ELSEVIER

Journal of Neuroscience Methods 70 (1996) 141–152

JOURNAL OF NEUROSCIENCE METHODS

Measure and statistical test for cross-correlation between paired neuronal spike trains with small sample size

Xuesi M. Shao [a,*], Yang Tsau [b,1]

[a] *Systems Neurobiology Laboratory, Department of Physiological Science, University of California at Los Angeles, Los Angeles, CA 90095-1527, USA*
[b] *Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT06510, USA*

Received 18 October 1995; revised 20 June 1996; accepted 20 June 1996

Abstract

Recent development of multi-unit recording techniques such as optical recording and multi-electrode arrays makes it possible to record neuronal activities from tens or hundreds of neurons simultaneously. To analyze functional connections between these neurons, cross-correlation analysis has been most commonly applied to the hundreds to thousands of pairs of these neurons. However, conventional cross-correlation data needs statistical tests for significance especially when the sample size of recorded spike trains is small. Here, a multiple hypergeometric model based on a transformation of the cross-correlogram data to a $2 \times J$ table has been suggested. The exact p value for significance can be obtained by the generalized Fisher's method with small sample size and a cross-correlation coefficient for the strength of cross-correlation can be obtained based on the R-square analogue for nominal data. For large sample size, χ^2 test can be applied based on the same transformation. Examples of real spike train data set and simulation show that the methods are applicable to the data of multi-unit activity with only tens of spikes. These methods are especially useful when thousands of cross-correlograms need to be screened quickly and automatically.

Keywords: Multi-unit activity; Spike train analysis; Cross-correlation coefficient; Cross-correlation statistical test; Generalized Fisher's method; Exact p value; Multiple hypergeometric model

1. Introduction

Recent development of multi-unit recording techniques such as optical recording (London et al., 1987; Zecevic et al., 1989; Parson et al., 1991; Wu et al., 1994; Tsau et al., 1994) and microelectrode array (Villa and Abeles, 1990; Gochin et al., 1991; Lindsey et al., 1992; Nelken et al., 1994; Wilson and McNaughton, 1994; Vaadia et al., 1995; Ylinen et al., 1995) enables us to record neuronal activities from several or hundreds of neurons simultaneously. Questions regarding which neurons are functionally related to other neurons, and how strongly or how directly they are related are often raised. One way to study these issues is to analyze the timing relation between the multiple spike trains. This time correlation may have some implications for the mechanism underlying it, e.g., the existence and nature of their synaptic connections. The cross-expectation density or cross-correlation function (Perkel et al., 1967; Moore et al., 1970) is used extensively for analyzing the time relation between two spike trains. A number of methods have been developed to test and quantify peaks or troughs of the cross-correlograms in order to detect and measure the strength of the cross-correlation (Sears and Stagg, 1976; Graham and Duffin, 1981; Abeles, 1982; Aertsen and Gerstein, 1985; Epping and Eggermont, 1987; Melssen and Epping, 1987; Shao and Chen, 1987; Palm et al., 1988; Eggermont, 1992, 1994; Tsau et al., 1994). These methods are bin by bin based. They take each bin of the cross-correlogram as an independent Poisson variable (Abeles, 1982). For small to medium sample sizes, the

* Corresponding author. Tel.: (310) 825-1586; Fax: (310) 206-9184; E-mail: mshao@ucla.edu

[1] Present address: Georgetown Institute of Cognitive and Computational Sciences, Georgetown University Medical School, 3970 Reservior Rd. N.W., Washington, DC 20007, USA. Tel.: (202) 687-1617; Fax: (202) 687-0617.

PII S0165-0270(96)00112-4

95% confidence limit would be too wide since the variance of a Poisson variable is equal to its mean. Some modifications of these procedures take normal approximation for Poisson variable (Sears and Stagg, 1976; Graham and Duffin, 1981; Abeles, 1982), they work well only when the sample sizes are very large. In the situation where one or both of the paired spike trains have low discharge frequencies with limited recording time, the methods would be inaccurate or infeasible. Relatively short recording periods are often obtained in optical recording from the nervous system (Tsau et al., 1994). In addition, most of these methods for testing the cross-correlation consider the height

