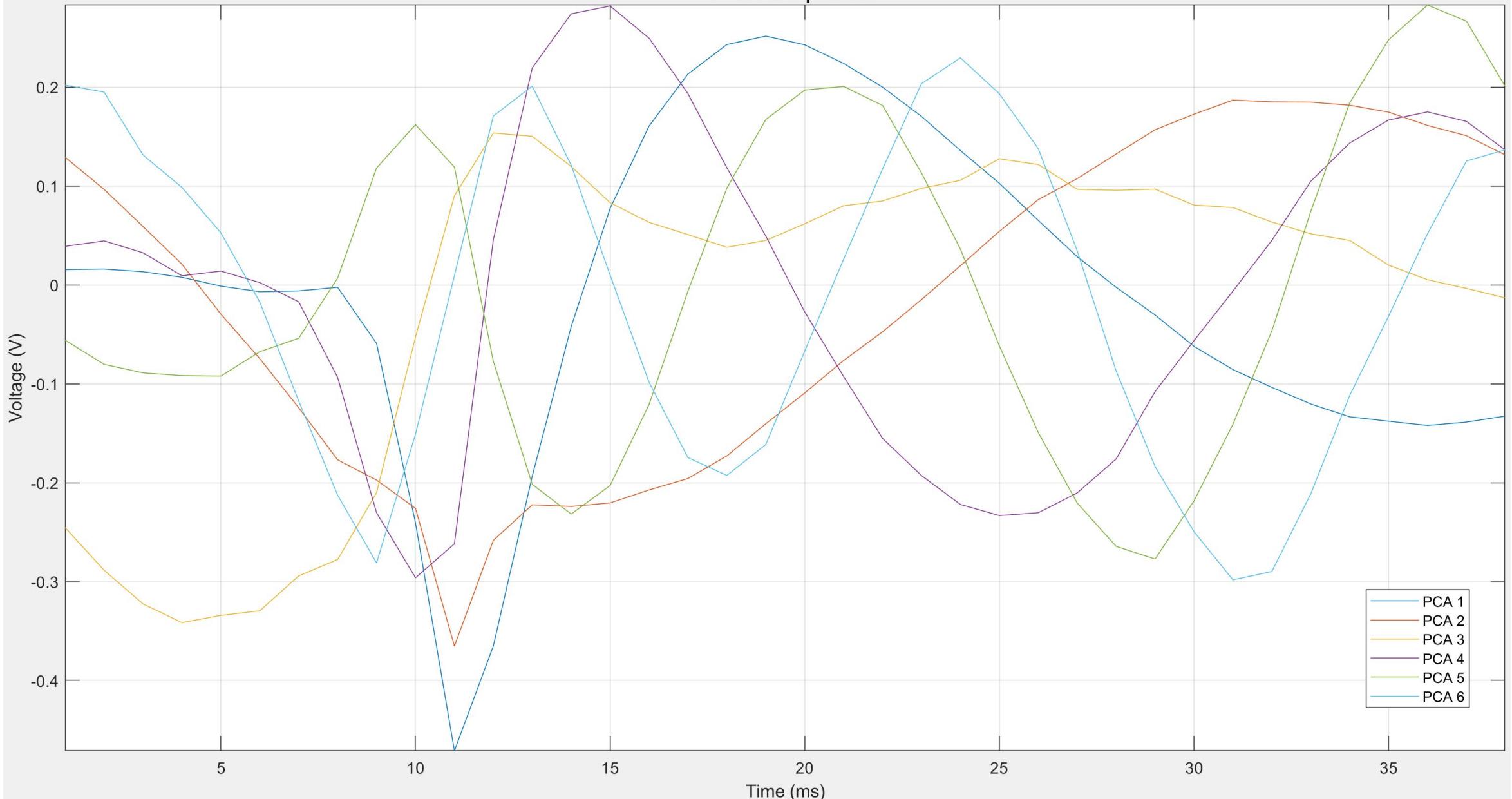


Figure 4c: First 6 eigenvectors from our “shapes” matrix, A, using channel 8. The sixth eigenvector (PCA 6) looked similar, but not coincident, with PCA component 5. The dissimilarity is evidence that

