

# The Representation of Neural Data using Visualization

Mr. Martin Walter\*  
mwalter@plymouth.ac.uk

Dr. Liz Stuart\*  
lstuart@plymouth.ac.uk

Prof. Roman Borisyuk\*\*†  
rborisyuk@plymouth.ac.uk

\*School of Computing, Communication and Electronics, University of Plymouth, Plymouth, Devon, UK

†Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142 290, Russia

## Corresponding Author:

Mr Martin Walter

Room B310, Portland Square

School of Computing, Communication and Electronics

University of Plymouth

Plymouth

Devon

PL4 8AA

United Kingdom

Telephone: +44 (0) 1752 23 2620

Fax: +44 (0) 1752 23 2540

Email: mwalter@plymouth.ac.uk

## 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).

## **Abstract**

Currently, the focus of research within Information Visualisation is steering towards genomic data visualisation [21] due to the level of activity that the Human Genome Project [11] has generated. However, the Human Brain project [10], renowned within Neuroinformatics, is equally challenging and exciting. Its main aim is to increase current understanding of brain function such as memory, learning, attention, emotions and consciousness. It is understood that this task will require the “integration of information from the level of the gene to the level of behaviour”.

The work presented in this paper focuses on the visualisation of neural data. More specifically, the data being analysed is multi-dimensional spike train data. Traditional methods, such as the ‘raster plot’ and the ‘cross-correlogram’, are still useful but they do not scale up for larger assemblies of neurons.

In this paper, a new innovative method called the Tunnel is defined. Its design is based on the principles of Information Visualisation; overview the data, zoom and filter data, data details on demand [18]. The features of this visualisation environment are described. This includes data filtering, navigation and a ‘flat map’ overview facility. Additionally, a ‘coincidence overlay map’ is presented. This map washes the Tunnel with colour, which encodes the coincidence of spikes.

## **Keywords**

*Information visualisation, neural data, spikes, spike trains, analysis, visualisation environment.*

## 1 Introduction

Solution of many problems in the field of Neuroscience is associated with the theoretical comprehension of a large body of experimental neural data. Specifically, investigation of information processing in the nervous system is associated with the analysis of these vast quantities of neural data. More specifically, this data is simultaneously recorded multi-dimensional spike train data. Much of the research effort in this area is steered towards the principle of synchronisation of neural activity [4][6].

The experimental evidence that is currently available requires further, in-depth analysis in order to extract inherent information. Analysis of neural data such as multi-dimensional spike trains using traditional tools like raster plots and cross-correlograms is increasingly complex due to the vast amount of data involved. Hence, new methods of analysing this data are required.

Information visualisation is one of the fields of computer science that deals with innovations in the representation of vast quantities of data. One of the fundamentals of this field is the ability of the investigator to interact with the data being analysed, in order to achieve greater insight. Thus, the investigator should be able to navigate throughout the data, to identify and explore specific subsets of interest. When visualising large datasets, the issue of efficient navigation is amplified. It is important for the

user to be able to move quickly to points of interest without becoming disoriented within the data. By limiting the ways in which the user can navigate, this problem can be alleviated. The user can be constrained to follow predetermined paths throughout the data space. Subsequently, reversing or re-tracing your steps [15] becomes trivial. In addition, providing the user with different frames of reference can also help [17].

In addition to navigation capabilities, the investigator should also have control over the representation itself. Thus, in order to steer the analysis, they should be able to manipulate the data by applying appropriate techniques. For example, it may be appropriate to sort or organise the data using statistical or other mathematical routines.

Traditionally, analysis of multi-dimensional spike train data has not supported real-time user interaction. In 1996, Shneiderman [18] identified user interaction as one of the essential components of information visualisation. Shneiderman also introduced the “information-seeking mantra” that highlighted user requirements in this area. It specified that users should have the capability 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. In many cases, different visualization are utilised to represent data at differing

levels of detail. Thus, resulting in the creation of a number of different views of the data. Ideally, these multiple views should be linked for consistency [13][3][14].

In this paper, an interactive method for exploring neural data is presented. This representation of data is based on current information visualization and virtual reality principles. It supports multiple views of the neural data as well as real time interaction by means of a 'toolbox' that facilitates zooming, filtering and manipulation of the data.