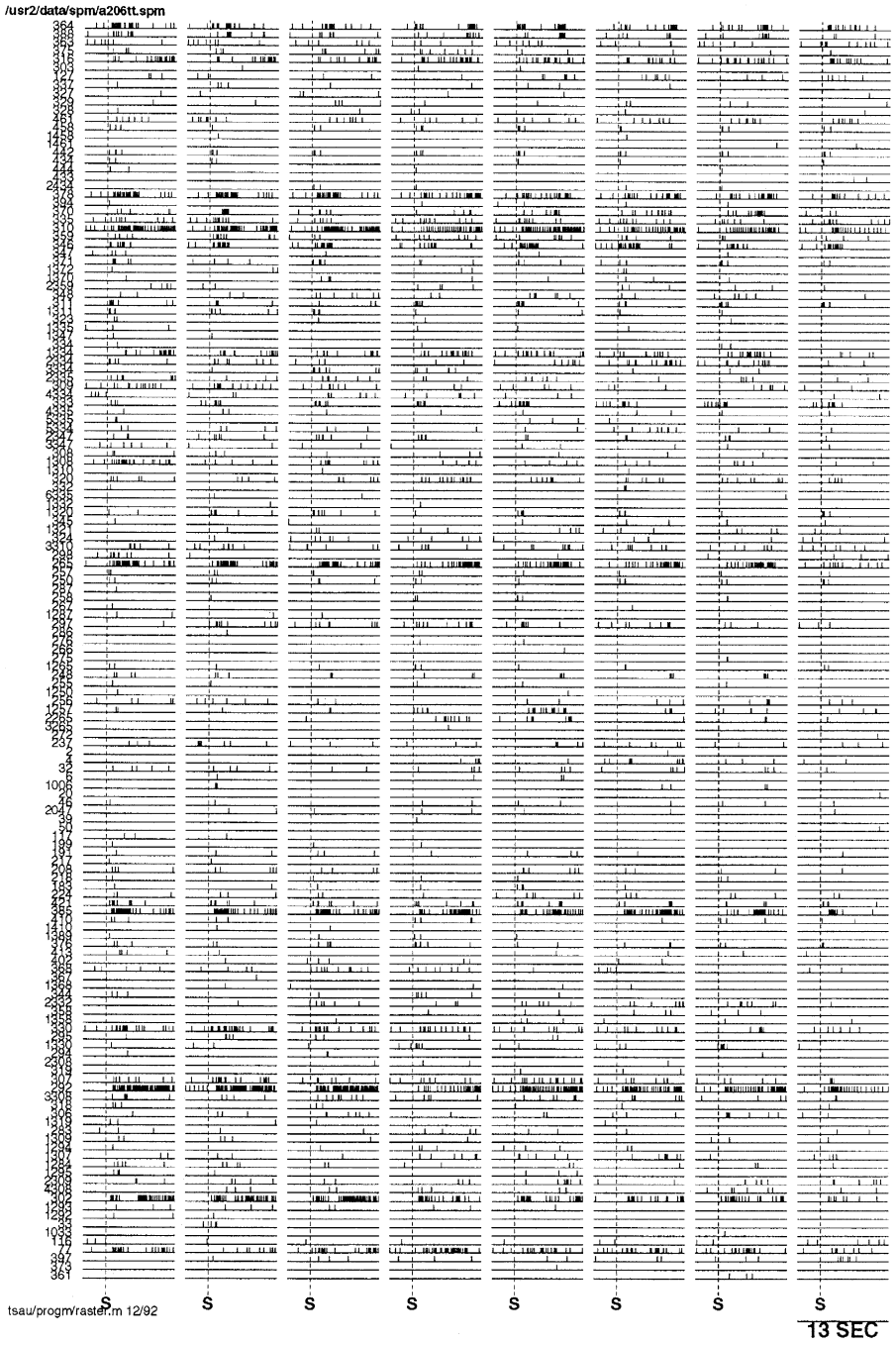

Fig. 1. Raster diagrams of action potential activity detected in an abdominal ganglion during 8 presentations of a light touch separated by 10–15 min intervals. Each vertical panel shows a 13 s recording. The numbers to the left of each line are identification numbers derived from their signal locations on the photodiode array which was used for the optical recording (for more details see Wu et al., 1994; Tsau et al., 1994). The light touch stimuli are indicated by the vertical dashed lines with 'S's. Experiment no. A206.

of the peaks or troughs but do not take the width into account except in the methods of peak integration based on the normalized cross-covariance function (NCCVF) suggested by Shao and Chen (1987). For example, if several neighboring bins are close to the confidence limit, they should be considered significant but most test procedures do not detect that. While measures of cross-correlation based on the area of peaks of certain scaled cross-correlogram have been proposed (Levick et al., 1972; Shao and Chen, 1987; Epping and Eggermont, 1987). Furthermore, all these bin by bin test methods have the multiple comparison problem for each cross-correlogram. The type I error rate for the correlogram is much higher than the significance level α assigned for each bin (Miller, 1981; also see Section 4).

Fig. 1 shows 149 spike trains recorded optically from the abdominal ganglion of *Aplysia* (Wu et al., 1994). There are $149!/[2!(149-2)!] = 11\,026$ possible pairs and we want to know if any pairs are correlated and, if so, how strongly they are correlated. Quite often it is difficult to tell whether there is a correlation between pairs by looking at the cross-correlation histograms or the raw data. Fig. 2 shows two NCCVF histograms for paired neural spike trains. For the top pair (no. 311 vs. no. 364), where no. 311 has 50 spikes and no. 364 has 195 spikes, there is an apparent mode around time point zero in the NCCVF (cross-correlogram shown in Fig. 3, no. 24), we can measure the strength of cross-correlation by the integration of the mode (Shao and Chen, 1987), and no statistical test is needed. However, for the pair no. 434 vs. no. 1311 (lower panel, cross-correlogram shown in Fig. 3, no. 8), the measurement would be invalid because of the small number of spikes in the trains (10 for neuron no. 434 and 29 for no. 1311). We are not able to say that there is correlation between them with any confidence although we do see a peak in the figure. In this situation, we first need a statistical test which works well with a small sample size and second, would like a summary measure for the strength of the association.

In this paper, we develop procedures for measuring and testing the cross-correlation by transferring the paired neural spike train data to a 2 by J contingence table. Taking each column of the table as a binomial random variable, we can obtain the overall probability of the table. Then, a test based on the generalized Fisher's exact test and a measure for cross-correlation based on the R-square for nominal data are suggested. Application of these proposed methods to the data sets of multi-unit recording and computer simulation illustrates their performances in terms of sensibility, reliability, error rates and the relationship between the measured correlation and the specified 'true' correlation. We also applied the proposed procedures to the 'shift predictor' (Perkel et al., 1967; Epping and Eggermont, 1987) of the same data set in order to identify the co-activation effect by the stimuli.

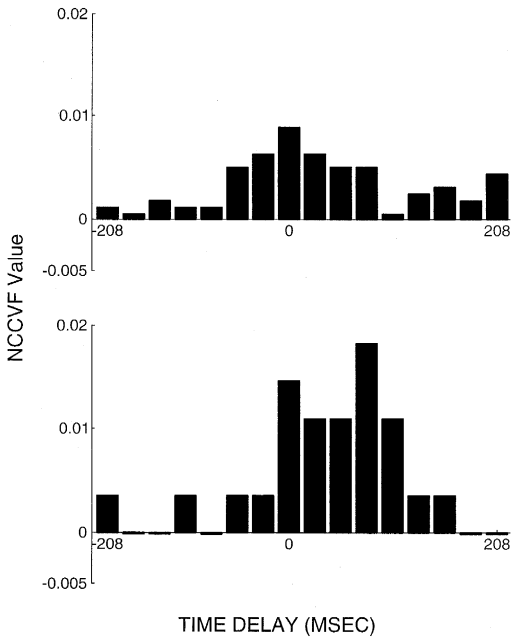

Fig. 2. Two normalized cross-covariance functions (NCCVF) from the pairs no. 311 vs. no. 364 (upper) and no. 434 vs. no. 1311 (lower). The spike train no. 311 has 50 spikes; no. 364, 195; and no. 434 has 10 spikes; no. 1311, 29. The NCCVF was defined by Shao and Chen (1987). Negative values indicate negative cross-correlation. The binwidth is 26 ms. The total recording time is 85 s.

2. Methods

2.1. Transformation of the paired spike train data

For a pair of spike trains (A and B), if every spike recorded from neuron A tends to be followed with certain relatively constant time by a change in discharge rate or by a systematic discharge pattern in spike train B, we call it a time correlation or time-locked correlation between train A and B. We assume that the discharge events of the two chains are jointly stationary Poisson processes (Cox and Lewis, 1966). We discretize the time axis as follows. We divide the total sample time from 0 to T into unit time intervals Δt. Δt is taken so small that the probability of more than one event occurring during an interval is zero. First we use the events in spike train A to trigger a fixed time window consisting of J time units. Index j could be from 0 to J or from $-J/2$ to $J/2$ as we usually do in constructing a cross-correlogram. Then we construct a $2 \times J$ table with cells ij $(i = 0,1;\ j = 1,2,\ldots,\ J)$. The time interval for each cell corresponds to the concept of bin width in cross-correlogram. Each cell $1j$ in the first row counts the number of spikes of train B falling in the time interval j. Each cell $0j$ in the second row counts the number of 0 events at the corresponding time interval j.

