Information Visualization (2004) 3, 245–256

www.palgrave-journals.com/ivs

The representation of neural data using visualization

Martin Walter[1]
Liz Stuart[1]
Roman Borisyuk[2,3]

[1]School of Computing, Communications and Electronics, University of Plymouth, Plymouth, Devon, U.K.; [2]Plymouth Institute of Neuroscience, University of Plymouth, Plymouth, Devon, U.K.; [3]Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow Region 142 290, Russia

Correspondence: Martin Walter, Room B310, Portland Square, School of Computing, Communications and Electronics, University of Plymouth, Plymouth, Devon PL4 8AA, U.K.
Tel: +44 1752 23 2620;
Fax: +44 1752 23 2540;
E-mail: mwalter@plymouth.ac.uk

Received: 5 September 2003
Revised: 15 February 2004
Accepted: 17 March 2004
Online publication date: 10 June 2004

Abstract
Currently, the focus of research within Information Visualization is steering towards genomic data visualization due to the level of activity that the Human Genome Project has generated. However, the Human Brain project, 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 visualization 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 Visualization; overview the data, zoom and filter data, data details on demand. The features of this visualization 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.
Information Visualization (2004) **3,** 245–256. doi:10.1057/palgrave.ivs.9500071

Keywords: Neural data; spikes; spike trains; analysis; visualization environment

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, these data are simultaneously recorded as multi-dimensional spike train data. Much of the research effort in this area is steered towards the principle of synchronization of neural activity.[1,2]

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 these data are required.

Information Visualization[3] is one of the fields of computer science that deals with innovations in the representation of vast quantities of data. This field is already recognized for its current and potential contribution to large-scale projects such at the Human Genome Project[4] and the Human Brain Project.[5] Information Visualization has also led the development of many useful visual representations for hierarchical and temporal data. These visualizations include techniques such as treemaps,[6] space-filling visualizations based on radial layouts[7] and the use of helical structures.[8] Other fundamental techniques such as parallel coordinates,[9] have already been successfully applied to the analysis of multi-dimensional spike train datasets,[10] as well as the innovative use of other basic geometric primitives.[11]

One of the fundamentals of the field of Information Visualization 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 visualizing large data sets, 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[12] becomes trivial. In addition, providing the user with different frames of reference can also help.[13]

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 organize 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[14] identified user interaction as one of the essential components of Information Visualization. Shneiderman also introduced the 'information–seeking mantra' that highlighted user requirements in this area. This mantra specified that users should have the capability to overview data, zoom and filter these data and to obtain details-on-demand. This mantra was widely adopted throughout the Information Visualization community as a basis for defining user requirements. In many cases, different visualizations are utilized to represent data at differing levels of detail. Thus, resulting in the creation of a number of different views of the same data. Ideally, these multiple views should be linked for consistency.[12,15,16]

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.

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 weak electrical impulses.

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 are 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 300 to 800 ms, 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 these data but it has also been established that each spike from a single neuron is identical.[17] 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 synchronization of spike trains and the coupling of neurons.

Coupling

Direct Synaptic coupling and *Common Input* coupling are the general cases of coupling where synchronization 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.

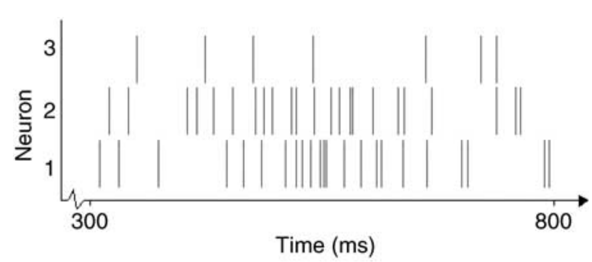

Figure 1 An example of a typical spike train recording for three neurons over a period of 500 ms.

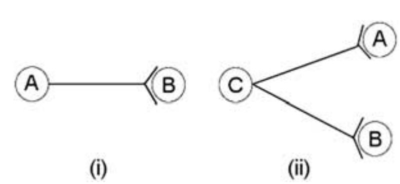

Figure 2 An example of (i) direct synaptic coupling and (ii) common input coupling.

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.

Multi-dimensional spike train data

One of the basic principles that underlie information processing in the brain is the principle of synchronization of neural activity. Research[1] indicates that the synchronization principle may be useful in devising various systems of information processing.

The current capability to record neural activity has led to the production of large quantities of experimental data. These data are in the form of multi-dimensional spike train recordings. Investigation of these 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.

Traditional methods of analysis

A number of methods exist to analyze 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.

The Raster Plot

The raster plot[18] 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 synchronized 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.

The cross-correlogram

The cross-correlogram[18,19] quantifies the synchronization 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'.

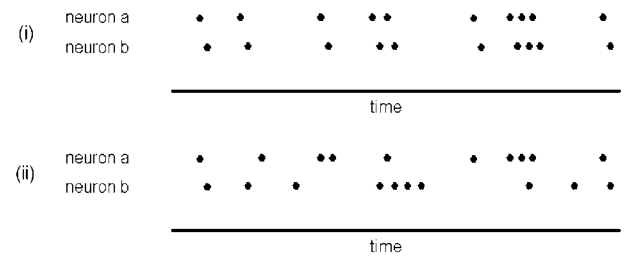

Figure 3 An example of two raster plots denoting (i) correlated spike trains, and (ii) spike trains with no apparent correlation.