## **2 Neural Data**

There are many different types of neurons in the mammalian nervous system, each of which performs a different task. For example, excitation of motor neurons controls muscle fibres resulting in the contraction of muscles. These neurons communicate via relatively small electrical impulses.

### **2.1 Spikes and Spike Trains**

In general, a neuron accumulates electrical stimulus, from other neurons coupled to it, until some internal threshold is reached. Once its threshold is reached, the neuron initiates an action potential. When a neuron initiates action potentials over time, we say that the neuron is firing. Note that action potentials are more commonly referred to as spikes and a series of these spikes over time is known as a spike train.

Spike train data is one of the main types of data collected during Neurophysiological experimentation. It is a record of the activity of a collection of neurons under investigation. Figure 1 shows a section, from 300ms to 800ms, of a typical spike train recording for three neurons. In this figure, a horizontal plot represents the spike train of each neuron. This horizontal plot denotes the occurrence of spikes, at specific times, by a vertical line.

It is well established that information is encoded in this data but it has also been established that each spike from a single neuron is identical [16]. Hence, 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. Thus, research is focused on the analysis of multi-dimensional spike train data to reveal information about the synchronisation of spike trains and the coupling of neurons.

### **2.2 Coupling**

*Direct Synaptic* coupling and *Common Input* coupling are the general cases of coupling where synchronisation may occur between the firing of two neurons.

Direct Synaptic coupling is illustrated in Figure 2 (i) where neuron A is coupled so that it stimulates neuron B. If neuron A fires, then neuron B has an increased probability of firing. Common Input coupling is illustrated in Figure 2 (ii), where neurons A and B

are both coupled so that they receive stimulation from a third neuron C, resulting in the correlation of their input. Thus, if neuron C fires then both neurons A and B have increased probabilities of firing.

### 2.3 Multi-Dimensional Spike Train Data

Solutions to many of the problems in the field of Neuroscience are associated with understanding large bodies of experimental data. For example, basic brain functions such as memory, learning, attention, emotions and consciousness remain unclear in many respects. More specifically, one of the basic principles that underlie information processing in the brain is the principle of synchronisation of neural activity. Research [4] indicates that the synchronisation principle may be useful in devising various systems of information processing.

The current capability, to record neural activity within the brains of mammals, has led to the production of large quantities of experimental data. This data is in the form of multi-dimensional spike train recordings. Investigation of this data focuses on the synchrony between spike trains and the coupling of neurons.

Traditional methods of analysis are still employed by Neurophysiologists in the absence of more substantial software support. However, this type of analysis is both time consuming and complex due to the quantity of data currently available.

## 3 Traditional Methods of Analysis

A number of methods exist to analyse multi-dimensional spike train data. Two of the most commonly used methods are the *raster plot*, which directly plots the spike train data, and the *cross-correlogram* used to analyse the correspondence between spike trains. Most current methods are designed for use with a pair of neurons and do not scale-up to deal with larger numbers of neurons.

### 3.1 The Raster Plot

The raster plot [1] is one of the original methods for viewing and analysing spike train recordings. Each train is displayed as a line of dots, where each dot representing the presence of a spike at that time from the stimulus. Raster plots can be used to compare a number of recordings from the same neuron. This aids in the identification of similarities between these trials. In addition, raster plots can be used to view a number of spike trains from different neurons.

Figure 3 shows two raster plots each displaying spike train data from two neurons. From Figure 3 (i) it can be deduced that the spike trains of neurons a, and b, are synchronised and thus, neurons a, and b, are likely to be coupled. From Figure 3 (ii), note that the spike trains of neuron a and b are not obviously correlated and thus neurons a and b are less likely to be coupled.

### 3.2 The Cross-Correlogram

The cross-correlogram [1][12] quantifies 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'. Due to the inherent delay in neural circuits, a time frame for correlation must be specified. This 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 process is repeated for each subsequent spike in the reference train. The results of individual comparisons are summed up to give the overall correlation.

This overall correlation is then plotted as the 'cross-correlogram', see Figure 3-2, for the two spike trains and shows the correlation of the target train with the reference train.