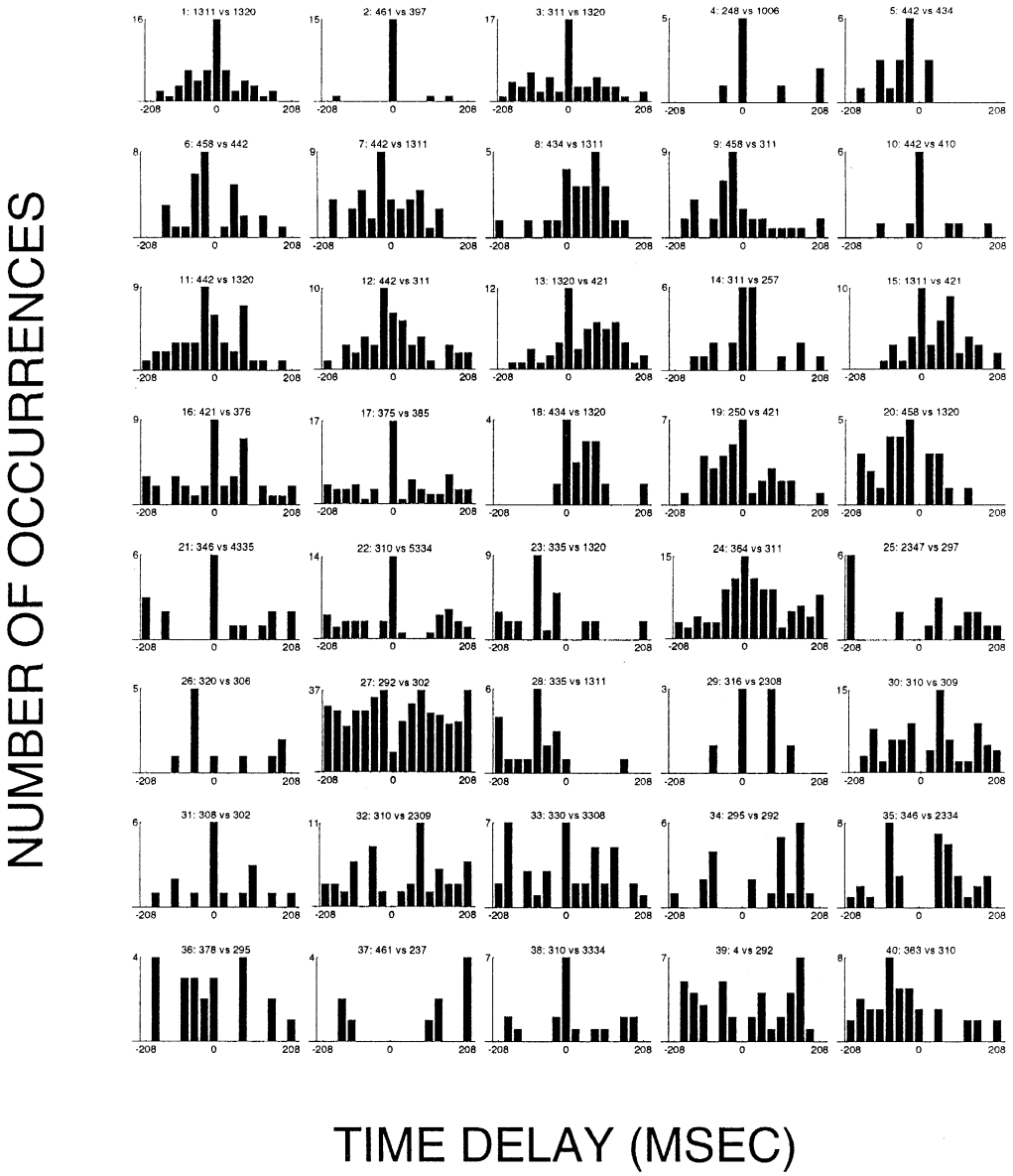

Fig. 3. 40 cross-correlograms selected by the generalized Fisher's or χ^2 test with the significance level $\alpha = 0.01$ from the data illustrated in Fig. 1. The cell pairs are indicated by the ID numbers which are given on the top of each panel. The bin width used for the calculation of cross-correlation was 16 data points (26 ms). The cross-correlograms have been sorted according to their correlation coefficient scores (CCS).

For a triggered window, either one event occurs in a particular time interval j or no event occurs in that interval, then there is either one count in row one or one count in row two of the column. Having gone through all trigger spikes (n) of train A, we get the number version of the cross-correlogram in the first row. Thus, row two is the

Table 1
A tabled cross-correlogram data from paired neurons no. 434 vs. no. 1311

	Time delay																
	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	Total
Events in B	1	0	0	1	0	1	1	4	3	3	5	3	1	1	0	0	24
No event	9	10	10	9	10	9	9	6	7	7	5	7	9	9	10	10	136
Total	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	160

complement of row one. Each column may be modeled as a binomial random variable $B(p,n)$. Where p is an unknown parameter – the probability of an event occurring in the particular interval and n is the column sum. It equals the total trigger number which is also the total number of spikes in spike train A used for constructing the table.

For example, spike train pair no. 434 vs. no. 1311 of Fig. 1 consists of 10 spikes in A and 29 in B. The total recording time was 85 s and we took $\Delta t = 26$ ms and $J = 16$. Then we constructed a 2×16 contingency table as Table 1 (see Section 4 for choices of bin width and J).

This kind of table illustrates the relation between two variables: the response variable – the occurrence of spikes in train B, and the explanatory variable – the peri-event time with respect to events in train A. It reflects the time correlation between the events of the two trains, or in other words, the correlation problem of two point processes becomes the correlation between the two variables. This concept can be extended to a train of point stimuli (random or regular) and a recorded neural spike train. Then the correlation of two spike trains becomes a correlation of one spike train and the stimuli; the table corresponds to a peri-stimulus time histogram (PSTH) instead of cross-correlogram. The statistical test for independence and measurement of association are based on this table. In practice, neurophysiological experiments very often generate tables with many empty cells in the first rows (see Table 1), which are referred to as sparse tables. The traditional chi-square test for the $2 \times J$ table would not be appropriate since the chi-square distribution is an asymptotic distribution. It is inaccurate when the contingency table is sparse. Therefore, we suggest a generalized Fisher's exact test.

2.2. Exact p value for testing the cross-correlation with generalized Fisher's method

A $2 \times J$ table like Table 2 can be used for exact analysis for the two variables X and t. In our cross-correlation analysis context, X represents the occurrence of spikes in train B and t represents peri-event time with respect to events in train A as mentioned above. Although t is time and is ordered, we treat it as unordered for now (see Section 4). Here the column sums are fixed and equal to each other, $n_1 = \ldots = n_J = n$. Then the data in each column can be considered as a realization of a binomial random variable, and the joint probability of the entire table is a product of J binomials

$$\Pr\left[Y_{ij} = y_{ij}; j = 1,....,J; i = 0,1 \mid y_{1j} + y_{0j} = n \right]$$

$$= \prod_{j=1}^{J} \binom{n}{y_{1j}} P_{1j}^{y_{1j}} \left(1 - P_{1j}\right)^{n - y_{1j}} \quad (1)$$

where p_{1j} are the conditional probabilities that a spike of train B occurs in time interval j given that a spike of train A occurs at time 0. Now we wish to test whether there is any significant association between X and t. The null hypothesis is that they are independent of each other. Under the null hypothesis, we have the conditional probabilities $p_{11} = ,... = p_{1J} = p_1$. There is only one unknown parameter p_1. r_1 is the sufficient statistic. If we further condition on row sums r_1 and r_0 (thus we condition on both row and column sums), the conditional probability

$$\Pr\left[Y_{ij} = y_{ij}; j = 1, \ldots, J; i = 0,1 \mid y_{1j} + y_{0j} = n; \right.$$

$$\left. \sum_{j=1}^{J} y_{1j} = r_1 \right] = \frac{\prod_{j=1}^{J} \binom{n}{y_{1j}} P_1^{y_{ij}} \left(1 - P_1\right)^{n - y_{1j}}}{\binom{N}{r_1} P_1^{r1} \left(1 - P_1\right)^{N - r_1}}$$

$$= \frac{\prod_{j=1}^{J} \binom{n}{y_{1j}}}{\binom{N}{r_1}} \quad (2)$$

This simple expression represents the multiple hypergeometric distribution and it does not depend on unknown parameters (Agresti, 1990). Thus, it permits exact inference. For hypothesis testing, the exact p value for an observed table can be determined by the 'Minimum Likelihood Method' (Gibbons and Pratt, 1975). This approach orders all possible tables with the fixed row and column margins on a probability scale. Based on the idea that the smaller the probability of the table, the more the table deviates from the null hypothesis, the two-sided p value is defined as the total probability of those tables whose null conditional probabilities are less than or equal to the probability of the observed table.