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 4) 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[20] a correlation exists between the two trains. Consequently, it is likely that the two neurons are connected. In Figure 4(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 4(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

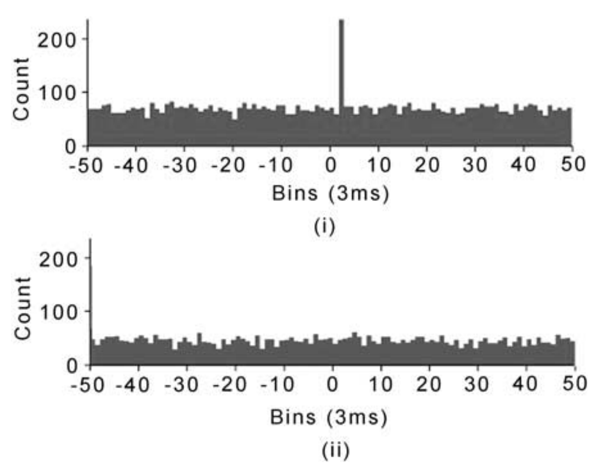

Figure 4 An example of a cross-correlogram from (i) two connected neurons and (ii) two detached neurons.

significantly sized neuronal assembly, numerous pairwise results would be generated. This large quantity of resultant data poses its own analysis problem, as all pairwise results must be analysed to understand any relationship between the underlying neurons.

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.[21–23] The gravity transform algorithm can be used to study the dependencies in firing of multi-dimensional spike trains. Recent work by Stuart *et al*[10] has enhanced the output from the original gravity transform algorithm using visualization 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.

An incremental approach to providing support

The ultimate system for this type of problem would be an intelligent adaptive system that built-up knowledge and expertise based on experimentation that provided positive and negative feedback to the system. This ideal system would receive a data set from the user as input and would produce an assembly, perhaps even two or three, with a certainty value or percentage associated with each topology suggested.

However, the main milestone separating the current support available for this type of analysis from this 'ideal' system is a reliable human-centred approach that accurately specifies the topology of a network of neurons from a multidimensional spike train data set.

This human-centred approach will not comprise a single, perfect representation that is suddenly discovered. On the contrary, this human-centred analysis approach will be a toolbox of many different representations, each with individual strengths and weaknesses. However, the combined functionality of this suite of tools will enable the experienced user to accurately identify the topology of a network of neurons from multi-dimensional spike train data sets as required by Neurophysiologists.

Hence, our work is focused on the design, implementation and testing of individual representations which, when coherently combined together, truly support the analysis of multi-dimensional spike train data. This approach is a unique and significant development in this area as it is based on the principles of Information Visualization and Software Engineering. This paper presents a new representation called the spike train 'Tunnel', that enables the user to investigate synchrony in the data sets.

The Tunnel environment

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

The Tunnel visualization

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

Each of the numbered horizontal bands, that comprise this Tunnel visualization, encodes the spike train of the corresponding neuron. The two 'end' bands, bands 1 and 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 data set. Synchrony is detected by perception of the position, intensity and frequency of light sources at different parts of the Tunnel.

In the Tunnel visualization, 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 minimize the possible side-affects of disorientation during navigation, motion is restricted to being along the Tunnels length. Thus, the user is restricted to forward and reverse motion.

Filtering data

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

Figure 6 illustrates filtering of another 200 ms data set. In this data set, 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

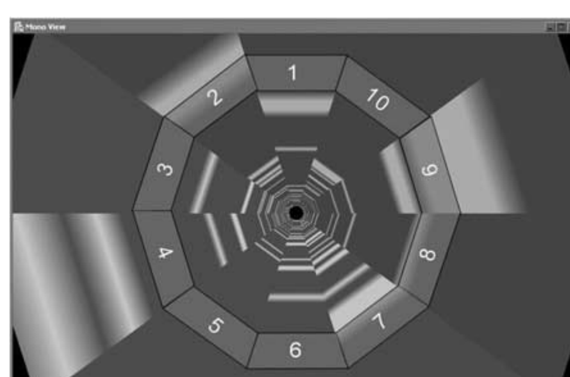

Figure 5 A snapshot of the Tunnel representation of the randomly generated data set over 200 ms.

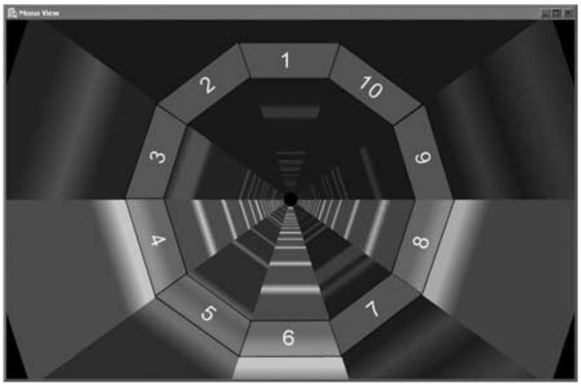

Figure 6 A snapshot of dimming (filtering functionality) within the Tunnel environment of a 200 ms data set.

designed to enable investigators to highlight spike trains of interest while maintaining context within the data set.

Coincidence sorting

As stated previously, in section on ‘Spikes and spike trains’, 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 200 ms data set, based on an assembly of 10 neurons, was generated. In this assembly, neurons three and 10 fire every 12 ms and neuron eight fires every 7 ms. To illustrate the progressive sorting feature of the Tunnel, neurons four and six fire every 7 and 12 ms.

The Tunnel visualization of this unsorted data set is illustrated in Figure 7. Note the first coincident spikes (at 12 ms) on spike train three and four, and the following non-coincident spike (at 14 ms), 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 visualization 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 over-rides 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 10 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 10 displaces spike train eight.