If the cross-correlogram has a significant peak [5] a correlation exists between the two trains. Consequently, it is likely that the two neurons are connected. In figure 3-2 (i) the cross-correlogram of two spike trains, from connected neurons, clearly shows a significant peak. In contrast, noting the different scale, the cross-correlogram in figure 3-2 (ii) shows no significant peaks, indicative of neurons that are not connected.

The cross-correlogram is a very useful and commonly used tool for representing the relationship between the spike trains of a pair of neurons. However, for any significantly sized neuronal assembly, numerous pair-wise results would be generated. This large quantity of resultant data poses its own analysis problem, as all pair-wise results must be analysed to understand any relationship between the underlying neurons.

### 4 Current Methods of Analysis

Other innovative methods also exist for the analysis of the spike trains of large groups of interconnected neurons, also known as assemblies. One notable method is the 'Gravity Transform', originally developed by Gerstein and Aertsen [7][8][2]. The gravity transform algorithm can be used to study the dependencies in firing of multi-dimensional spike trains. Recent work by Stuart et al. [19] has enhanced the output from the original gravity transform algorithm using visualisation techniques including parallel coordinates.

However, much work is still needed in this area in order to fully support interactive exploration of these large multi-dimensional data sets. For this reason, the spike train 'Tunnel' was developed.

### 5 The Tunnel Environment

The spike train Tunnel is a method of analysing the firing patterns of multiple neurons. This environment presents different views of the dataset and an

additional overlay that encodes spike coincidence. It enables the user to focus on a specific subset of the dataset using a set of interaction tools. Different frames of reference are provided to enable investigators to track their location within the data space.

### **5.1 The Tunnel Visualisation**

The Tunnel is a cylindrical environment that supports user interaction. Figure 5 shows the Tunnel visualisation of a randomly generated dataset over 200ms.

Each of the numbered horizontal bands, that comprise this Tunnel visualisation, encodes the spike train of the corresponding neuron. The two 'end' bands, band 1 and band 10 in Figure 5, are adjacent to each other, thus forming the cylindrical environment. Note that, time is represented down through the Tunnel.

Overall, illumination inside the environment represents the firing of neurons in the currently displayed portion of the dataset. Synchrony is detected by the perception of the position, intensity and frequency of light sources at different parts of the Tunnel.

In the Tunnel visualisation, the investigator is able to 'fly' through the Tunnel to arrive at sections of the Tunnel (subsets of the data) that are of specific interest. The user has control over both, the speed and direction of the flight. However, to minimise the possible side-effects of disorientation during

navigation, motion is restricted to being along the Tunnel's length. Thus, the user is restricted to forward and reverse motion.

### **5.2 Filtering Data**

To meet the user requirements previously discussed in section 1, the Tunnel has filtering functionality known as 'dimming'. Whilst in the filtering mode, dimming may be switched, on or off, for each of the spike trains individually.

Figure 6 illustrates filtering of another 200ms dataset. In this dataset, the spike trains of neurons 4, 6 and 8 are identical. The remaining spike trains were all randomly generated. In this figure, all spike trains are dimmed with the exception of spike trains 4, 6 and 8. This filtering is designed to enable investigators to highlight spike trains of interest while maintaining context within the dataset.

### **5.3 Coincidence Sorting**

As stated previously, in section 2.1, a key concept of spike train analysis is the identification of coincidence between spikes. Within the Tunnel environment, the investigator is able to progressively sort the order of the spike trains to view spike coincidence. The user selects a spike on a reference train. Subsequently, the spike trains in the Tunnel are reordered, so trains with spikes coincident to the selected spike are adjacent. Trains that do not have any coincident

spikes are inherently moved away from the reference train.

To illustrate coincidence sorting another 200ms dataset, based on an assembly of ten neurons, was generated. In this assembly, neurons three and ten fire every 12ms and neuron eight fires every 7ms. To illustrate the progressive sorting feature of the Tunnel, neurons four and six fire every 7ms and 12ms.

The Tunnel visualization of this unsorted dataset is illustrated in Figure 7. Note the first coincident spikes (at 12ms) on spike train three and four, and the following non-coincident spike (at 14ms), solely on train four. Further, the selected spike on the reference train is highlighted in yellow. Subsequent to sorting, spike trains four and eight are moved adjacent to the reference train, six, due to coincidence with the selected spike. This reordering is shown in Figure 8.