We can assign a critical value $\alpha = 0.01$ or 0.05 before the test. Having the exact p value by the above computation, we reject the null hypothesis of independence, if $p < \alpha$, and conclude that the time correlation of the paired neurons is significant. Otherwise it is not significant.

If the sample size of the spike train record is large (a lot of spikes either from high firing rate of the neurons or from long recording time), Fisher's exact test would be

Table 2
A $2 \times J$ contingency table

	$t = 1$	...	$t = J$	Total
$X = 1$	y_{11}	...	y_{1J}	r_1
$X = 0$	y_{01}	...	y_{0J}	r_0
Total	n_1	...	n_J	N

very time consuming. However the table is no longer sparse. The ordinary χ^2 test can be used:

$$\chi_2 = \sum_{ij} \frac{\left(y_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{3}$$

where E_{ij} is the expected count for cell ij calculated from the table margins. The statistic χ^2 tends to the chi-squared distribution.

2.3. Algorithms

Performing the exact analysis for a $2XJ$ table like Table 2 is very time consuming. Great effort has been made to find faster algorithms (Verbeek and Kroonenberg, 1985). Using the network algorithm (Mehta and Patel, 1980, 1983), a statistical package StatXact is commercially available for doing these exact tests (CYTEL Software Corp., Cambridge, MA). We may input our cross-correlogram data as Table 1 to StatXact and we obtain exact p values for independence with the command FI/EX. Sometimes, we get a message 'problem, too large' if the table size or the numbers of the fixed margins are too big. Then the program provides Monte Carlo methods to deal with this problem (Lynch et al., 1991). StatXact also has an option to carry out an ordinary χ^2 test.

Since the number of columns is generally very large, e.g. 16, 32, or 64, the computation time is a prominent problem even on a workstation. We have developed a better algorithm based on the fact that our tables always have equal column sums. The improved network algorithm that is special for the exact test of unordered $2 \times J$ tables with equal column sums will be published elsewhere. Tested with our cross-correlogram data, the program based on the improved algorithm is found to be much faster than StatXact, and reasonably fast for those data which elicit the message 'problem too large' in StatXact.

2.4. The measurement for time correlation

After determining that the association between the response and explanatory variables is significant and thus a correlation between the paired spike trains exists, we want to know the strength of this correlation. We need to have a single number as a summary description of cross-correlation displayed by the $2 \times J$ table or the cross-correlogram.

A measure of association between nominal variables, which is an analogue to the R-square (the coefficient of determination) for continuous variables, was suggested by Light and Margolin (1971). It is a measure of the proportion of total variation in the response variable attributed to the explanatory variable. It is the ratio of SSB (sum of squares between groups) and SST (total sum of squares). For our tabled paired spike train data, the total variation for the binary response is

$$\mathrm{SST} = \frac{N}{2} - \frac{r_1^2 + \left(N - r_1\right)^2}{2N} \tag{4}$$

Taking the variation of each column as variation within groups, the sum of squares within groups is

$$\mathrm{SSW} = \frac{N}{2} - \frac{1}{2} \sum_j \frac{1}{n_j} \sum_i y_{ij}^2 . \tag{5}$$

In our context, $n_1 = \ldots = n_j = n$ and since SSB = SST − SSW, the ratio SSB/SST, after some cancellation, is

$$R^2 = \frac{\frac{1}{n} \sum_{j=1}^{J} \left(y_{1j}^2 + y_{0j}^2\right) - \frac{r_1^2 + r_0^2}{N}}{N - \frac{r_1^2 + r_0^2}{N}} . \tag{6}$$

Practically, we tend to use its positive square root, $r = \sqrt{R^2}$, for which we give the name cross-correlation coefficient. It takes values between 0 and 1: $r = 0$ indicates independence between t and X; $r = 1$ if the knowledge of t completely determines X. Actually, a positive or negative sign can be attached to r if an excitatory or inhibitory relation is shown in the cross-correlogram.

There are many different kinds of measures of association for unordered $\mathrm{I} \times J$ tables. They are different in value and in interpretation. Eq. (6) is the most interpretable one. Under multinomial sampling, it is equivalent to a measure based on the proportional prediction suggested by Goodman and Kruskal (1954). With cross-correlogram analysis, $I = 2$ and Eq. (6) reduces to Φ^2 which is another measure of association based on the χ^2 statistic

$$\Phi_2 = \frac{\chi^2}{N},$$

where χ^2 is defined by Eq. (3).

Since R^2 is the proportional reduction in variance, it allows comparison across experiments. So does the cross-correlation coefficient r as the ordinary correlation coefficient for continuous variables.

3. Application and simulation

The algorithms have been implemented on a Silicon Graphics Indigo workstation (UNIX Irix 5.2) using F77 and MATHEMATICA (Wolfram Research, Champaign, IL; version 2.2). The spike train data used in this paper was from optical recordings from the abdominal ganglion of *Aplysia*, where activity of over 100 neurons can be monitored simultaneously (Zecevic et al., 1989; Wu et al., 1994). The experiments and optical recordings have been described (Tsau et al., 1994). In this paper, the cross-correlation data we tested was fast cross-correlation (Tsau et al., 1994).

For the data shown in Fig. 2, we calculated the exact p values which are 0.00977 and 0.0005, respectively, for the

neuron pairs no. 434 vs. no. 1311 (lower panel) and pair no. 311 vs. no. 364 (upper panel) by the generalized Fisher's or χ^2 test. The cross-correlation coefficients r are 0.4316 and 0.2235, respectively. The exact p values show that the correlations for both pairs are significant, but the strength of correlation of pair no. 434 vs. no. 1311 is larger than that of the other pair. This result is surprising, since, by just looking at the cross-correlograms (cross-correlogram nos. 8 and 24 in Fig. 3), it would appear the strength of correlation of pair no. 311 vs. no. 364 is larger because the correlation peak is higher. While the normalized cross-covariance function (NCCVF) figures imply that of pair no. 434 vs. no. 1311 is larger (Fig. 2, lower panel). From the point of view of probability, the NCCVF figures gave us the correct impression because the NCCVF is normalized by the trigger number and the mean discharge rate (Shao and Chen, 1987). Here, our r values are consistent with the NCCVF figures.