1-7 PCA components

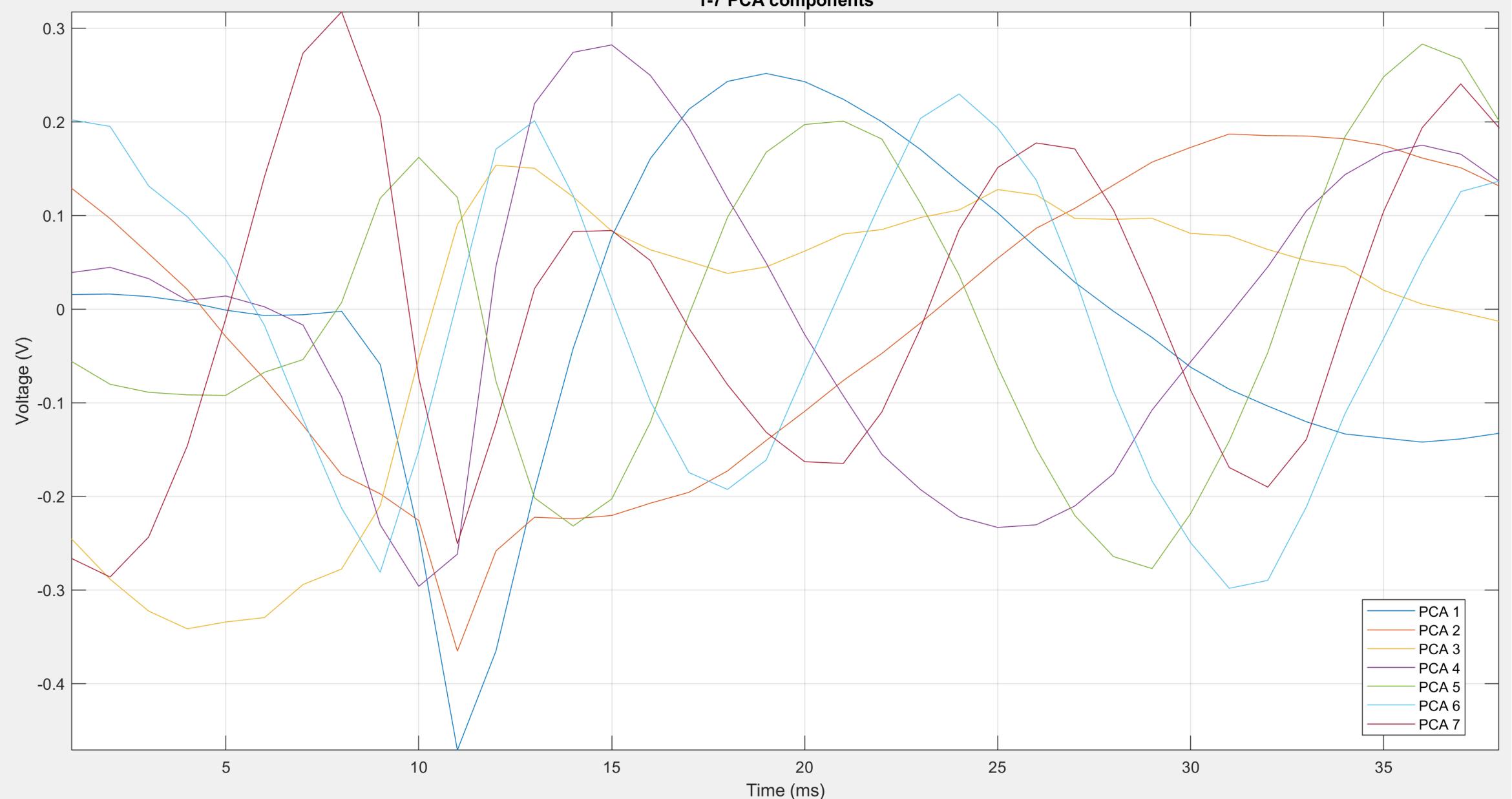


Figure 4d: First 7 eigenvectors from our “shapes” matrix, A. These also appear to be quite dissimilar, which provides even further evidence that perhaps the first seven PCA components are of significance.

Histograms of values for first 3 PCA components

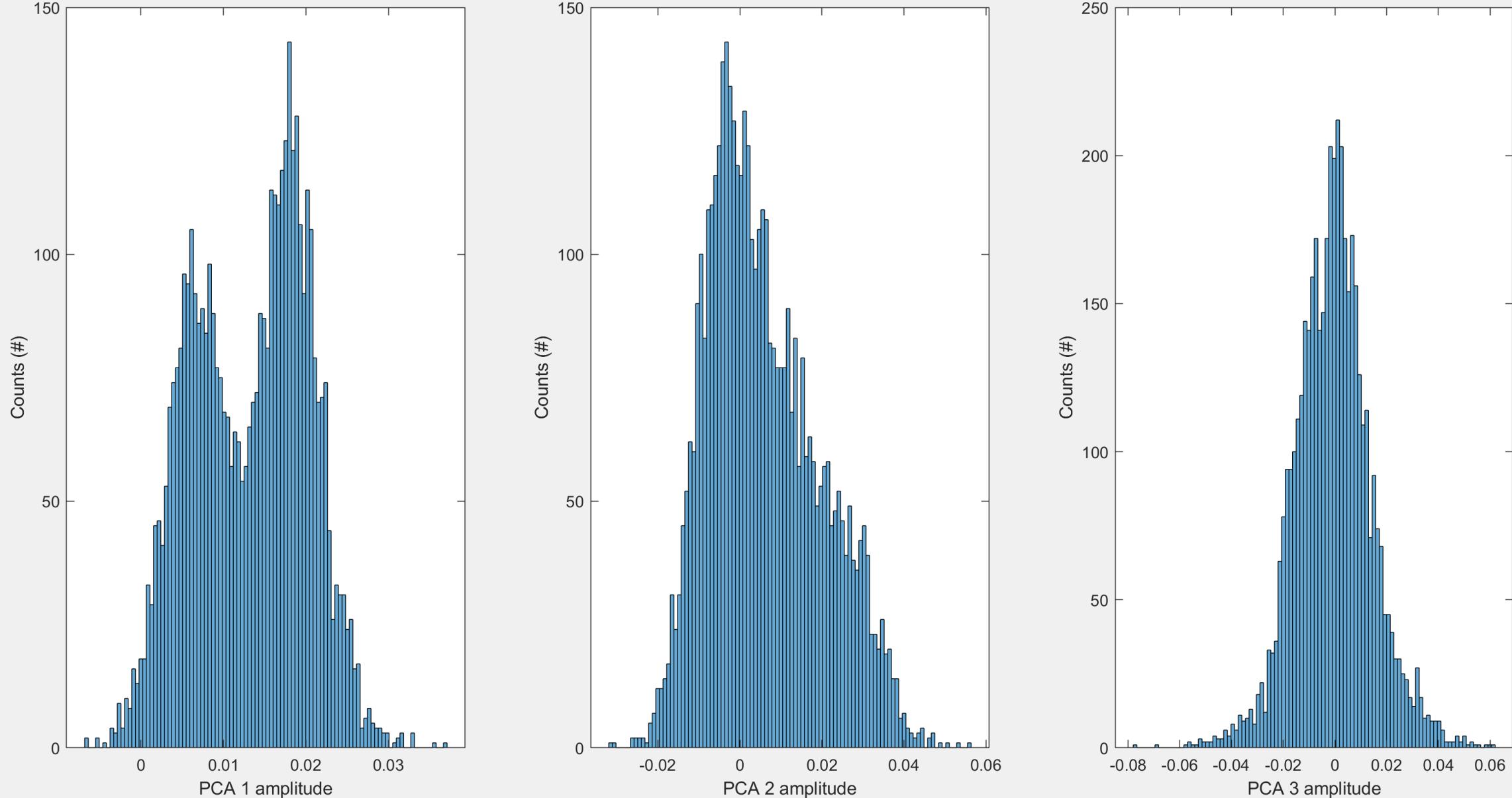


Figure 5a: in question 15, we made histograms for select PCA components. The first eigenvector had bimodal peaks and not centered around 0, so we could be reasonably confident that PCA component 1 was significant and distinct.