The Tunnel visualisation provides a progressive sorting facility. When a successive sort is applied, the order of sorted spike trains is initially preserved. However, as this ordering is applied, any train with a spike correlated to the currently selected spike overrides this order. This supports 'fine-tuning' of the spike train order within the Tunnel.

As a result of this type of successive sort, highly correlated spike trains are nearer to each other. This is demonstrated in Figure 9. Spike train three, four and ten are near to train six due to their correlation with the currently selected spike (highlighted in blue).

Note that, the previous ordering of train four adjacent to train six persists as it correlated to both the first and second selected spikes. Since train three only correlated to the currently selected spike, it cannot displace train four. In contrast, note that train ten displaces spike train eight.

#### **5.4 Coincidence Summary**

In addition to individual spike coincidences, the overall spike coincidence of the dataset is also of interest. Thus, the Coincidence Summary was developed. This representation derives a summary of neuron firing and colour codes this data.

Each spike train in the dataset is divided into a number of equal time slots,  $n$ , commonly referred to as bins. The size of bin is specified by the user, but is usually relative to the neural transmission delay time.

Each spike train has an associated array made up of  $n$  elements, where each element of the array is associated with a bin. Each bin is inspected and if one, or more, spikes occur inside the bin, the corresponding element in the associated array is set equal to one, otherwise zero. Note, the total number of spikes in each bin is not important, but the presence, or absence, of a spike within that bin.

When all of the associated arrays have been calculated, an intermediate summary array, of  $n$  elements, is created. It is computed from the associated arrays, such that, the  $i$ th element of this



intermediate summary array is equal to the sum of the  $i$ th element of each associated array.

Subsequent to computation, if an element of the intermediate summary array is less than two, there is no spike coincidence. Thus, these elements are set equal to zero in the final summary array. Note that the maximum value of an element in this final summary array is equal to the total number of spike trains in the dataset. This array is used to create the Coincidence Summary Visualisation.

The dataset used to create the visualisation in Figure 11 comprised ten spike trains each lasting 200ms. Each of the spike trains were created by appending four, 50ms trains together, such that the first and third segments were low in spike frequency in comparison to the second and fourth segments. Thus, for this 200ms dataset, the final summary array is made up of 67 elements where there are sixty-six 3ms bins and one 2ms bin. Due to the way in which the dataset was generated, elements 1-17 and 35 –51 of the final summary array will be relatively low in value with respect to elements 18-34 and 52-67.

This data is encoded for the Coincidence Summary Visualisation using colour based on the Hue map shown in Figure 10.

The Coincidence Summary Visualisation (CSV) for the Tunnel is illustrated in Figure 11. Recall, that the red labels of the trains bear no relation to coincidence, they have been added to this paper for clarity.

Note that the majority of the visualisation viewed at this position in the Tunnel primarily displays the first segment of the final summary array. Due to the relatively low values of the first segment, this is represented by different hues of blue.

It is possible to distinguish the second segment of the final summary array, at the centre of the visualisation. Due to the relatively high values in this segment, it is represented by mainly red and yellow hues.

### **5.5 Combining them Together**

In addition to viewing the CSV, it is also possible to superimpose the Tunnel visualisation onto the CSV. An example of this is shown in Figure 12 where the data used to generate the CSV in Figure 11 has been superimposed by the corresponding Tunnel visualisation of this dataset.

This combination of the CSV and the Tunnel visualisation increases the complexity of the display but also helps identify coincidence between spike trains.

Note that in this combined visualisation, the colour used to represent a spike is defined by the corresponding colour in the CSV. This relates to the overall firing activity at that time in the Tunnel.

### **5.6 The 'Flat Map' Representation**

Investigators can also overview the data using the 'flat map' representation as shown in Figure 13. This is similar to the raster plot discussed in section 3.1 but it

has additional functionality. It enables the investigator to select a subset for analysis, thus it has the capability to zoom in upon “interesting” data. In order to zoom, the investigator highlights an area of the ‘flat map’. The red boundary line indicates the section that will be accentuated. The boundary line enables the selection to be fine-tuned. Moreover, it enables the user to track their location within the data set.