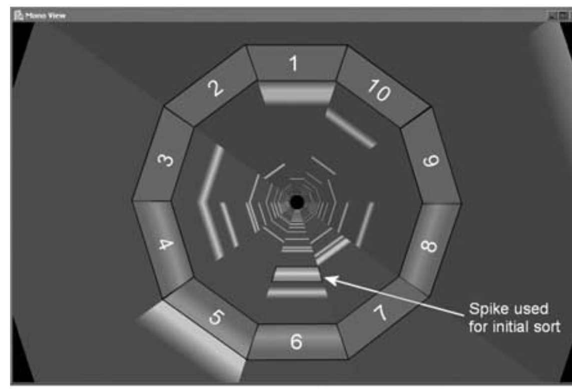

Figure 7 A snapshot of the Tunnel visualization for the unsorted data set.

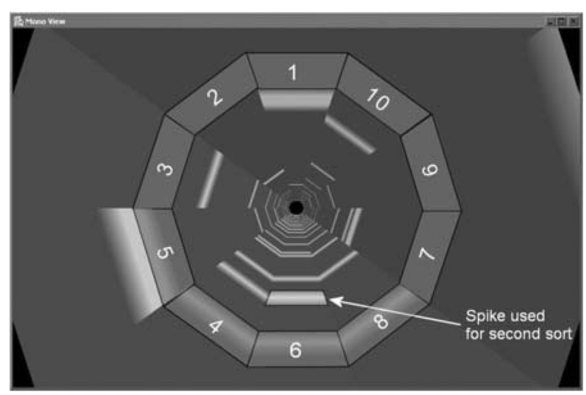

Figure 8 A snapshot of the Tunnel visualization depicting sorting.

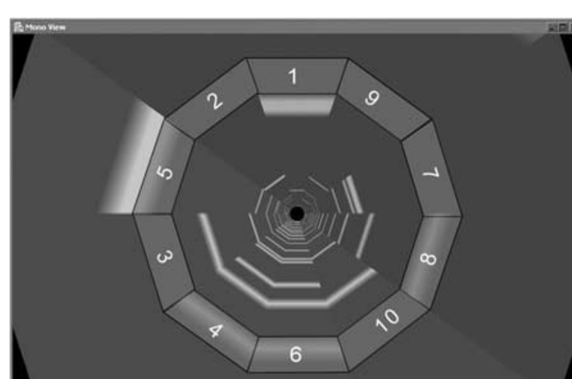

Figure 9 A snapshot of the Tunnel visualization depicting progressive sorting.

Coincidence summary

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

Each spike train in the data set 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 ith element of this intermediate summary array is equal to the sum of the ith 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 data set. This array is used to create the Coincidence Summary Visualization.

The data set used to create the visualization in Figure 11 comprised 10 spike trains each lasting 200 ms. Each of the spike trains were created by appending four, 50 ms 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 200 ms data set, the final summary array is made up of 67 elements where there are 66 3 ms bins and one 2 ms bin. Due to the way in which the data set 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.

These data are encoded for the Coincidence Summary Visualization using colour based on the Hue map shown in Figure 10.

The Coincidence Summary Visualization (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 visualization 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 visualization. Due to the relatively high values in this segment, it is represented by mainly red and yellow hues.

Combining them together

In addition to viewing the CSV, it is also possible to superimpose the Tunnel visualization 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 visualization of this data set.

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

Note that in this combined visualization, 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.

Figure 10 Colour coding for Coincidence Summary visualization.

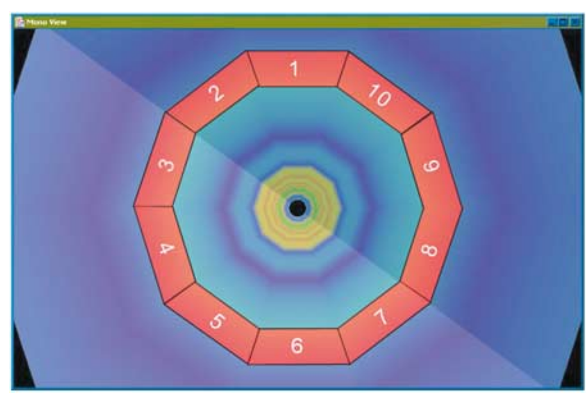

Figure 11 A snapshot of the Coincidence Summary visualization for a 200 ms data set.

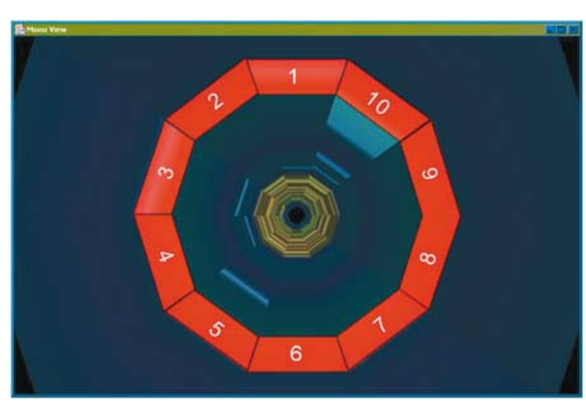

Figure 12 A snapshot of the Tunnel visualization superimposed onto the CSV.

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 earlier, however, it has additional functionality. This functionality enables the user to select a 1,000 ms subset of the data for analysis. It is this selected 1,000 ms subset that is subsequently displayed in greater detail within the Tunnel representation.

