Spike Train Correlation Visualization

Martin A. Walter*
mlwalter@plymouth.ac.uk

Liz J. Stuart*
lstuart@plymouth.ac.uk

Roman Borisyuk*†
rborisyuk@plymouth.ac.uk

*Centre for Neural and Adaptive Systems, School of Computing, University of Plymouth, Plymouth, Devon, UK
†Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142 290, Russia

Abstract

The current ability to record neural activity within the brains of mammals has led to the production of a large body of experimental data. The analysis and comprehension of this data is key to the understanding of many basic brains functions, for example learning and memory.

The main constituent of this data is multi-dimensional spike train recordings. As the analysis of these datasets, by traditional means, becomes more complex and time consuming the need for better methods of data analysis increases.

This paper presents an innovative method for analysis of the relationships within large multi-dimensional spike train datasets. This method, called the 'Correlation Grid,' is based on the Information Visualisation principles; overview the data, filter and zoom the data and obtain details-on-demand [1]. The features of the Correlation Grid are described, including filtering and statistical sorting methods.

1 Introduction

Explanations to many questions in the field of Neuroscience are dependent on the theoretical understanding of a large body of experimental neural data. Specifically, this understanding is fundamental to the exploration of information processing within the nervous system. A primary component of this data is simultaneously recorded multi-dimensional spike trains. Significant research in this area is steered towards the principle of synchronisation of neural activity [2][3].

Further, in-depth, analysis of the available experimental evidence is required in order to extract inherent information. The analysis of neural data, such as multi-dimensional spike trains, is increasingly complex using traditional tools, like cross-correlograms, due to the vast quantity of data involved. Consequently, new analysis methods are required to deal with this data.

The specific computer science field of Information Visualisation is focused on innovations in the representation of vast quantities of data. A guiding principle of Information Visualisation is that the investigator should have control over the data representation that they are using. Moreover, the investigator should be able to manipulate the data by applying relevant techniques, in order to steer that direction of the analysis. For example, it may be appropriate to use statistical or other mathematical routines to sort and organise the data.

The “information–seeking mantra”, introduced by Shneiderman [1] in 1996, highlighted user requirements in this area. It proposed that users should have the ability to overview data, zoom and filter this data and to obtain details-on-demand. This mantra was widely adopted throughout the Information Visualisation community as a basis for defining user requirements. Frequently different levels of detail are viewed using different visualizations. Resulting in a number of different views of the data. For consistency, these multiple views should be linked [4][5][6].

In this paper, a method of dealing with the analysis of relatively large numbers of spike trains, involving cross-correlation, is proposed.

2 Neurophysiological data

Within the mammalian nervous system, there are many different types of neurons, each of which performs a different task. These neurons communicate via small electrical impulses.

A neuron accumulates electrical charge from other neurons attached to it. When this accumulation reaches an internal threshold, the neuron will initiate an action potential. When a neuron initiates action potentials over time, it is said that the neuron is firing. For example, the
application of pressure to the skin causes pressure-sensitive neurons to fire.

Note that action potentials are often referred to as spikes and a series of the spikes, over time, is called a spike train.

2.1 Spike trains

Spike train data is the primary data recorded during experimental Neurophysiology. This data is a record of the spiking activity of a collection of neurons under investigation. Figure 2-1 shows a section, from 200ms to 700ms, of a typical spike train recording for three neurons. In this figure, a horizontal plot represents the spike train of each neuron. Each horizontal plot denotes the occurrence of spikes, at specific times, by a vertical line.

Figure 2-1: An example of a typical spike train recording for three neurons over a period of 500ms.

It has been established that information is encoded in this data. However, it has also been shown that spikes from a single neuron are identical [7]. Thus, the form of individual spikes is believed to carry little information. Instead, it is the spiking frequency and thus, inter-spike-intervals that carry information. Moreover, the synchronisation of spikes within this data is of prime interest. Research [2] has shown that spike synchrony is key to information processing within the brain.

2.2 Coupling

Connections between neurons fall into to general groups; these are Direct Synaptic coupling and Common Input coupling. In both cases the firing patterns of the neurons will be synchronised.

If neuron A fires, then neuron B has an increased probability of firing. Figure 2-2 (ii) illustrates Common Input coupling, where neuron A stimulates both neurons B and C, resulting in the correlation of their input. Thus, if neuron A fires then both neurons B and C have increased probabilities of firing.

2.3 Multi-dimensional spike train data

Explanations to many questions in the field of Neuroscience are dependent on the comprehension of large sets of experimental data. For example, many fundamental brain processes, such as memory, learning and attention, are still unclear in many respects. One of the central theories explaining how the brain processes information is the synchronisation of neural activity. Research [2] suggests that the synchronisation principle may be useful in describing various brain systems.

Much experimental data has been recorded, from the mammalian brain. The majority of this data is in the form of multi-dimensional spike train recordings. Investigation of this data centres on the synchronisation of spikes between spike trains, to derive the coupling of the underlying neurons.

Neurophysiologists still utilise traditional analysis methods due to the absence of more substantial software support. However, due to the quantity of data involved these methods are both time consuming and complex.