We applied the generalized Fisher's test to a set of data from 149 cells which form 11 026 cross-correlation pairs (Fig. 1). For those records with a large number of spikes, we used the χ^2 test (Eq. (3)) instead. The criterion for using the χ^2 test was set as the row sum of the first row $r_1 >= 50$. The p value of the χ^2 test with $J-1$ degree of freedom is calculated using MATHEMATICA. It took about 20 h to perform all 11 026 tests on our machine. Fig. 3 shows the 40 cross-correlation histograms which are significant and selected by the test (at the level of $\alpha = 0.01$). They show cross-correlation peaks or troughs. The ratio 40 out of 11 026 pairs is about 0.4% which means that only a very small percentage of the total cross-correlation pairs are significantly associated even though most cells are driven by the stimulus, suggesting that they might be associated in some loose way rather than the tight synaptic connections.

To distinguish the possible co-activation by stimulus on both neurons of the pairs, we applied Fisher's test to the 'shift predictor' (Perkel et al., 1967) of the 40 selected pairs shown in Fig. 3. The results show that shifting one trial of one of the paired spike trains reduced the cross-correlation tremendously. Most pairs become insignificant except nos. 3, 11, 12 and 15 of Fig. 3 while their peaks are lower and the bases of the peaks become relatively wide. These results suggest that the correlation detected from most pairs in our data set is due to effective connection rather than the direct effect of the stimuli. The proposed methods are not applicable to the difference cross-correlation histogram (Epping and Eggermont, 1987) resulting from subtraction of the shift predictor from the 'raw' correlogram because this subtraction may result in negative numbers in the first row of the $2 \times J$ tables. There is no definition for negative number in both Fisher's exact test and the R^2 value. Further development for these procedures is needed.

A spike train simulator developed by Aertsen and Gerstein (1985) can generate paired spike trains with a specified correlation (or 'Synaptic') strength β. We used such a simulator (the software was kindly provided by Dr. Gerstein) with a strength (probability) 0.15 to generate a long data set (1000 spikes in train A as the trigger train to construct the cross-correlogram and the 2×32 table. Fig. 4B). We see a major peak from the correlogram. The p value is 0 (it is very very small; the computer takes it as 0) and the r value is 0.1515. Then we broke down the pair of trains into 20 sections with each section having 50 spikes for its trigger train. Fisher's test shows that 18 out of the 20 are significant at the level of 0.05. Fig. 4A is one of the 18, the p value of it is 0.002345 and the r value is 0.195.

To reveal further the detectability of Fisher's test and the properties of the measure for cross-correlation with

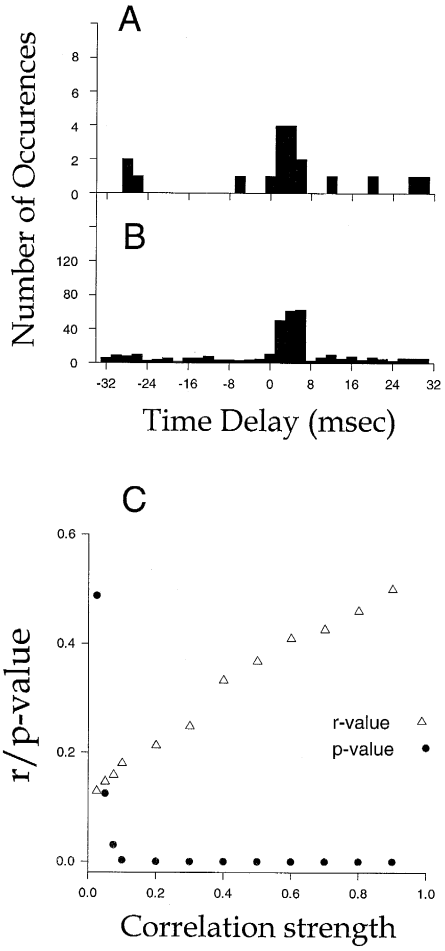

Fig. 4. Performance of the suggested methods with randomly generated spike trains by simulation. (A) Cross-correlogram of a short section with 50 spikes in spike train A as trigger ($n = 50$) from a long sample (shown in B). The exact p value calculated using Fisher's procedure is 0.002345 and the r value, 0.195. (B) Sample data set with 1000 spikes in the trigger train A which was generated with the simulator described by Aertsen and Gerstein (1985). The synaptic strength β was set to 0.15 and the constant delay, 2 mS; the maximum of the uniformly distributed random interval was set to 6 ms. The exact p value is 0 and the r value is 0.1515. (C) The p and r values vs. different levels of β in a series of simulations. $n = 60$. The firing rates for both of the paired trains are 4 ± 1 s^{-1} and bin width is 2 ms for A–C.

small sample size, we generated short spike trains with β set at different levels. Fig. 4C shows that with the trigger trains A of 60 spikes, the exact p value starts falling below 0.05 level when $\beta = 0.075$. It drops close to 0 very quickly as β increases. It implies that the test can reliably detect pretty small excitatory synaptic connection with a small sample. Fig. 4C also shows that the r value is a monotonic increasing function of the specified correlation strength and it is a good measure because it looks virtually linear. As illustrated by Aertsen and Gerstein (1985), inhibitory connection is much harder to detect. An example from our simulation shows that with β as large as -0.8 and spike number of 1000 in the trigger train A, a significant cross-correlation with $p = 0.0278$ can be detected.

To evaluate the false positive error rate, we generated 40 independent ($\beta = 0$) pairs of spike trains with the simulator. Their p values were calculated. None of them is significant at the level of 0.05.

For comparison purposes, another selection method of the largest correlation coefficient score (CCS) was applied (Tsau et al., 1994). However, it must be used cautiously because when cross-correlograms have very few occurrences within the time window and one or two spike trains in a pair have only a few spikes, the CCS can be very large which is apparently not a valid measure. Therefore, those cross-correlograms with few occurrences have to be deleted before selection. Fig. 5 shows 39 cross-correlograms with the largest scores when the criterion for deleting pairs is set to 7 occurrences, where 18 out of 39 are identical to