To define the required subset the user interacts directly with the Flat Map resulting in a red boundary line indicating the current portion to be viewed in detail in the tunnel. Note that this detailed viewing is a dynamic process, effectively the user flies through the segment. A green marker on the Flat Map indicates to the user their progression within the segment.

Figure 13, depicts the flat map for the demonstration data set used in section on 'Coincidence sorting'. Although the tunnel is capable of displaying 1,000 ms at any given time, this demonstration data set is only 200 ms long in order to clarify the concepts presented. Thus, it is the majority of this 200 ms data set that is shown in the tunnel representation shown in Figure 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.[14] Thus, the user should be able to easily return to previous states of the visualization.

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

Empirical testing

In this section, the results of two trials are presented. Each of the trials use a data set that was generated using an advanced integrate and fire model of neurons with particular coupling between elements.[1] The parameters of this model were chosen to imitate the general neurophysiological characteristics of cortical neurons. All connection strengths in these simulations were chosen to be positive.

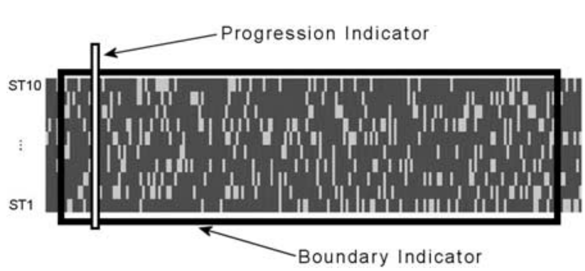

Figure 13 The Flat Map representation of the data segment depicted in the Tunnel representation of Figure 7.

Trial one

In this trial, an assembly of 10 neurons, as shown in Figure 14, was simulated for a period of 20,000 ms. The mean inter-spike interval (ISI) was 70 ms, the standard deviation of the ISI of the data set was 53 ms and its coefficient of variation was 0.76.

Clustering The random ordering of spike trains within the Tunnel poses a number of problems when attempting to analyse their interconnection(s). Thus, it is preferable to have trains with a high temporal correlation adjacent to each other. To achieve this, the stripes of the tunnel (representing each spike train) are reordered using a clustering algorithm.

A range of different mathematical clustering algorithms were analysed to perform the clustering of the spike trains. These algorithms included a nearest-neighbour algorithm (the minimum of measures between objects in two groups), a furthest neighbour algorithm (the maximum of measures between objects in two groups), and a centroid clustering algorithm.[24] The most suitable algorithm is the furthest neighbour method as this algorithm creates tight clusters between objects and all objects inside clusters have limited dissimilarity.

The smaller group of trial one A Tunnel representation for the trial one data was compiled and subsequently sorted using the selected clustering algorithm. This is shown in Figure 15, in which the trains 1, 3, 5 and 7 have been highlighted, for the segment of the data set between 3,000 and 4,000 ms. This is a useful feature of the tunnel environment.

Observe the simultaneous spiking activity on trains 3, 5 and 7. Further, note that these spikes are preceded by a spike on train 1. It was also possible to see this spiking pattern occurring further along the tunnel. Although it is not easy to see this from Figure 15, it is possible to see a pattern of spikes around the second and third spikes on train 1. Subsequently, it is reasonable to infer that a functional relationship exists between these neurons. Furthermore, since the spikes on train 1 precede those of the other trains, it is likely that neuron 1 is a common

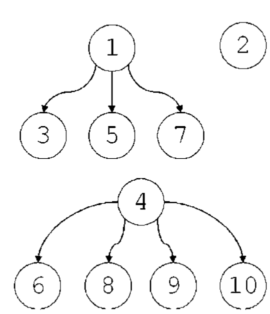

Figure 14 The assembly of 10 neurons used to generate the spike train data sets used in trial one.

input to neurons 3, 5 and 7. This is confirmed by referring back to Figure 14.

The larger group of trial one Figure 16 shows the same snapshot as Figure 15, however, in Figure 16 the spike trains 4, 6, 8, 9 and 10 are illuminated. Observe the simultaneous spiking activity on trains 6, 8, 9 and 10. Furthermore, note that these spikes are also preceded by a spike on train 4. Again, it is reasonable to infer that neurons 6, 8, 9 and 10 are connected and that neuron 4 is a common input to them. This is also confirmed by referring back to Figure 14.

Unconnected neuron By observing train 2 along a number of sections in the tunnel, it is possible to see that there is not notable synchrony with any other trains. This is due to the fact that neuron 2 is unconnected in the assembly in Figure 14.

Summary Overall, the simple assembly depicted in Figure 14 can be deduced solely using the Tunnel representation.

Trial two

In this trial, an assembly of 15 neurons, as shown in Figure 17, was simulated for a period of 20,000 ms. The mean inter-spike interval (ISI) was 75 ms, the standard deviation of the ISI of the data set was 53 ms and its coefficient of variation was 0.7. The ISI histogram for spike train 6 is shown in Figure 18(a) and its autocorrelation is shown in Figure 18(b). The raster plot of the first 3,000 ms of this data set is also shown in Figure 19. The

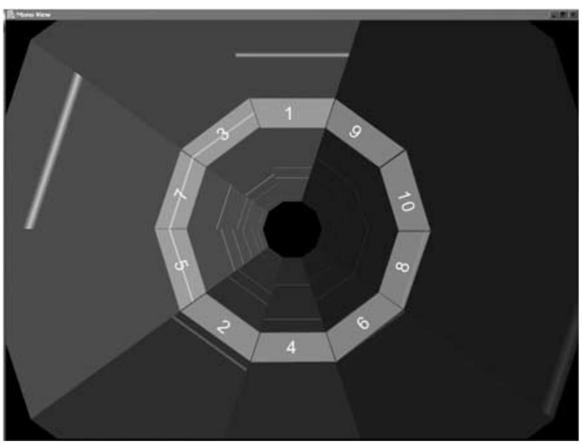

Figure 15 A snapshot of the Tunnel representation depicting the 10 spike trains of trial one, in which the order of the trains is defined by the clustering algorithm and spike trains 1, 3, 5 and 7 are highlighted.