From Figure 13, all ten spike trains, ST1 to ST10, are distinguishable but it is difficult to draw any conclusions about their pair-wise correlation to each other.

### **5.7 Undo/Redo Facility**

In addition to user interaction and navigation, the Tunnel supports a comprehensive undo/redo facility. Shneiderman asserted that the ability to back track adaptation to the visualization was key to the refinement of understanding [18]. Thus, the user should be able to easily return to previous states of the visualisation.

Towards this end, the environment tracks all changes to the spike train order enabling the user to selectively undo/redo refinements to aid understanding.

## **6 Future Work**

The work presented in this paper is part of an Information Visualisation project, at the Centre for Neural and Adaptive Systems, University of Plymouth, called Visualisation of Inter-Spike Associations (VISA)

[VISA]. Specifically, this paper has described an innovative visualisation technique for the analysis of multi-dimensional spike train data called the Tunnel visualisation.

Currently, the ‘flat map’ overview of the Tunnel identifies the section the user has chosen to zoom in upon, but it does not track the users position. Hence, an exocentric frame of reference [17] needs to be developed to identify the current position of the user.

Currently, the user has the ability to reorder the bands that comprise the Tunnel to support the investigation of spike coincidence. The bands of the Tunnel also form the ‘flat map’ overview. It has already been established that the usefulness of multiple views is significantly increased if views are linked [13]. This needs to be developed for the Tunnel visualisation.

The initial order of the spike trains is dependent on their order in the input file. Future work will see the implementation of cluster analysis [22]. Based on the cross-correlation of the spike trains, this analysis will be used to find the optimal initial spike train order. This will improve the speed at which correlation can be identified.

Currently, the user may only select one, continuous, area of the tunnel to zoom. It could prove useful to facilitate the zooming of a number of disjoint sections of the Tunnel, simultaneously. Allowing the user to view, and compare, a number of sections of interest.

In this document, snapshots of the Tunnel have spike train numbers appended to aid identification. This labelling will be built into the system. A number of identification methods exist and currently two are under consideration. To identify a train, the user could mouse over the train to see the number. Alternatively, a semi transparent overlay of the train number could

be developed. More over, it may be appropriate to implement both methods, allowing the user to select between them.

Empirical testing is currently under way to evaluate the usability of this visualization method. Subsequently, it is likely that the Tunnel will be integrated into the VISA Toolbox.