Histograms of values for PCA components 4-6

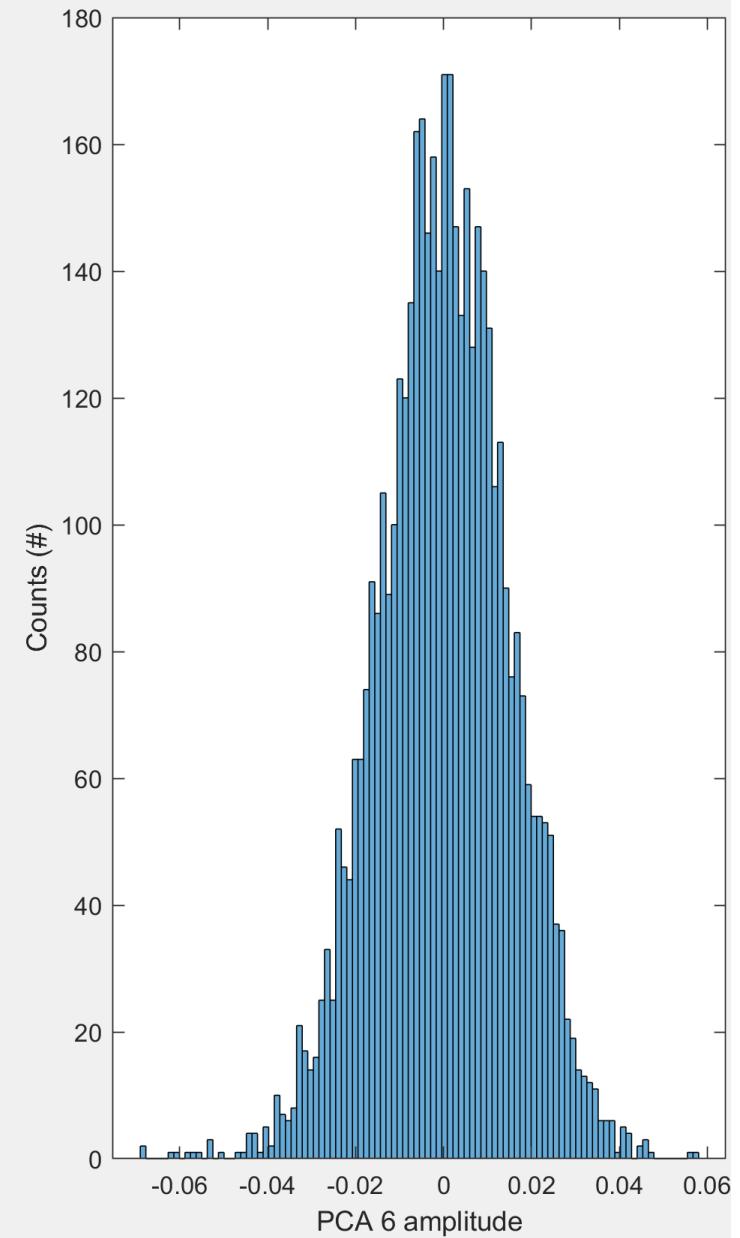
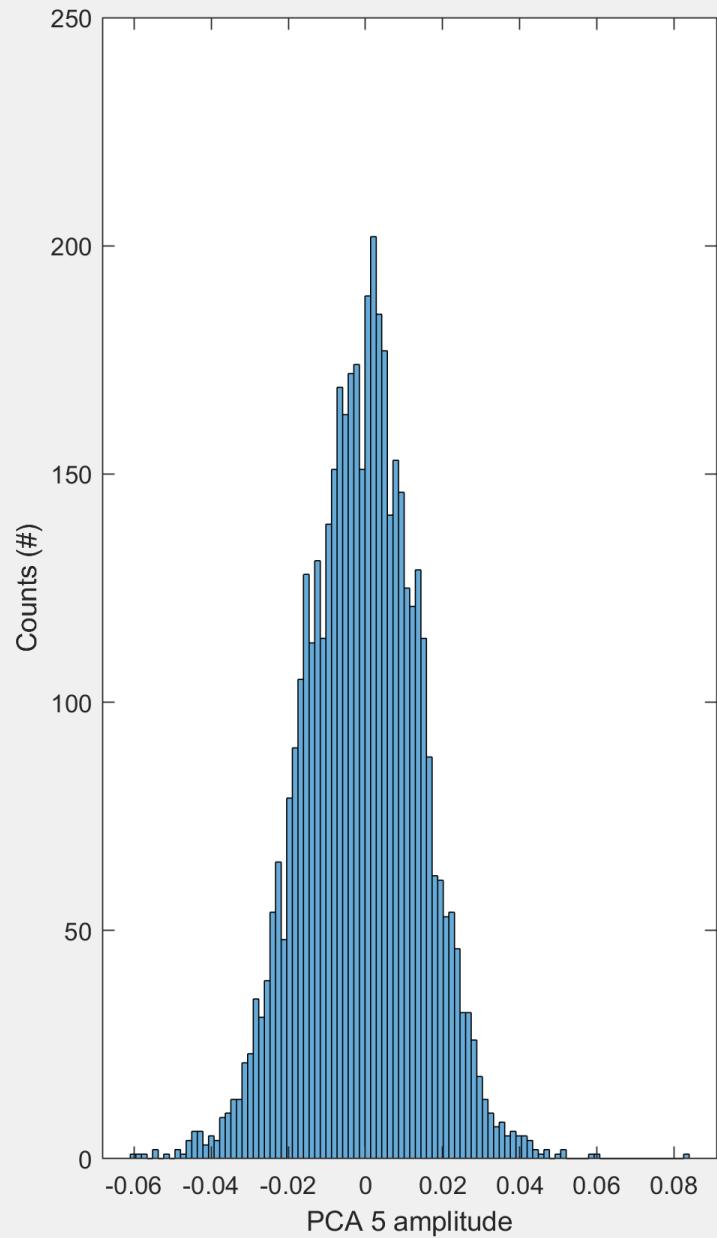
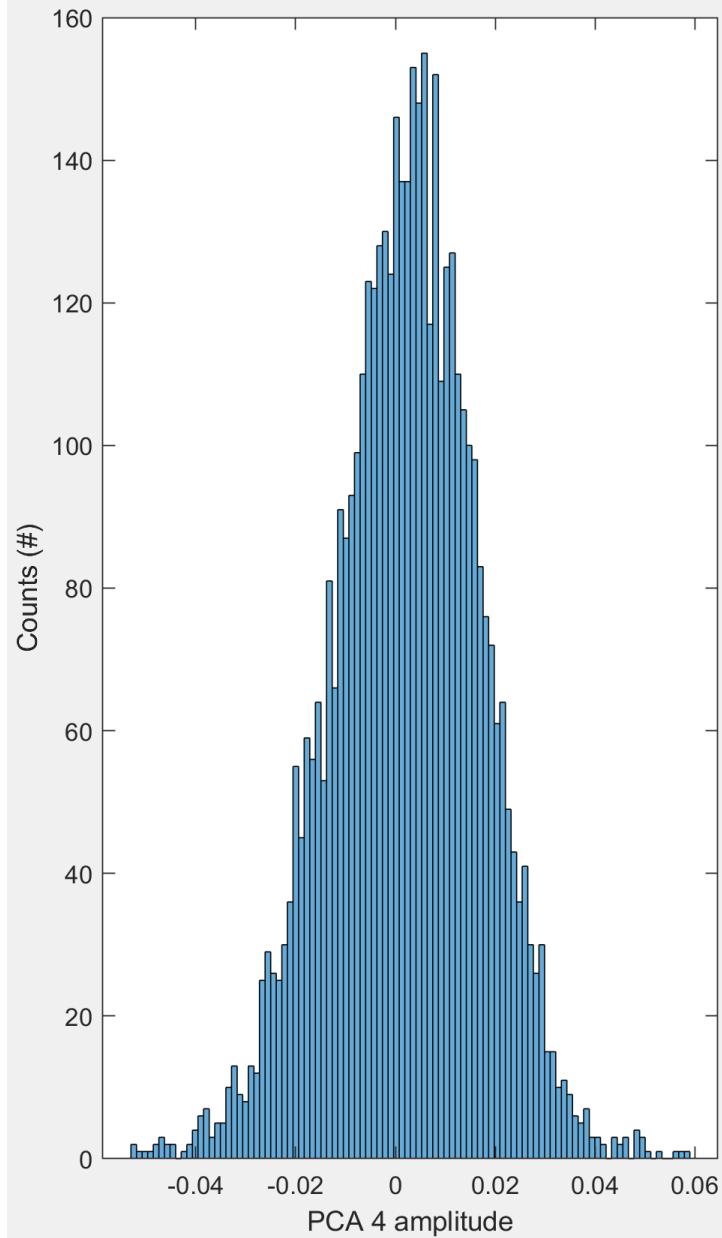