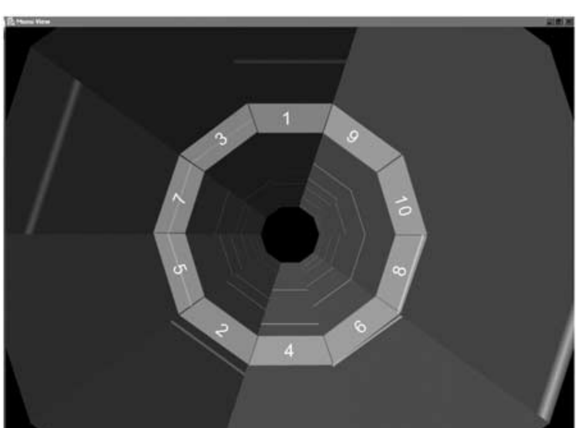

Figure 16 A snapshot of the Tunnel representation depicting the 10 spike trains of trial one, in which the order of the trains is defined by the clustering algorithm and spike trains 4, 6, 8, 9 and 10 are highlighted.

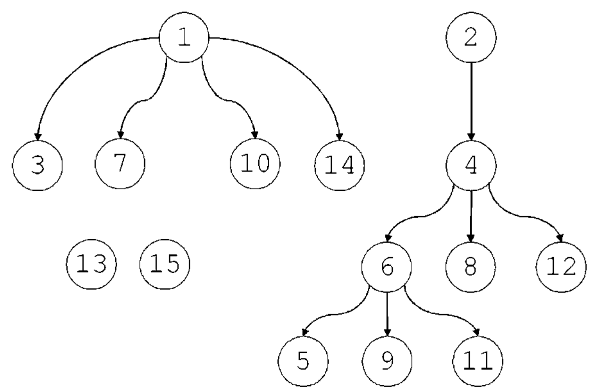

Figure 17 The assembly of 15 neurons used to generate the spike train data sets used in trial two.

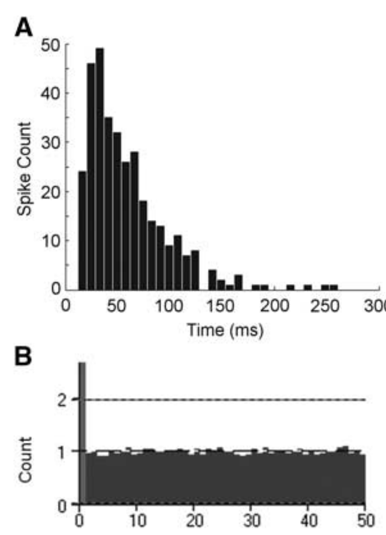

Figure 18 This figure depicts (a) the ISI histogram of spike train 6 and (b) the autocorrelation of spike train 6 from the data set used in trial two.

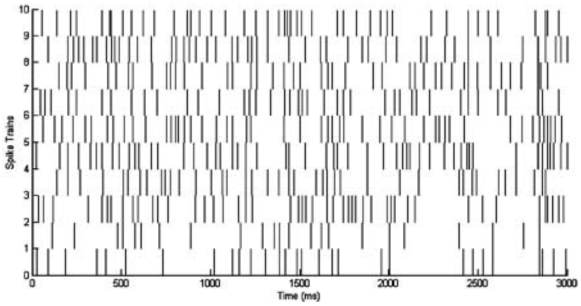

Figure 19 Raster A raster plot depicting the first 3,000 ms of the data set generated for trial two.

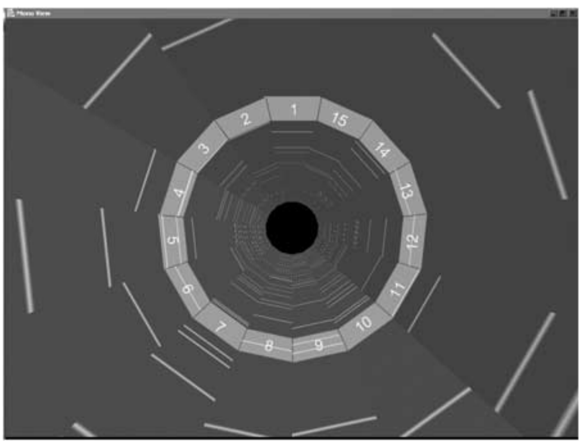

Figure 20 A snapshot of the Tunnel representation with 'random' ordering of the 15 spike trains of the second trial.

initial spike train tunnel for these 15 spike trains is shown in Figure 20, in which the order of the stripes is based on the default order of the spike trains in the data file. Recall that the user sees up to 1,000 ms of the data set at a time in this representation, hence, Figure 20 displays a snapshot of the first 1,000 ms of the data set.

Note, that from this random ordering of the trains it is difficult to extract useful information. However, when viewing several snapshots, or 'flying' through this section of the tunnel, it is possible to identify similarity between trains.