## 7 References

- [1] AWISZUS F 1997. Spike train analysis. *Journal of Neuroscience Methods*, 74. pages 155-166
- [2] BAKER, S.N., AND GERSTEIN, G.L., 2000. Improvements to the Sensitivity of the Gravitational Clustering for Multiple Neuron Recordings. *Neural Computation* 12, 2597 - 2620.
- [3] BALDONADO M. Q. W, WOODRUFF A., AND KUCHINSKY A 2000. Guidelines for using multiple views in information visualization. *Advanced Visual Interfaces*, pages 110-119.
- [4] BORISYUK R.M., AND BORISYUK G.N., 1997. Information coding on the basis of synchronisation of neuronal activity. *BioSystems* 40, 3-10
- [5] BRILLINGER D. R. 1979. Confidence intervals for the crosscovariance function. *Selecta Statistica Canadiana*, V, pages 1-16.
- [6] FRIES P., NEUENSCHWANDER S., ENGEL A.K., GOEBEL R. AND SINGER W., 2001. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience* 4(2), 194-200.
- [7] GERSTEIN, G.L. AND AERTSEN, A.M., 1985. Representation of Cooperative Firing Activity Among Simultaneously Recorded Neurons. *Journal of Neurophysiology* 54(6), 1513-1528
- [8] GERSTEIN, G.L., PERKEL, D.H., AND DAYHOFF, J.E., 1985. Cooperative Firing Activity in Simultaneously Recorded Populations of Neurons: Detection and Measurement. *Journal of Neuroscience* 5(4), 881-889.
- [9] GRÜN S., DIESMANN M., GRAMMONT F., RIEHLE A. AND AERTSEN A. 1999. Detecting unitary events without discretization of time. *Journal of Neuroscience Methods* 94:67-79
- [10] HBP Human brain project web site [WWW Document] <http://www.nimh.nih.gov/neuroinformatics/index.cfm> (accessed 04 August 2003)
- [11] HGP Human genome project [WWW Document] [http://www.ornl.gov/TechResources/Human\\_Genome/](http://www.ornl.gov/TechResources/Human_Genome/) (accessed 04 August 2003)
- [12] MuLAB <http://mulab.physiol.upenn.edu/index.html>
- [13] NORTH C, 2001 Multiple views and tight coupling in visualization: A language, taxonomy, and system. *Proc. CSREA CISST 2001 Workshop of fundamental Issues in Visualization*, pages 626-632
- [14] ROBERTS J. C. 1998. On encouraging multiple views for visualization. *IV'98 – Proceedings International Conference on Information Visualization*, pages 8-14 IEEE Computer Society, July 1998
- [15] ROBERTSON G. G., CARD S. K. AND MACKINLAY J. D. 1993. Information visualization using 3D interactive animation. *Communications of the ACM*, 34(4):59-71.
- [16] ROBINSON D. editor 1998. *Neurobiology*. Springer, The Open University. ISBN: 3-540-63546-7
- [17] SALZMAN M., DEDE C., LOFTIN B. AND ASH K. 1998. VR's frames of reference: A visualization technique for mastering abstract information spaces. *Third International Conference on Learning Sciences*, pages 249-255
- [18] SHNEIDERMAN B, 1996. The eyes have it: A task by data type taxonomy for information visualizations. *Proc. IEEE Symposium Visual Language, VL*, pages 336-343, 3-6.

- [19] STUART L., WALTER M. AND BORISYUK R. 2002. Visualization of synchronous firing in multi-dimensional spike trains. *BioSystem* 67:265-279
- [20] Visualization of Inter-spike Association project website [WWW Document] <http://www.plymouth.ac.uk/infovis> (accessed 04 August 2003)
- [21] WONG P. C AND ANDREWS K. 2002 InfoVis 2002 Proceeding of the IEEE Symposium on Information Visualization
- [22] ZAHN C. T. 1971. Graph-Theoretical Methods of Detecting and Describing Gestalt Clusters. *IEEE Transactions on Computers* C-20:1, pages 68-86

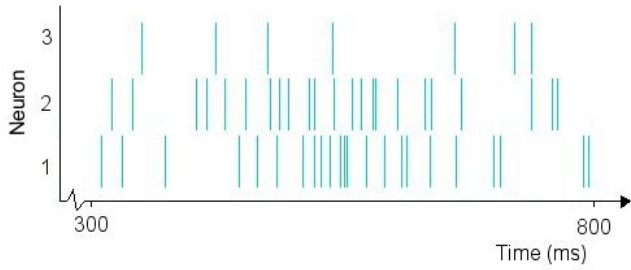


Figure 1: An example of a typical spike train recording for three neurons over a period of 500ms.

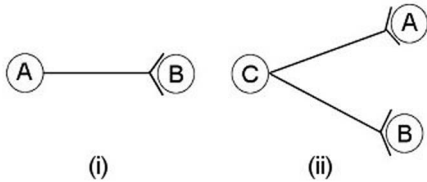


Figure 2: An example of (i) direct synaptic coupling and (ii) common input coupling

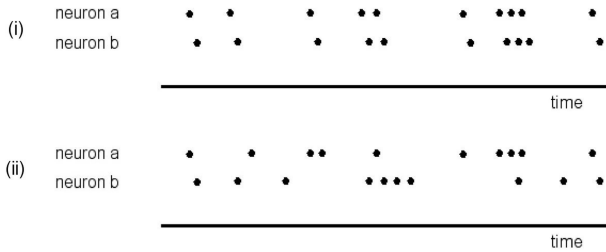


Figure 3: An example of two raster plots denoting (i) correlated spike trains, and (ii) spike trains with no apparent correlation.