Figure 5b: we made histograms for the next 3 PCA components. Since the PCA component means were centered at 0, we could be reasonably certain that PCA components 4-6 are noise and not distinct clusters.

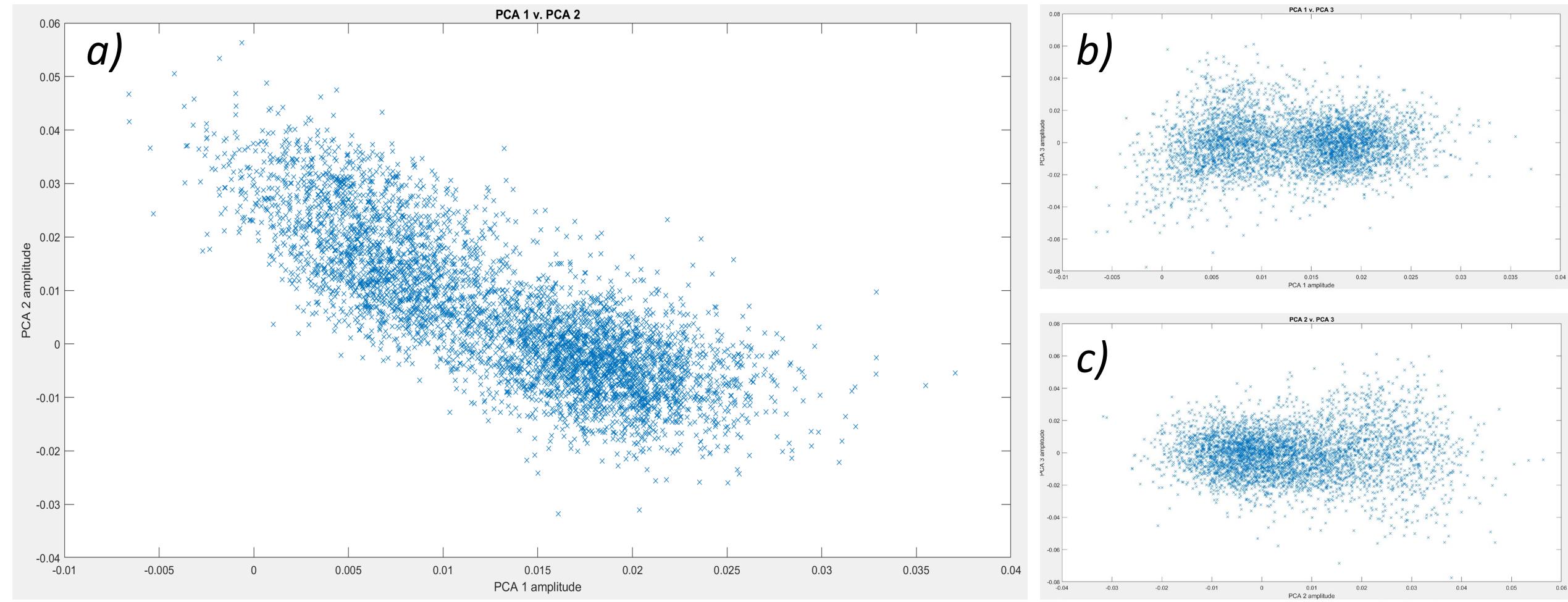
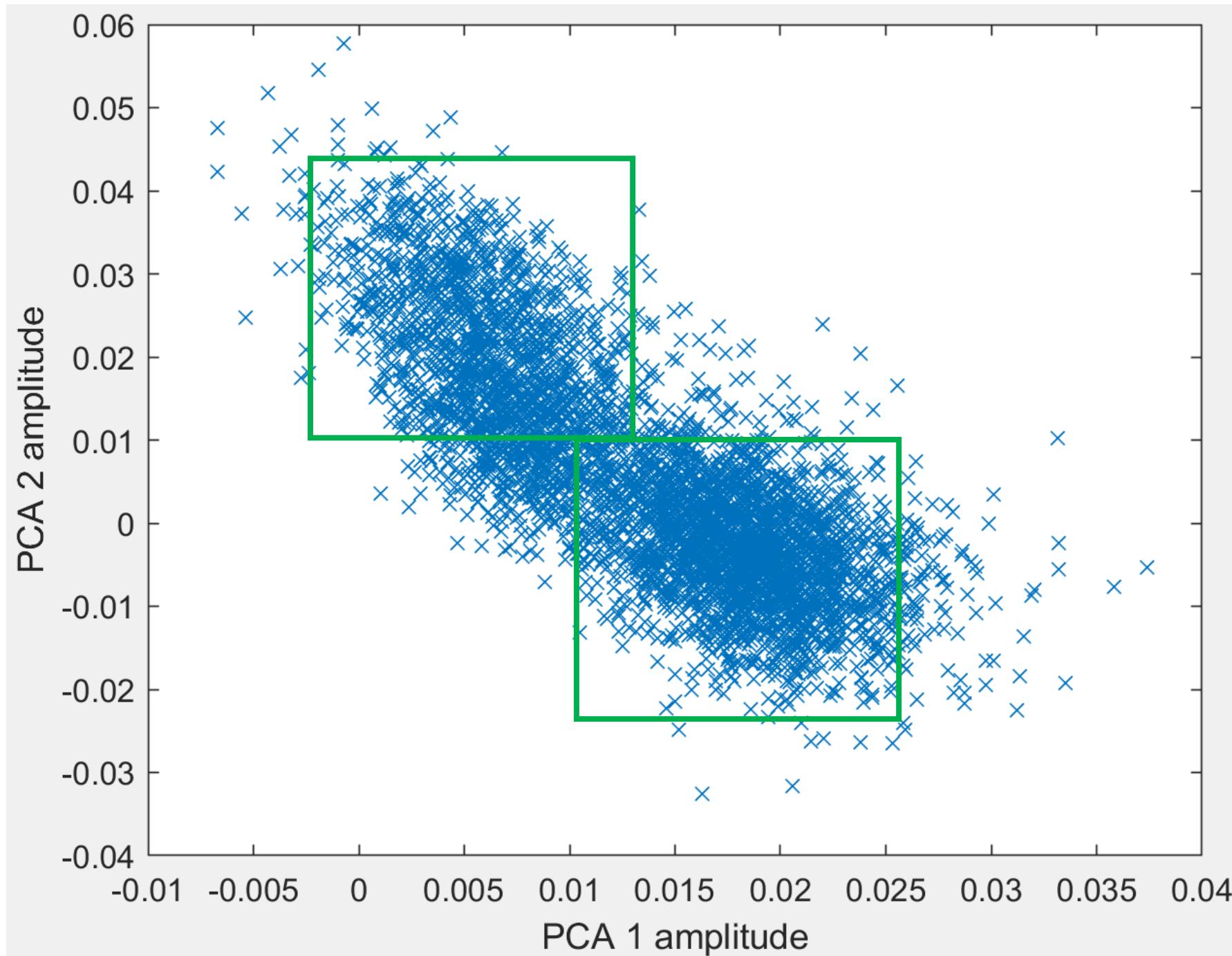