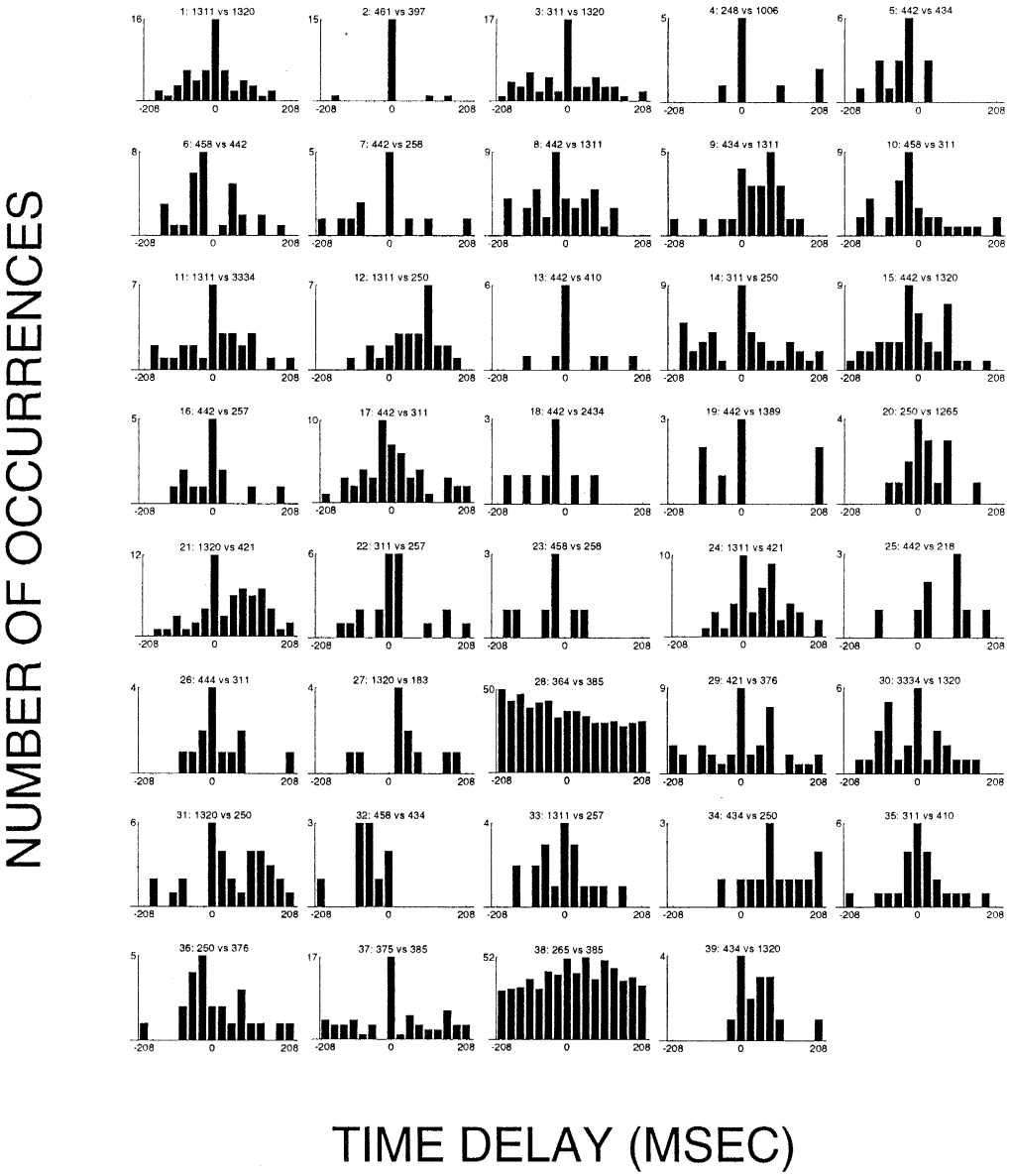

Fig. 5. 39 cross-correlograms selected by the top CCS among the cross-correlograms with more than 7 occurrences within the time window, as used by Tsau et al. (1994). The data is illustrated in Fig. 1. The cell pairs are indicated by the ID numbers which are given on the top of each panel. The bin width is 26 ms. The cross-correlograms have been sorted according to their CCS.

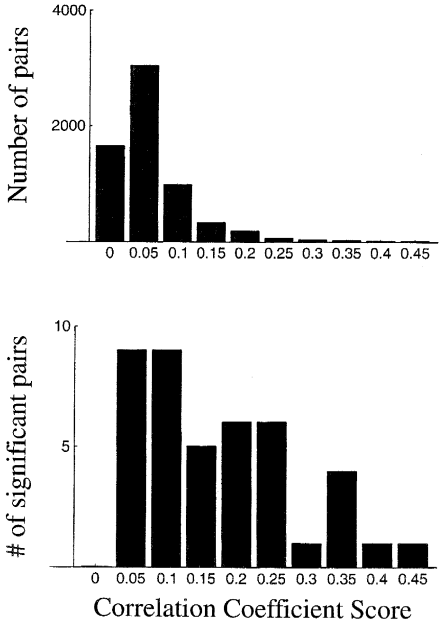

Fig. 6. Distribution of the number of cross-correlograms over CCS. The scale for each score is the lower limit (e.g., 0 represents 0–0.0499). Upper: distribution of total 11 026 pairs. The number for 0.45–0.499 is 11, and for 0.4–0.4499 it is 16, which indicates that the number is much larger than that of significant pairs (lower panel). Lower: distribution of 40 pairs selected by the generalized Fisher's or χ^2 test.

those in Fig. 3 selected by Fisher's or χ^2 test. 15 out of 39 pairs with the largest scores are identical to those in Fig. 3 when the criterion is set to 4 occurrences (cross-correlograms not shown), suggesting that deleting pairs with low occurrences helps selecting significant pairs by CCS. In fact, 6000 out of 11 026 pairs have CCS below 0.2 (Fig. 6, upper panel), and they are usually not selected by the largest score method. However, 13 out of 40 significant pairs selected by Fisher's test are below 0.2 but not less than 0.05 (Fig. 6, lower panel), showing that the pairs significantly correlated could have relative small correlation coefficient scores. On the other hand, there are only 2 significant pairs (Fig. 6 lower panel) among 27 pairs with the highest scores (upper panel), suggesting that a large score is not determinative for statistical significance.

4. Discussion

We have described the cross-correlation analysis methods for paired neural spike trains obtained from hundreds of units simultaneously monitored. By transferring the spike train data into a $2 \times J$ table, a statistical test based on the multiple hypergeometric model has been suggested. Also, we proposed a cross-correlation coefficient based on the R-square analogue to be a measure for the strength of association between the paired neural spike trains. These methods are conceptually novel, since both significance and the strength of cross-correlation are evaluated from the entire cross-correlogram instead of from the confidence limit of each bin (Abeles, 1982) or from its peaks or troughs (Graham and Duffin, 1981). The multiple comparison problem (see below) no longer exists in our test based on the single p value of the entire table. The test procedures before were based on estimations of some unknown parameters of the variable (e.g. the mean bin count and its variance). The test proposed here does not include unknown parameters, thus, the probability calculated using Eq. (2) is exact. From this we can obtain the exact p value of the table. The suggested methods are especially useful when the sample size is small and when we deal with sparse tables. The tests can be used as a fast screening method to pick up the related pairs. When the related pairs of neurons have been picked, further studies like cross-correlograms, NCCVF or joint PSTH (Gerstein and Perkel, 1969, 1972) can then be used to analyze the direction or the temporal properties of the correlation. The detected time correlation between paired neurons does not necessarily indicate direct synaptic connection between the two neurons. Explanation of the detected time correlation should be made cautiously (Moore et al., 1970).