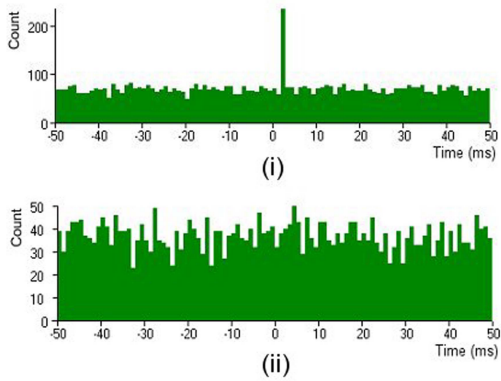


Figure 4: An example of a cross-correlogram from (i) two connected neurons and (ii) two detached neurons.

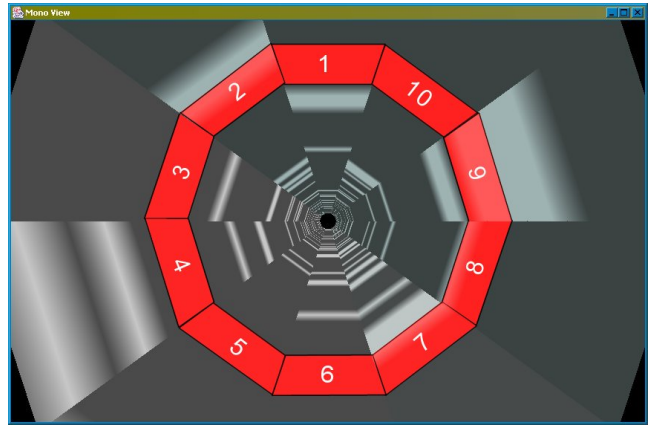


Figure 5: A snapshot of the Tunnel representation of the randomly generated dataset over 200ms.

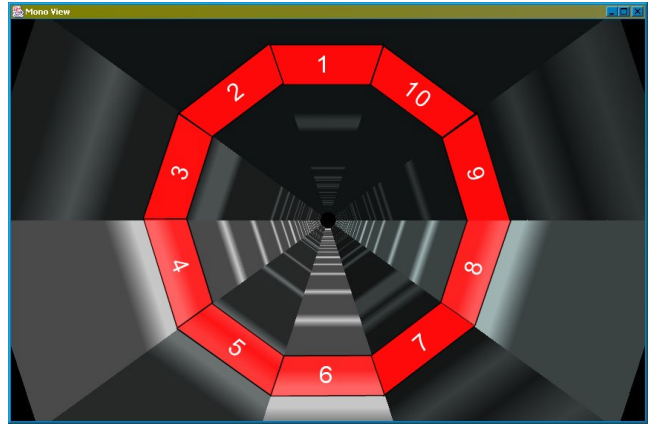


Figure 6 A snapshot of dimming (filtering functionality) within the Tunnel environment of a 200ms dataset.