Figure 5c: Select combinations of channel 8's PCA components were plotted against each other. Because the “shapes” matrix, A, from our svd-PCA step is rank-ordered, we know the eigenvector columns are arranged in such a way that column 1 explains the most variance of the snippet data, followed by column 2, etc. The scatterplot in a) appears as if there are two distinct clusters. In scatterplots b) and c), the grouping is increasingly less obvious. This is consistent with the counts from the histogram.

Figure 5c: In problem 15, the amplitudes of the first two PCA eigenvectors were plotted against each other to determine how to cluster the data. Clustering occurs in our model's scaling/coefficient space, so we used the matrix W from the singular value decomposition (svd)-PCA step. There appear to be two clusters, but the proximity between clusters might lead to misclassification errors.

From a visual inspection of the scatterplot of PCA components 1 and 2, there appear to be two candidate clusters:

Cluster 1: $\epsilon_{11} = 0.01 < W(:,1) < \epsilon_{12} = 0.025$
Cluster 2: $\epsilon_{21} = 0.01 < W(:,2) < \epsilon_{22} = 0.045$



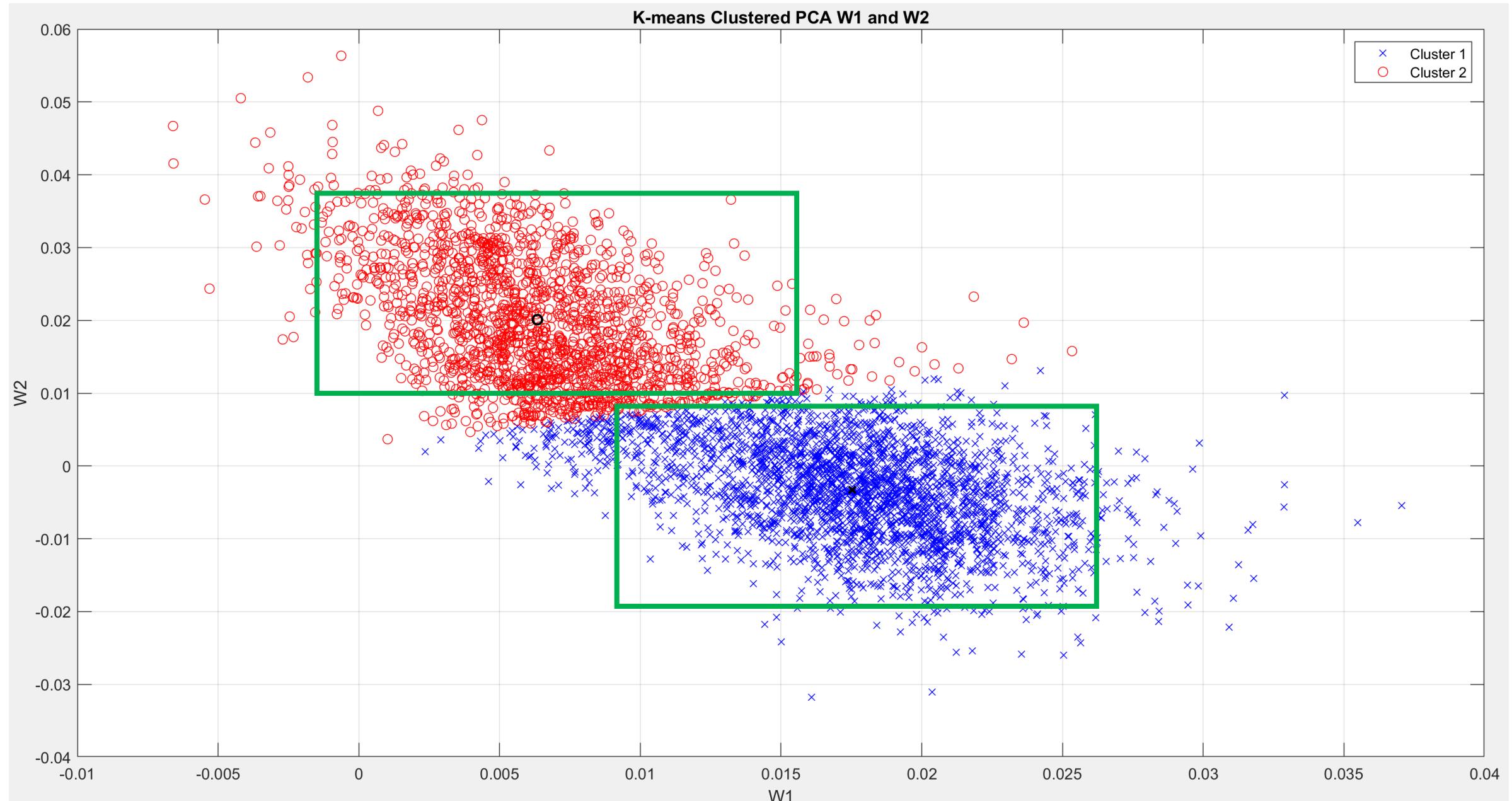


Figure 5d: We used `kmeans` to determine how to sort the PCA component scaling coefficients (W) into 2 clusters. However, since a parameter to `kmeans` is how many clusters we think there are, of course there will be two distinct clusters.

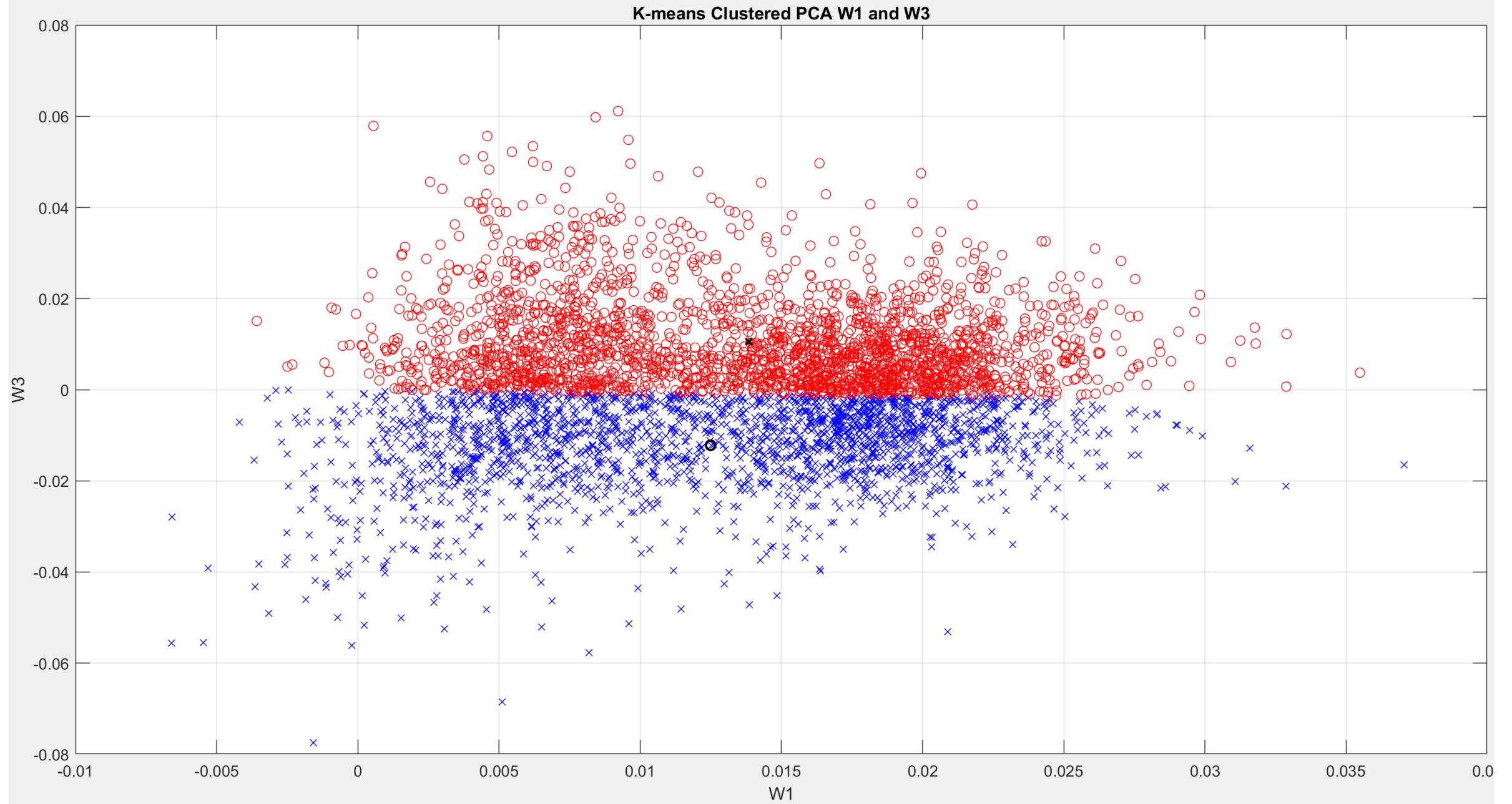


Figure 5e: We used kmeans to determine how to sort the PCA component scaling coefficients (W) into 2 clusters. However, since a parameter to kmeans is how many clusters we think there are, of course there will be two distinct clusters.