3 Traditional methods of analysis

Currently there are a variety of methods to analyse multi-dimensional spike train data. The cross-correlogram is one of the most common and measures the temporal correlation of one spike train with another. Like the cross-correlogram, most current methods are designed for use with only a pair of neurons, and these methods do not scale-up to deal with larger assemblies.

3.1 Pair-wise correlation

The cross-correlogram [8][9] is used to measure the synchronisation between the spike trains of two neurons. One spike train is designated to be the reference train. The other is known as the target. A time frame for correlation must be specified, this is to account for the inherent delay in neural circuits. The time frame, or correlation window, consists of a number of equal time segments, called bins.

The correlation window is centred over the first spike of the reference train. The number of target train spikes that fall within each bin is calculated. This is repeated for each subsequent spike in the reference train.
The result of each comparison is added to the previous to give the overall correlation; an example of this is illustrated in Figure 3-1.

![Figure 3-1 An example of a Cross Correlogram calculation using a six-bin window](image)

The overall correlation is then plotted using a cross-correlogram.

### 3.2 The cross-correlogram

The resultant cross-correlation, plotted as the cross-correlogram, shows the correlation of the target train with respect to the reference train. This is illustrated in Figure 3-2, for two spike trains recorded from neurons coupled as shown in Figure 2-2(i).

![Figure 3-2 Example cross-correlogram for two connected neurons](image)

Should a significant peak [10], as shown in Figure 3-2, be evident in this plot a correlation exists between the two trains. In this case the two neurons are likely to be connected. In contrast when no peaks are evident from the cross-correlogram, see Figure 3-3, it is likely that the neurons are not connected.

![Figure 3-3 Example cross-correlogram for two un-connected neurons](image)

It takes one cross-correlogram to represent the relationship between the spike trains of two neurons. Thus, for any significantly sized neuronal assembly, numerous cross-correlograms would be required to analyse the connectivity of the assembly. New problems are posed by this large quantity of data. This is because all pair-wise results must be analysed to understand any relationships between the underlying neurons.

### 3.3 Peak significance

It is possible for a peak in the output of a cross-correlogram to be ‘false’. These false peaks are random occurrences and due to the high spike frequency of the train.

A number of methods have been proposed to deal with this problem including the Brillinger normalisation and confidence interval [10].

The Brillinger method normalises the results of the cross-correlogram, giving each bin, in the output, a mean value of one. This normalisation supports the comparison of cross-correlograms with different train lengths, as the quantity of spikes on each train is taken into account. In addition, a confidence interval, based on the data, is calculated. Any peaks that are within this interval are defined to be random. In contrast, peaks greater than this interval are considered to show significant correlation.

Figure 3-4 illustrates a Brillinger normalised cross-correlogram, for a pair of common input neurons. Note the majority of bins have values around one, falling within the Brillinger confidence interval. Thus, these bins do not show statistically significant correlation. In contrast, the central peak is much greater than the confidence interval, showing a significant correlation between the input spike trains. Thus indicating a high likelihood of the neuronal coupling.

![Figure 3-4 Example Brillinger normalised cross-correlogram with confidence interval](image)

### 4 Current methods of analysis

A number of innovative methods exist for the analysis of spike trains from large groups of interconnected neurons, called *assemblies*. The ‘Gravity Transform’, originally developed by Gerstein and Aertsen [11][12][13] is one notable method. The Gravity Transform algorithm can be used to study the firing dependencies of multi-dimensional spike trains. Recent work by Stuart et al. [14] has enhanced the output of the original Gravity Transform algorithm using visualisation techniques including parallel
coordinates to support the display of the Gravity Transform output.

However, a great deal work is still necessary in this area in order to fully support exploration of large multi-dimensional data sets. For this reason, the spike train ‘Correlation Grid’ has been developed.

5 Correlation grid

The Correlation Grid presents users with an overview of cross-correlogram results, for a number of spike trains.

For a given dataset, of $n$ spike trains, all unique cross-correlograms are generated, where the user specifies the bin and window size. Subsequently, the cross-correlograms are normalised using the Brillinger method. Finally, the results of these cross-correlograms are displayed as an $n$-by-$n$ grid of grey scale cells, representing the individual correlations between all pairs of spike trains.

The grid encodes the ‘height’ of the largest peak in each cross-correlogram. The peaks are in coded from white, representing no peak, to black, representing the largest peak in the grid.

The user can select whether to view ‘all peaks’ or just significant peaks. Significant peaks are those that fall outside of the Brillinger confidence interval specified for the grid. In addition, the individual cross-correlograms can be view by selecting the corresponding cell in the grid.

To demonstrate the Correlation Grid, a dataset of ten spike trains was generated, over 2000ms, for the assembly of neurons shown in Figure 5-1.

![Figure 5-1 Neuron assembly for test data set](image)

In this assembly, the spike trains of neurons one, three, five and seven will be correlated, as their corresponding neurons are connected. Like wise the spike trains of neurons two, four and six will correlate. In contrast, the spike trains of neurons eight, nine and ten will not correlate with any others as they are unconnected, thus have a random firing pattern.