Figure 21 shows the same snapshot of the Tunnel as Figure 20 except that in this figure the trains 5, 6, 9 and 11 are highlighted. Note the spike-pair, the first two spikes, on train 6. This spike-pair is followed by a similar spike-pair on trains 5, 9 and 11, each with different short delays. Thus, it is possible to infer that a relationship exists between spike trains 5, 6, 9 and 11. This synchrony of spiking is attributable to the connections from neuron 6 to neurons 5, 9 and 11; the lower sub-group in the assembly shown in Figure 17. In Figure 21, the trains 5, 6, 9 and 11 have been highlighted to aid clarity in this diagram. Even though identification of relationships in this manner is possible, it is difficult and time consuming.

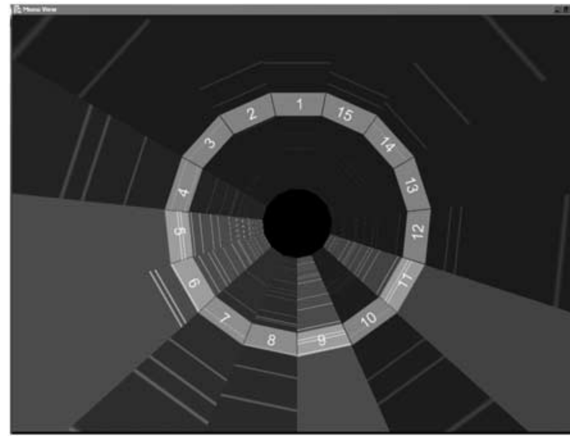

Figure 21 A snapshot of the Tunnel representation with 'random' ordering of the 15 spike trains of the second trial, in which trains 5, 6, 9 and 11 are highlighted.

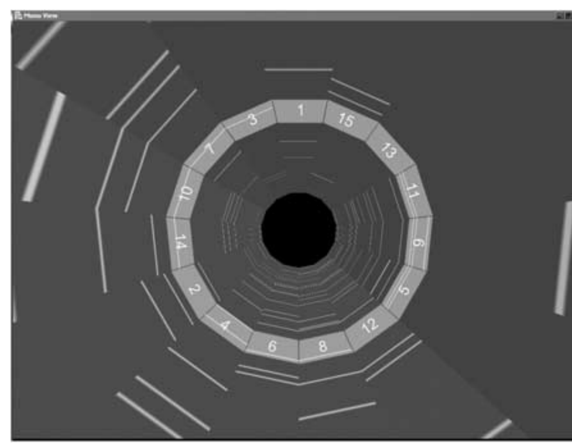

Figure 22 A snapshot of the Tunnel representation depicting the 15 spike trains of trial two, in which the order of the trains is defined by the clustering algorithm.

The result of clustering the spike trains of this data set, and subsequently, reordering the tunnel stripes, is shown in Figure 22.

The smaller group of trial two Subsequent to clustering, it is easier to identify trains that are related. In Figure 23, spike trains 3, 7, 10 and 14 are highlighted. Note how the clustering algorithm has assigned trains 3, 7, 10 and 14 adjacent to one another. Also note the coincident spikes on all four trains, mid-way down through the tunnel. Additionally, the occurrence of synchronous spikes on trains 3, 7, 10, 14, was observed in other parts of the tunnel. As these neurons tend to spike coincidently, it is reasonable to suggest that these neurons are all receiving a similar input. Further note that it is possible to see this synchronous firing is commonly preceded by the occurrence of a spike on train 1 as shown in Figure 23.

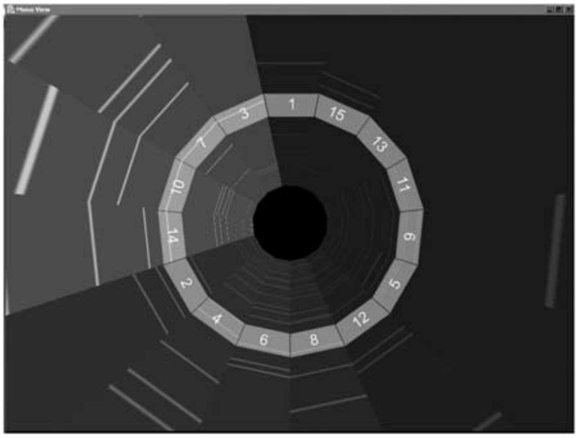

Figure 23 A snapshot of the Tunnel representation depicting the 15 spike trains of trial two, in which the order of the trains is defined by the clustering algorithm and spike trains 3, 7, 10 and 14 are highlighted.

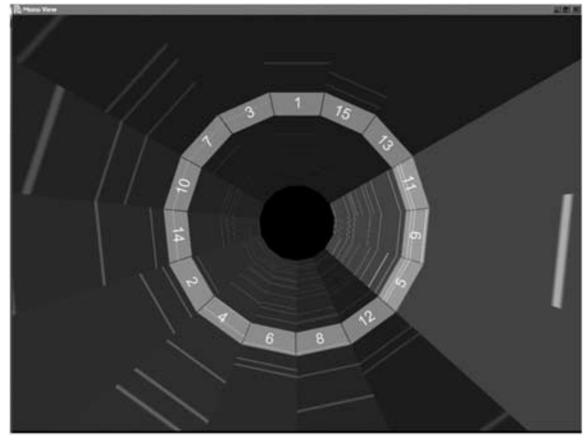

Figure 25 A snapshot of the Tunnel representation depicting the 15 spike trains of trial two, in which the order of the trains is defined by the clustering algorithm and spike trains 5, 9 and 11 are highlighted.