K-means Clustered PCA W2 and W3

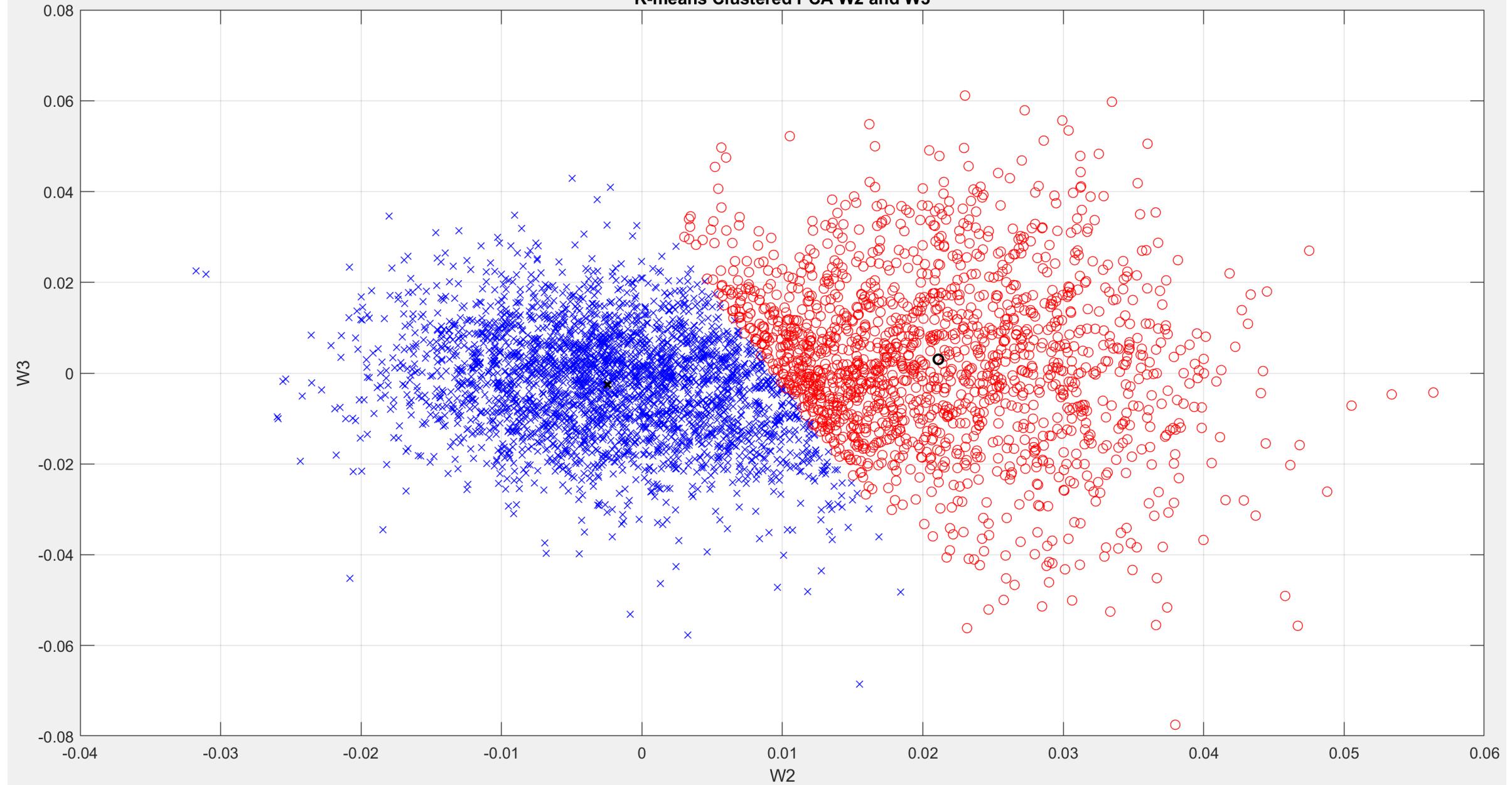


Figure 5f: We used kmeans to sort the second and third PCA component scaling coefficients (W2 and W3) into 2 clusters.

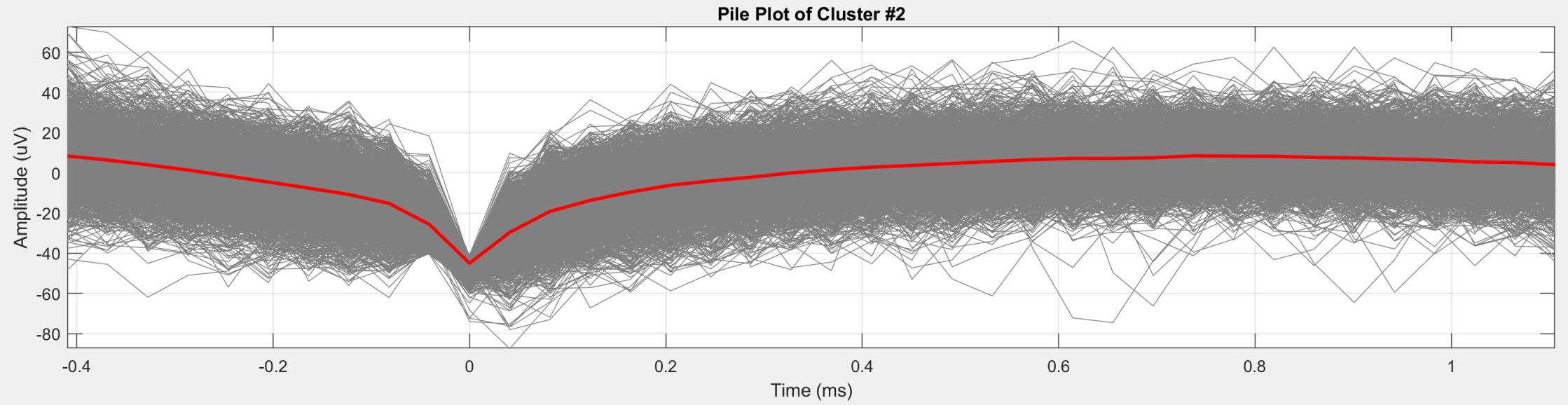
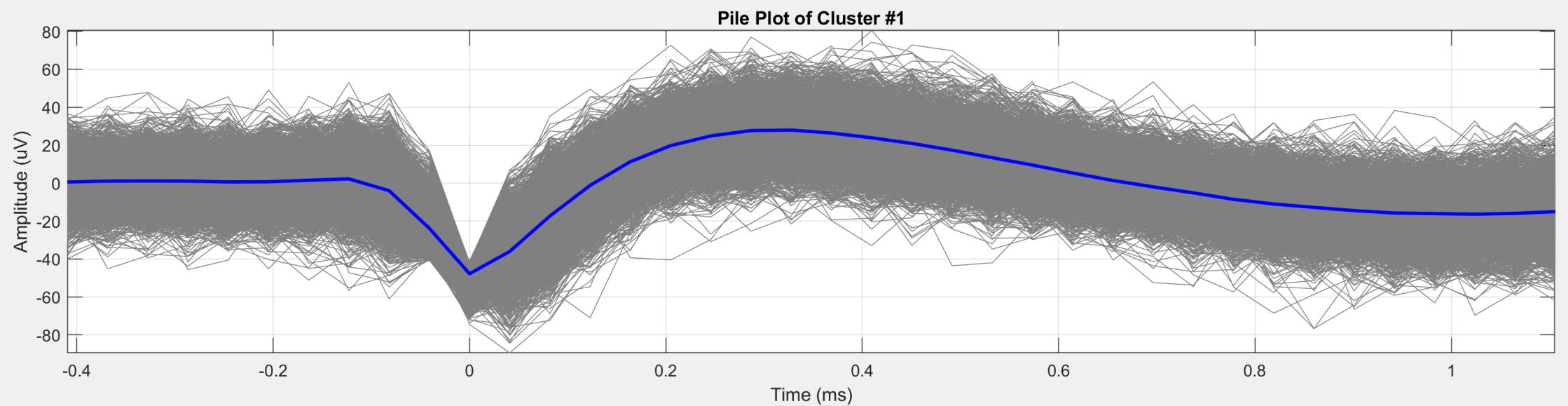