Although the proposed methods allow cross-correlation analysis to be carried out with small sample size, we face the common problems in statistical inference: (1) for the test, if $p \leq \alpha$, we can conclude that there is cross-correlation between the paired spike trains at the significance level of α. While if $p > \alpha$, we should not simply conclude that there is no cross-correlation. $p > \alpha$ would happen either when there is really no cross-correlation or just because the sample size is too small. A larger sample size always has greater power to detect a correlation; (2) for the estimation of the strength of correlation, the same calculation can be applied to both small and large samples. If the recorded pair of neurons are correlated more or less like the simulation model proposed by Aertsen and Gerstein (1985), a larger sample tends to smooth the cross-correlogram. On the other hand, the cross-correlogram from smaller sample would be noisier (compare Fig. 4A with B). This 'noisy effect' would be added to the estimation of r resulting in overestimation of the strength of cross-correlation, so that although r as a measure can be used to compare the correlations across different experiments, their sample sizes should not be too different. Finally, acquiring as large a sample as possible during the experiment is always suggested.

The basic problems of multiple comparison were originated from general methods of analysis of variance (ANOVA) in statistics. Here, for the cross-correlogram analysis with the bin by bin based methods (Abeles, 1982; Graham and Duffin, 1981) the problem is: for a J bin cross-correlogram we are doing J tests simultaneously. The significance level of the entire cross-correlogram is not the same level as we assign for the individual test anymore. For instance, in the case of $J = 16$, if we take $\alpha = 0.01$, the null hypothesis is that there is no correlation between the two spike trains. The correlogram is consid-

ered significant if any one or more bin counts are higher than the confidence limit line; and it is a type I error if the null hypothesis is true. Thus, the type I error rate would be 1 minus the probability that all 16 bins are within the confidence limit. That is

$$\alpha' = 1 - (1 - 0.01)^{16} = 0.149.$$

Although by Tukey's procedure (Miller, 1981), the family type I error rate of 0.01 can be kept to increase the protection, the significance level for each bin (i.e., the horizontal line for every bin of the cross-correlogram) will be at the level of

$$\alpha = 1 - (1 - 0.01)^{1/16} = 0.000628.$$

This is an even wider confidence limit than the one that is already too wide. Much power for detecting the possible related pairs will be lost.

A basic assumption of the test is that the paired spike trains are jointly stationary Poisson processes. This may not be true in different practical situations. Thus, we should bear in mind that if the autocorrelation is strong in either train, error in the statistical test may arise and the error is usually a false positive. Inspecting the raster of spike trains is an important step before doing statistical tests. Roughly speaking, if the spike trains show periodic bursts, or if they are from pacemakers, the test may be invalid.

For the kind of $2 \times J$ tables in this paper, we take the peri-event time as an unordered variable although discretized time is ordered. For our purpose of statistical testing and measuring the strength of correlation, we do not consider any linear trend of discharge over peri-event times and obviously, the correlation of spike trains is not a certain simple monotonic function of time. Taking the time as ordered variable is necessary only when we want to study the correlative discharge as a function of time by some kind of regression procedure.

For constructing the cross-correlogram, we usually take one of the paired neural spike trains, say train A, as the trigger. If we reverse the two trains, i.e. take train B as the trigger, we will get a cross-correlogram which is symmetrical to the original one. If the number of spikes in train B is very different from that of train A, we will have different trigger numbers for the two symmetric cross-correlograms. Then we will have two $2 \times J$ tables with the same counts in the first row cells but different column sums. When we conduct the statistical tests for independence, we will get slightly different p values. We may face a dilemma, although rarely, that for a paired spike trains A and B, the association is significant for one side, e.g., A to B, but not significant for B to A. For consistency, we suggest always using the spike train which has the smaller spike number of the two as the trigger. The reason is, from the statistical point of view, the trigger number takes the role of sample size. We are using n (trigger number) samples to observe the relationship between the two variables. If we use the larger number as trigger, what we are doing is just increasing the sample size by repeating some data in the same data set.

By definition, when we discretize the time axis, we take Δt so small that more than one event occurring in Δt is impossible. Ideally, we should take Δt as the bin width to construct the $2 \times J$ table. Then J would be a very big number if we want to look at a wide time window. It would take a long time to compute such a large $2 \times J$ table. In many practical situations, the neuronal discharge rate is low. Thus, the probability of 2 or more spikes falling in one bin in a single trial is very small. Then, taking a bigger bin width that includes more than one Δt would be a good idea. The choices of bin width and the number of columns J should be compromised with several factors such as how long the fast interaction of the neurons is and how wide a time window the experimenter wants to observe based on his data, the firing rate of the neurons and the computational time.

As mentioned above, if the tests suggested in this paper are going to be performed as a screening method to find correlated pairs of neurons, again the multiple comparison issue should be stressed. For practical screening for k simultaneously recorded spike trains we are trying to conduct $h = k!/2!(k-2)!$ tests. If we take all tests as a family, under the null hypothesis of no time correlation between those neurons, we can, if we wish, control the type I error rate at a fixed level α' for all tests jointly. In other words, we want the family confidence coefficient of $1 - \alpha'$ for the h tests. By assuming that all tests for the h cross-correlograms are independent of each other, the significance level for every individual test should be

$$\alpha = 1 - (1 - \alpha')^{1/h}.$$

Then those tests with the exact p value $< \alpha$ are considered significant (for instance, take $\alpha' = 0.05$, and $h = 20$, then $\alpha = 1 - (0.95)^{1/20} = 0.00256$). This is the idea from Tukey (Miller, 1981). An alternative called Bonferroni's procedure is even simpler (Woolson, 1987). The procedure requires that each test be made at the $1 - \alpha'/h$ confidence level. According to the Bonferroni's inequality, this procedure actually gives a lower bound on the true coefficient. Since the test results are not usually independent, the two procedures are too conservative (the results of these two procedures are very close. Bonferroni's is a little bit more conservative). Because of this conservative nature, the family confidence coefficient should be specified at a lower level (e.g., 0.9 or 0.8). These procedures just provide increased protection to the null hypothesis while the power of the tests is decreased. This power problem becomes extremely serious when the family is big. Therefore, the procedures should be used cautiously: (1) the number of tests h should not be too large; (2) the investigator really wants to control the familywise error rate rather than per-test error rate (Woolson, 1987). If the

neurophysiologist insists on considering each test separately and controlling the per-test error rate, the multiple comparison can be ignored and α is the significant level for each test (O'Brien, 1983). Since we are performing more tests, a higher error rate is acceptable for the family. In other words, giving up the fixed low family error rate (0.05 or 0.01) is sensible. The key point is the null hypothesis we are testing. The issue here is different from the problem for each cross-correlogram we mentioned in the beginning of this section. In the latter situation, the null hypothesis is that there is no correlation between the two spike trains, it is definitely a hypothesis for all bins in the correlogram as a family. The multiple comparison problem is unavoidable there.

If the $2 \times J$ table size and the values of the margins are large, the computing time would be a serious problem (Section 2). However, if this is due to the large numbers in the cells of the table, the table is no longer sparse. Then the ordinary χ^2 test would work well. For the sample size of this approximate test, as Cochran (1954) suggested, a minimum expected value of 1 is permissible as long as no more than about 20% of the cells have expected values below 5. For our 2×16 table constructed by paired spike train data, the expected values of the second row cells are almost always larger than 5. If the values of 10 cells in the first row are 5, r_1 would be larger than 50. So that we take $r_1 >= 50$ as our criterion for using χ^2 tests.