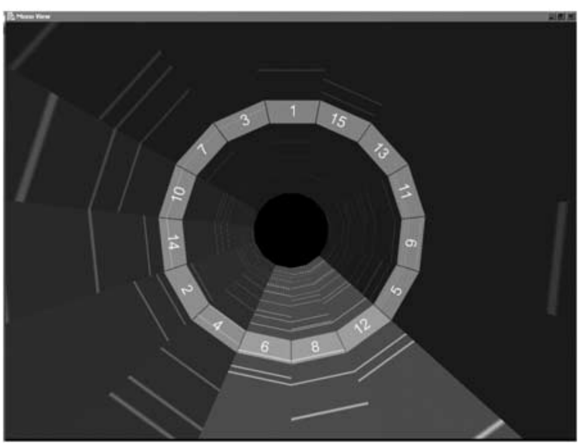

Figure 24 A snapshot of the Tunnel representation depicting the 15 spike trains of trial two, in which the order of the trains is defined by the clustering algorithm and spike trains 6, 8 and 12 are highlighted.

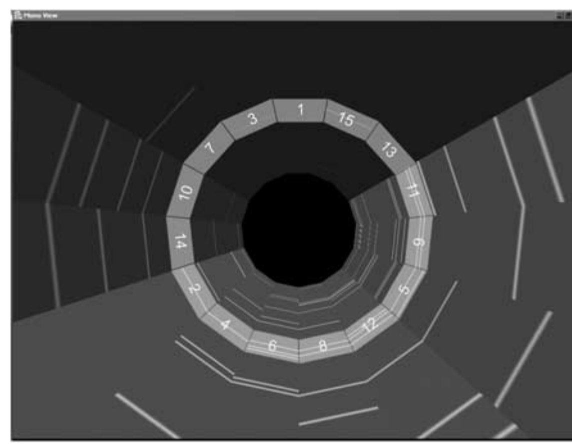

Figure 26 A snapshot of the Tunnel representation depicting the 15 spike trains of trial two, in which the order of the trains is defined by the clustering algorithm and spike trains 2, 4, 6, 8, 12, 5, 9 and 11 are highlighted.

Indeed, from Figure 17, it is noted that spike trains 3, 7, 10 and 14 are generated by a group of common input neurons; namely neurons 3, 7, 10 and 14 which receive common input from neuron 1.

The larger group of trial two In Figure 24, spike trains 6, 8 and 12 are highlighted to make it easier for the user to view the relationship between these trains. Closer examination also shows that this synchronous firing is commonly preceded by the occurrence of a spike on train 4. Thus, it is likely that neuron 4 is common input to neurons 6, 8 and 12.

In Figure 25, spike trains 5, 9 and 11 are highlighted and it is very clear that a relationship exists between these trains. From Figure 26, it is possible to note that the synchronous spiking in trains 5, 9 and 11 is preceded by a spike in train 6. As it is already known that neuron 4 is a common input to neurons 6, 8 and 12, it is deduced that this is a three level hierarchy of neurons.

Inspection of train 2 using Figure 26 and additional snapshots of the tunnel also revealed that a relationship exists between trains 2 and 4. Therefore, it is possible that neuron 2 connects to neuron 4 which subsequently connects to neurons 6, 8 and 12.

However, these relationships are not as clear from the Tunnel representation and it is likely that an additional representation, called the Correlation Grid[11] which is accurate in the identification of hierarchies, would be required to reinforce this aspect of the investigation.

Unconnected neurons With the clustered order of the spike trains in the Tunnel representation, it is clear that spike trains 13 and 15 have no correlation with any of the other trains in this data set. This is notable from the lack of temporal relationships shown for these spike trains in Figure 22, as opposed to the other groups.

Summary of trial two From the clustered version of the Tunnel representation, it is possible to extract the information relating to the inter-neuron and inter-group relationships. However, it may be necessary, to support such an inferred definition of a neural assembly with additional evidence such as data from the correlation grid technique. This will become more apparent from subsequent empirical trials.

Future work

The work presented in this paper is part of an Information Visualization project, at the Centre for Interactive Intelligent Systems, University of Plymouth, called Visualization of Inter-Spike Associations (VISA).[25] Specifically, this paper has described an innovative visualization technique for the analysis of multi-dimensional spike train data called the Tunnel visualization. Testing of this representation to date has concentrated on data sets of approximately 20 spike trains over 20,000 ms. In these cases, the assemblies of the neurons that produced the spike train data were known to the investigator. This initial feasibility testing has been beneficial, reinforcing the efficacy of the method. However, this testing has also highlighted a number of limitations to the current version of the Tunnel representation. Each of these weaknesses is identified and future plans to strengthen these areas are described.

In this paper, all of the snapshots of the Tunnel have spike train numbers appended to aid identification. This is a trivial problem which will be addressed by simply developing a subtle labelling system that does not infer with data representation. A number of identification methods exist and currently two are under consideration. To identify a specific train, the user could "mouse over" the train to see its number. Alternatively, a semi-transparent overlay of the train number could be developed. Moreover, it is likely to be more appropriate to implement both methods, allowing the user to select between them.

Currently, the 'flat map' overview of the Tunnel identifies the section of the data that the user has chosen to zoom in upon, but it does not track the user's position. This is not helpful to the user and thus, an exocentric frame of reference[13] will be developed for the Tunnel, to enable the user to identify their current position within the Tunnel.