Figure 6a: The pile plots for the first two PCA component clusters are shown above, where the means for each are clearly distinct.

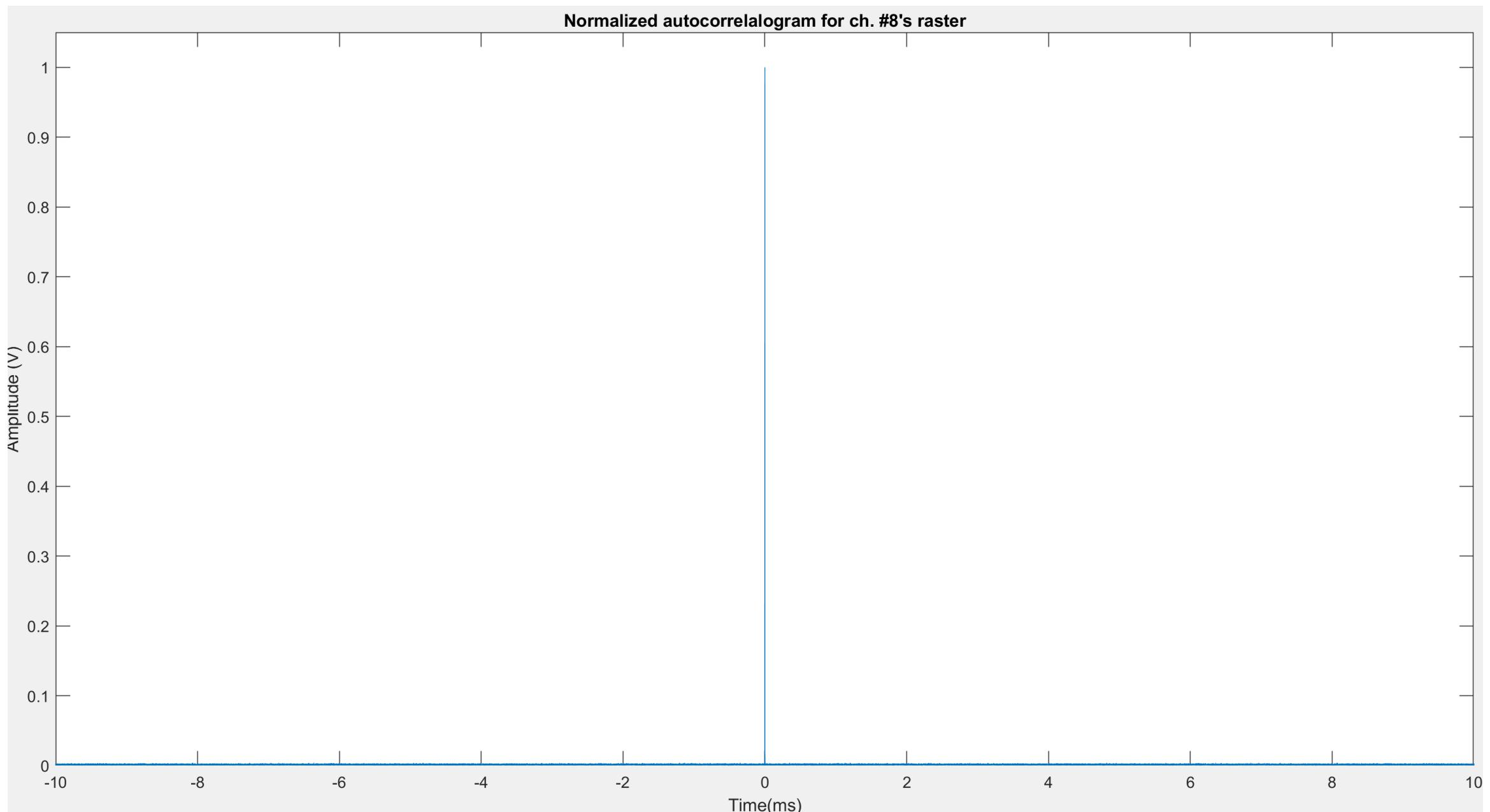


Figure 7a: In question 18, a normalized autocorrelalogram was made for channel 8's raster data. This provides useful information on whether we have single-unit activity, since we expect only a peak at the zero-lag mark: time = 0