Acknowledgements

We thank Dr. Lawrence B. Cohen for valuable comments and suggestions. We thank Dr. George L. Gerstein for the spike train simulator software. X.M. Shao was supported by grant no. HL40959. Y. Tsau was supported by grant no. NS08437.

References

Abeles, M. (1982) Quantification, smoothing, and confidence limits for single-units' histograms, J. Neurosci. Methods, 5: 317–325.

Aertsen, A.M.H.J. and Gerstein, G.L. (1985) Evaluation of neuronal connectivity: sensitivity of cross-correlation, Brain Res., 340: 341–354.

Agresti, A. (1990) Categorical Data Analysis, Wiley, New York, pp. 59–66.

Cochran, W.G. (1954) Some methods of strengthening the common χ^2 tests, Biometrics, 10: 417–451.

Cox, D.R. and Lewis, P.A.W. (1966) The Statistical Analysis of Series of Events, Chapman and Hall, London, pp. 17–36.

Eggermont, J.J. (1992) Neural interaction in cat primary auditory cortex. dependence on recording depth, electrode separation, and age, J. Neurophysiol., 68: 1216–1228.

Eggermont, J.J. (1994) Neural interaction in cat primary auditory cortex. II. Effects of sound stimulation, J. Neurophysiol., 71: 246–270.

Epping, W.J.M. and Eggermont, J.J. (1987) Coherent neural activity in the auditory midbrain of the grassfrog, J. Neurophysiol., 57: 1464–1483.

Gerstein, D.L. and Perkel, D.H. (1969) Simultaneously recorded trains of action potentials: analysis and functional interpretation, Science, 164: 828–830.

Gerstein, D.L. and Perkel, D.H. (1972) Mutual temporal relationships among neuronal spike trains, Biophys. J., 12: 453–473.

Gibbons, J.D. and Pratt, J.W. (1975) P-values: interpretation and methodology, Am. Stat., 29: 20–25.

Gochin, P.M., Miller, E.K., Gross, C.G. and Gerstein, G.L. (1991) Functional interactions among neurons in inferior temporal cortex of the awake macaque, Exp. Brain Res., 84: 505–516.

Goodman, L.A. and Kruskal, W.H. (1954) Measures of association for cross classification, J. Am. Stat. Assoc., 49: 732–764.

Graham, K. and Duffin, J. (1981) Croos correlation of medullary expiratory neurons in the cat, Exp. Neurol., 73: 451–464.

Levick, W.R., Cleland, B.G. and Dubin, M.W. (1972) Lateral geniculate neurons of cat: retinal inputs and physiology, Invest. Ophthalmol., 11: 302–311.

Light, R.J. and Margolin, B.J. (1971) An analysis of variance for categorical data, J. Am. Stat. Assoc., 66: 534–544.

Lindsey, B.G., Hernandez, Y.M., Morris, K.F. and Shannon, R. (1992) Functional connectivity between brain stem midline neurons with respiratory-modulated firing rates, J. Neurophysiol., 67: 890–904.

London, J.A., Zecevic D. and Cohen L.B. (1987) Simultaneous optical recording of activity from many neurons during feeding in *Navanax*, J. Neurosci., 7: 649–661.

Lynch, J.C., Landis, J.R. and Localio, A.R. (1991) StatXact review, Am. Stat., 45: 151–154.

Mehta, C.R. and Patel, N.R. (1980) A network algorithm for the exact treatment of the 2Xk contingency table, Commun. Stat.-Simul. Comput., B9: 649–664.

Mehta, C.R. and Patel, N.R. (1983) A network algorithm for performing Fisher's exact test in rXc contingency tables, J. Am. Stat. Assoc., 78: 427–434.

Melssen, W.J. and Epping, W.J.M. (1987) Detection and estimation of neural connectivity based on crosscorrelation analysis, Biol. Cybern., 57: 403–414.

Miller, R.G. (1981) Simultaneous Statistical Inference, 2nd edn., Springer-Verlag, New York, pp. 12–25.

Moore, G.P., Segundo, J.P. and Perkel, D.H. (1970) Statistical signs of synaptic interaction in neurons, Biophys. J., 10: 876–900.

Nelken, I., Prut, Y., Vaadia, E. and Abeles, M. (1994) Population responses to multifrequency sounds in the cat auditory cortex: one- and two-parameter families of sounds, Hearing Res., 72: 206–222.

O'Brien, P.C. (1983) The appropriateness of analysis of variance and multiple comparison procedures, Biometrics, 39: 787–788.

Palm, G., Aertsen, A.M.H.J. and Gerstein, G.L. (1988) On the significance of correlations among neuronal spike trains, Biol. Cybern., 69: 1–11

Parson, T.D., Salzberg, B.M., Obaid, A.L., Raccuia-Behling, F. and Kleinfeld, D. (1991) Long-term optical recording of patterns of electrical activity in ensembles of cultured *Aplysia* neurons, J. Neurophysiol., 66: 316–333.

Perkel, D.H., Gerstein, G.L. and Moore, G.P. (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains, Biophys. J., 7: 419–440.

Sears, T.A. and Stagg, D. (1976) Short-term synchronization of intercostal motoneurone activity, J. Physiol., 263: 357–381.

Shao, X. and Chen, P. (1987) Normalized auto- and cross-covariance functions for neuronal spike train analysis, Intern. J. Neurosci., 34: 85–95.

Tsau, Y., Wu, J.-Y., Hopp, H.P., Cohen, L.B., Schiminovich, D. and Falk, C.X. (1994) Distributed aspects of the response to siphon touch in *Aplysia*: spread of stimulus information and cross-correlation analysis, J. Neurosci., 14: 4167–4184.

Vaadia, E., Haalmen, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. and Aertsen, A. (1995) Dynamics of neuronal interactions in monkey cortex in relation to behavioral events, Nature, 373: 515–518.

Verbeek, A. and Kroonenberg, P.M. (1985) A survey of algorithms for exact distributions of test statistics in rXc contingency tables with fixed margins, Comput. Stat. Data Anal., 3: 159–185.

Villa, A.E.P. and Abeles, M. (1990) Evidence for spatiotemporal firing patterns within the auditory thalamus of the cat, Brain Res., 509: 325–327.

Wilson, M.A. and McNaughton, B.L. (1994) Reactivation of hippocampal ensemble memories during sleep, Science, 265: 676–679.

Woolson, R.F. (1987) Statistical Methods for the Analysis of Biomedical Data, Wiley, New York, pp. 327–337.

Wu, J.-Y., Tsau, Y., Hopp, H.P., Cohen, L.B., Tang, A.C. and Falk, C.X. (1994) Consistency in nervous systems: trial-to-trial and animal-to-animal variations in the response to repeated application of a sensory stimulus in *Aplysia*, J. Neurosci., 14: 1366–1384.

Ylinen, A., Bragin, A., Nadasdy, Z., Jando, G., Szabo, I., Sik, A. And Buzsaki, G. (1995) Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms, J. Neurosci., 15: 30–46.

Zecevic, D., Wu, J.-Y., Cohen, L.B., London, J.A., Hopp, H.P. and Falk, C.X. (1989) Hundreds of neurons in the *Aplysia* abdominal ganglion are active during the gill-withdrawal reflex, J. Neurosci., 9: 3681–3689.