A Correlation Grid for this dataset was generated, with a bin size of 2ms and window size of 100. The resultant grid, showing al peaks, is shown in Figure 5-2.

![Figure 5-2 Example Correlation Grid showing all peaks (bin size 2ms, window size 100)](image)

On close examination of this Grid, it is possible to deduce that spike trains one, three, five and seven are correlated. Note the higher peaks (represented by darker greys) between these columns/rows. Likewise, it is possible to conclude that a relationship exists between the spike trains of neurons two, four and six.

The clarity of these relationships is greatly improved by filtering the grid so it only containing significant peaks. This filtered grid is shown in Figure 5-3. Recall the correlation between spike trains one, three, five and seven which is more prevalent in this filter Grid. Like wise the correlation between train two, four and six.

![Figure 5-3 Example Correlation Grid showing only significant peaks (bin size 2ms, window size 100)](image)

With the removal of the non-significant peaks the grid is less cluttered. It is now easier to identify the correlation between the connected neurons, shown in Figure 5-1.

Moreover the lack of correlation between spike trains is more apparent. From an examination of Figure 5-3 it is possible to see that trains eight, nine and ten have minimal, or no, correlation with other trains.

5.1 Cluster analysis

The identification of groups, or clusters, of correlations is key to understanding the relationships between the underlying neurons. It is possible to identify these clusters visually, however this is not easily expanded for problems with larger datasets.
To aid with the identification of correlation clusters, a statistical cluster analysis method has been implemented. This method uses the height of the most significant peak, if any exist, of each cross-correlogram to build a dendrograph for the correlation grid. This in turn is used to generate the initial display order of the spike trains.

The effect of clustering is shown in Figure 5-4. This grid was generated from the same data as used to create the grid in Figure 5-3, based on the neural assembly in Figure 5-1. Cluster analysis was performed on the grid, to determine the optimal sequence of spike trains in the grid. This analysis was performed using significant peaks only.

With the aid of filtering and clustering it is possible to clearly identify the groups in the neuronal assembly shown in Figure 5-1.

![Figure 5-4 Example Correlation Grid, showing only significant peaks and clustered (bin size 2ms, window size 100)](image)

From Figure 5-4 it is possible to identify the correlations in the neuron ‘ring’: neurons one, three, five and seven. They are shown in the top left portion of the grid. Additionally, the common input group; neurons two, four and six; are also grouped together in the centre of the diagram. Finally, from the same figure, the independent neurons, eight, nine and ten, are all in the lower right portion of the grid.

From closer examination of the cross-correlograms of the final cluster, of unconnected neurons, it is apparent that the peaks are relatively small. See Figure 5-6 (i), which shows the cross-correlation for neurons two and ten.

![Figure 5-5 Cross-correlograms for (i) neurons two and ten (ii) neurons two and four from the previous Grid](image)

From the display it is possible to see that the peaks are relatively small, only slightly greater than the confidence interval. In contrast the correlations of connected neurons, shown in Figure 5-6 (ii) for neurons two and four, show relatively large peaks. Thus, for neurons eight, nine and ten, it would be possible to infer that these peaks do not represent true correlation between the spike trains. Hence, these neurons are not coupled to any others in the assembly.

This overview of cross-correlograms facilitates improved fine-tuning of the correlation parameters. This fine-tuning is commonly needed to obtain the highest clarity of result.

Increasing the bin size, for example, means more spikes are taken into consideration. Thus, a peak must be higher to be classed as significant. This principle is illustrated in Figure 5-6 where the bin size for the grid has been changed to 3ms. The same dataset and previous window size were used to create this grid.

![Figure 5-6 Example Correlation Grid showing only significant peaks (bin size 3ms, window size 100)](image)
The grid in Figure 5-6 has been filtered to solely show significant peaks and clusters. Note that correlation only exists between the spike trains of the coupled neurons. For example, spike trains eight, nine and ten show no correlation with any other spike trains.

6 Future work

The work presented in this paper is part of an Information Visualisation project, at the Centre for Neural and Adaptive Systems, called Visualisation of Inter-Spike Associations (VISA) [15]. Specifically, this paper has presented work on an innovative method for analysis of large neural assemblies using cross-correlation. However, a number of areas still require attention.

Currently, users are unable to alter the order of the spike trains. The ability to reorder the grid and to fine-tune the display will be implemented.

Empirical testing is underway to evaluate the usability of this visualization method. Subsequently, it is likely that this tool will be integrated into the VISA Toolbox [15].

7 Acknowledgments

This research is supported by the Engineering and Physical Sciences Research Council (EPSRC) grant number GR/N04904. Additionally, the research of Roman Borisyuk is supported in part by the Russian Foundation of Basic Research (grant 99-04-49112) and by EPSRC (grant GR/N63888/01).

8 References

[15] Visualization of Inter-spike Association (VISA) project website http://www.tech.plym.ac.uk/soc/research/neural/staff/lstuar t/VISA/index.html