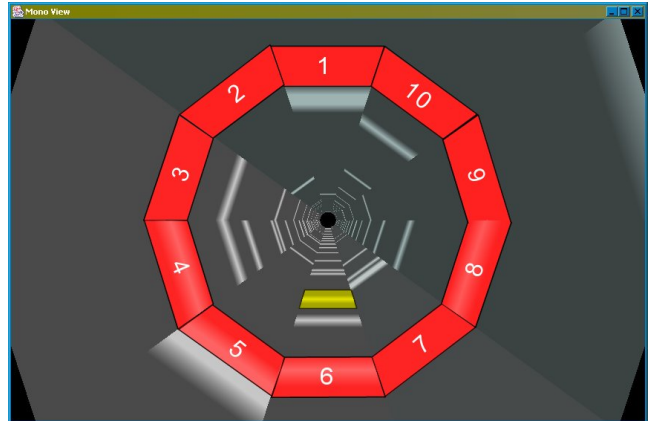


Figure 7 A snapshot of the Tunnel visualization for the unsorted dataset

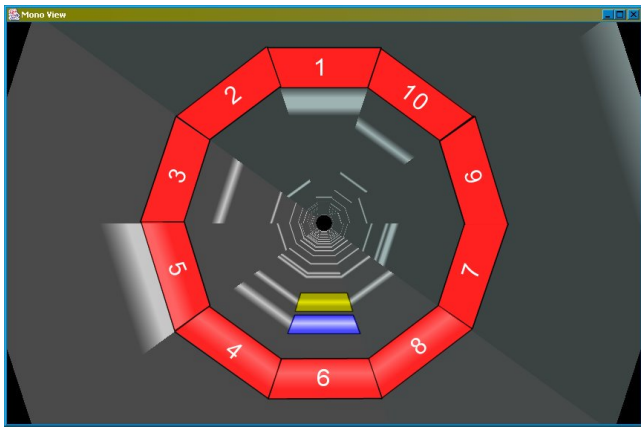


Figure 8 A snapshot of the Tunnel visualisation depicting sorting

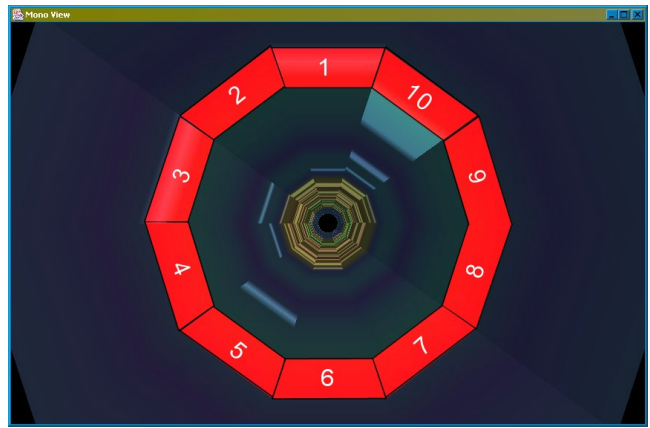


Figure 12 A snapshot of the Tunnel visualisation superimposed onto the CSV

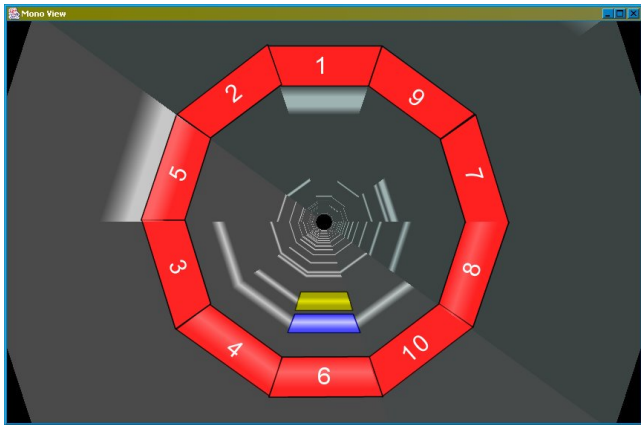


Figure 9: A snapshot of the Tunnel visualisation depicting progressive sorting

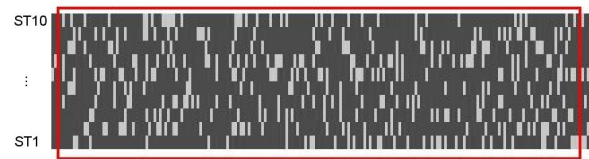


Figure 13 A zoomed section of the flat map denoting the ten spike trains, ST1 to ST10.



Figure 10 Colour coding for Coincidence Summary visualisation

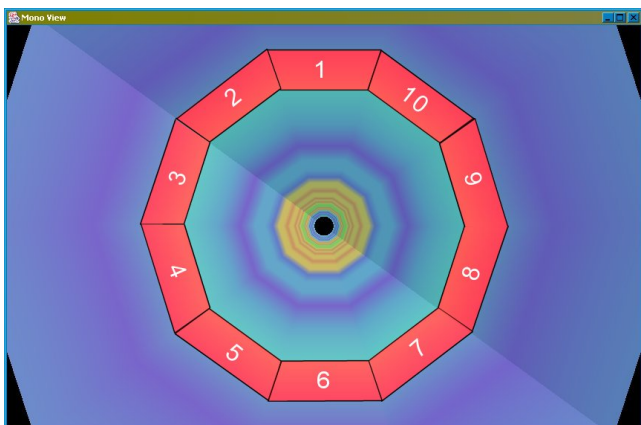


Figure 11 A snapshot of the Coincidence Summary visualisation for a 200ms dataset