Currently, the user may only select one, continuous, area of the tunnel to zoom. However, this representation does not provide the user with the facility to zoom in on a number of disjoint sections of the Tunnel, simultaneously. This is problem as the user is largely interested in synchrony and repeated patterns of activity within the whole data set. Thus, the user may need to view and compare a number of sections (subsets of the whole data set) of interest. An efficient method of shrinking areas of less interest in order to focus on areas of greater interest will be investigated.

Hitherto, the weaknesses of the Tunnel representation described are changes which need to be implemented to improve the general use of this representation specifically.

However, there is a more significant issue that also needs to be addressed. It is understood that the human-centred analysis system, see section on 'An incremental approach to providing support', that will provide the overall combined functionality required by Neurophysiologists, will comprise a range of representations. However, it also well established that the usefulness of multiple views is significantly increased if views are linked.[16] As the Tunnel representation is an integral part of the VISA system, its contribution to the overall analysis problem will be significantly advanced if it is used in connection with the other representations of VISA.

Thus, it is envisaged that future effort will be focused on meaningfully linking the Tunnel representation with other representations within the VISA system. In the future, it is likely that the VISA project will not only link representations together but it will provide a visual programming interface to enable users to "build" their own analyses interactively.

Finally, with respect to empirical testing, further testing is currently underway to evaluate the usability of this visualization method with larger numbers of neurons. Subsequent to this, significant testing will begin on real data recorded by Neurophysiologists who are keen to collaborate on this project.

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).

References

1 Borisyuk RM, Borisyuk GN. Information coding on the basis of synchronisation of neuronal activity. *BioSystems*, 1997; **40**: 3–10.

2 Fries P, Neuenschwander S, Engel AK, Goebel R, Singer W. Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience*, 2001; **4**: 194–200.

3 Wong PC, Andrews K. *Proceeding's of the IEEE Symposium on Information Visualization 2002*. IEEE Computer Society: New York, 2002.
4 HGP Human genome project [WWW Document]. http://www.ornl.-gov/TechResources/Human_Genome/ (accessed 04 August 2003).
5 HBP Human brain project web site [WWW Document]. http://www.nimh.nih.gov/neuroinformatics/index.cfm (accessed 04 August 2003).
6 Shneiderman B. Tree visualization with treemaps: a 2-d space-filling approach. *ACM Transactions on Graphics*, 1991; **11**: 92–99.
7 Stasko J, Zhang E. *Focus+contect display and navigation techniques for enhancing radial, space-filling hierarchy visualizations*. IEEE Symposium on Information Visualization 2000, IEEE Computer Society Press: New York, 57–65.
8 Hicks M. *A helix metaphor for customer behaviour visualisation*. IEEE International Conference on Information Visualization 2001 (London, UK, 2001), IEEE Computer Society Press: New York, 22–28.
9 Inselburg A, Dimsdale B. *Parallel coordinates: a tool for visualising multidimensional geometry*. Conference on Visualization 1990; 361–378.
10 Stuart L, Walter M, Borisyuk R. Visualization of synchronous firing in multi-dimensional spike trains. *BioSystem*, 2002; **67**: 265–279.
11 Walter M, Stuart L, Borisyuk R 2003 *Spike train correlation visualization*. IEEE International Conference on Information Visualization 2001 (London, UK, 2001), IEEE Computer Society Press: New York, 555–560.
12 Roberts JC. *On encouraging multiple views for visualization*. IEEE International Conference on Information Visualization 1998 (London, UK, 1998), IEEE Computer Society Press: London, 8–14.
13 Salzman M, Dede C, Loftin B, Ash K. *VR's frames of reference: a visualization technique for mastering abstract information spaces*. Third International Conference on Learning Sciences 1998; 249–255.
14 Shneiderman B. *The eyes have it: a task by data type taxonomy for information visualizations*. IEEE Symposium on Visual Languages 1996 (Las Alamos, USA, 1996), IEEE Computer Society Press: Chicago, 1996; 336–343.
15 Baldonado MQW, Woodruff A, Kuchinsky A. Guidelines for using multiple views in information visualization. *Advanced Visual Interfaces*, 2000; 110–119.
16 North C. *Multiple views and tight coupling in visualization: a language, taxonomy, and system*. Workshop of fundamental Issues in Visualization 2001; 626–632.
17 Robinson D (Ed). *Neurobiology*. Springer: The Open University, 1998.
18 Awiszus F. Spike train analysis. *Journal of Neuroscience Methods*, 1997; **74**: 155–166.
19 MuLab [WWW Document]. http://mulab.physiol.upenn.edu/index.html (accessed 04 August 2003).
20 Brillinger DR. Confidence intervals for the crosscovariance function. *Selecta Statistica Canadiana*, 1979; **V**: 1–16.
21 Baker SN, Gerstein GL. Improvements to the sensitivity of the gravitational clustering for multiple neuron recordings. *Neural Computation*, 2000; **12**: 2597–2620.
22 Gerstein GL, Aertsen AM. Representation of cooperative firing activity among simultaneously recorded neurons. *Journal of Neurophysiology*, 1985; **54**: 1513–1528.
23 Gerstein GL, Perkel DH, Dayhoff JE. Cooperative firing activity in simultaneously recorded populations of neurons: detection and measurement. *Journal of Neuroscience*, 1985; **5**: 881–889.
24 Bouguettaya A, Le Viet Q. Data clustering analysis in a multidimensional space. *Information Sciences*, 1998; **112**: 267–295.
25 Visualization of Inter-spike Association project website [WWW Document]. http://www.plymouth.ac.uk/infovis (accessed 04 August 2003). Author Biographics