Neuronal transients

KARL J. FRISTON*

The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, U.K.

SUMMARY

A recent letter to *Nature* (Vaadia *et al. Nature, Lond.* **373**, 515–518 (1995)) presented compelling results concerning neuronal interactions in monkey cortex. Vaadia *et al.* made two fundamental points: (i) it is possible that cortical function is mediated by dynamic modulation of coherent firing among neurons; and (ii) these time-dependent changes in correlations can emerge without modulation of firing rates. These observations have severe implications for models of neural coding and empirical approaches that are based on firing rates (e.g. neuroimaging). This communication presents a simpler explanation for the results presented in Vaadia *et al.*, by noting they are consistent with the correlated expression of stereotyped neuronal transients following (or preceding) a salient event. This re-formulation is important because: (i) correlations measured in terms of transients are not time-dependent, allowing prevailing models of neural coding to be 'reinstated'; and (ii) it suggests a powerful analysis based on singular value decomposition of firing rates.

1. INTRODUCTION

A fundamental phenomenon observed by Vaadia *et al.* (1995) is that, following behaviourally salient events, the degree of coherent firing between two neurons can change profoundly and systematically over the ensuing second or so. Furthermore, the mean firing rate (averaged over epochs) does not necessarily show any systematic change. One implication is that a 'better' metric of neuronal interactions could be framed in terms of dynamic changes in correlations, modulated on timescales of 100–1000 ms. This possibility touches on the distinction between temporal coding and rate coding as substrates of a putative neural code. This distinction, and the related debate (see, for example, Shadlen & Newsome 1995) centres on whether the precise timing of individual spikes can represent sufficient information to facilitate information transfer in the brain. The position adopted by Vaadia *et al.* adds an extra dimension to this debate: while accepting that spike trains can be considered as stochastic processes (i.e. the exact time of spiking is not vital), they suggest that temporal coding may be important in terms of dynamic time-dependent and behaviourally specific changes in the probability that two or more neurons will fire together. This dynamic modulation of coherent firing does not necessarily involve sustained changes in mean firing rate. In short they are proposing, very sensibly, a (second order) temporal coding model that is consistent with the stochastic behaviour of spike trains.

Although the above perspective appeals to our appreciation of the brain as a complex, highly non-linear dynamical system, it poses a problem for many useful models of neuronal interactions, for example: associative plasticity, self-organizing maps, feature detection and information theoretic accounts of neuronal activity and connectivity (for examples, see Willshaw *et al.* 1979; Kohonen 1982; Linsker 1988; Doldiak 1990; Gally *et al.* 1990; Lopez *et al.* 1990; Rubner & Schulten 1990; Miller 1992), where these models are expressed in terms of firing rates. This is because the temporal coding, implicit in a dynamic modulation of coherence, de-emphasizes the role of mean firing rates in mediating neuronal interactions. One resolution of this problem is provided by a simple explanation of temporally modulated coherence, that is based on the notion of neuronal transients. The aim of this paper is to describe this explanation and an analysis of separable or multiunit spike trains that ensues.

* Present address: The Wellcome Department of Cognitive Neurology, MRC cyclotron Unit, Hammersmith Hospital, 150 DuCane Road, London W12 ONN, U.K.

2. AN ALTERNATIVE VIEW

Imagine that two neurons respond to an event with a similar transient (a short-lived stereotyped time-dependent change in the propensity to fire); although the normalization procedures adopted in Vaadia *et al.* (1995) remove correlations induced by the mean expression of this transient, they will not remove correlations because of the covariation about this mean. For example, if two neurons respond to an event with decreased firing for 400 ms, and this decrease was correlated over epochs, then positive correlations between the two firing rates would be seen for the first 400 ms of the epoch, and then fade away, therein emulating a dynamic modulation of coherence.

This phenomenon can be made clear using a simple model: let $x_i(t)$ represent firing in unit i, at time t, following the onset of an event. Let $t_i(t)$ be an event-specific transient where its expression is modulated by a_i.

$$x_i(t) = a_i . t_i(t) + e_i(t), \qquad (1)$$

Proc. R. Soc. Lond. B (1995) **261**, 401–405
Printed in Great Britain

a_i simply represents the degree to which the transient $t_i(t)$ is expressed and is therefore a characterization of the event-related, short-term changes in the pattern of firing. $e_i(t)$ is a term reflecting non-specific activity, that is uncorrelated with the expression of the transient i.e. $\mathrm{Cov}(a_i,\ e_i(t)) = 0$ over many events. The non-specific cross-covariance between neurons i and j is given by $\mathrm{Cov}(e_i(t),\ e_j(t+h))$. If these covariances are stationary, this reduces to the cross-covariance function $g_{ij}(h)$. By direct calculation:

$$\mathrm{Cov}(x_i(t),\ x_j(t+h)) = \mathrm{Cov}(a_i,\ a_j)\,.\,t_i(t)\,.\,t_j(t+h) + g_{ij}(h), \tag{2}$$

This equation states that the covariance pattern can be expressed as the sum of non-specific covariances $g_{ij}(h)$ and a term that can be factorized into the covariance in the expression of the transients $\mathrm{Cov}(a_i,\ a_j)$ and the transients themselves $t_i(t)\,.\,t_j(t+h)$. The key thing to note is that the stimulus-related interaction between the two units is expressed as the covariance $\mathrm{Cov}(a_i,\ a_j)$ that is not a function of time. In other words dynamic modulation of covariances can be equivalently formulated as fixed covariances of dynamic transients. This alternative perspective essentially replaces the spike (a depolarization transient) with a neuronal transient. In terms of neuronal transients there is no dynamic modulation of correlations, or implicit temporal coding.

The firing rates averaged over epochs are:

$$E(x_i(t)) = E(a_i)\,.\,t_i(t) + E(e_i(t)), \tag{3}$$

and can, if $E(a_i) \approx 0$, show no modulation of mean firing rates following a stimulus event. In summary a highly structured fine-scale temporal pattern of correlations (shaped by the form of the underlying transients) can be observed even without modulation of mean firing rates (averaged over epochs). Of course in reality there is likely to be both a dynamic modulation of mean firing rate and of correlations. The point being made by Vaadia *et al.* is that one does not necessarily imply the other; and in some cases dynamic modulation of correlations can be expressed in the absence of changes in mean firing rate. In terms of neuronal transients this is equivalent to saying that the expression of transients is highly correlated in two neurons but the transients can be expressed as both increases and decreases in firing, such that their effect on mean firing rate cancels out (over a sufficient number of epochs).

The importance, of this perspective on dynamic correlations, is that conceptual and mathematical models that have proved themselves in application to firing rates, or the expression of depolarization events (i.e. x_i above) can be applied to the expression of higher-order transients (i.e. a_i above). In other words the rules that apply to x_i may also be applicable to a_i where x_i can be thought of as the degree to which a depolarization transient is expressed and, on a longer timescale, a_i reflects the degree to which a neuronal transient is expressed (the neuronal transient being composed of many depolarization transients). For example, associative plasticity (an increase in synaptic efficacy given the conjoint occurrence of pre and post-synaptic transients) would be manifest as the increased probability of some neuronal transient in V5, in response to a sensory transient in V2, only if the conjoint expression of both occurred more often than chance would predict. These sorts of hypotheses are amenable to empirical analysis using the techniques described in the next section. The forgoing may represent a 'revision of prevailing models' called for in Vaadia *et al.*

3. A MATHEMATICAL PERSPECTIVE

This section generalizes the above analysis and introduces singular value decomposition (SVD) as a technique that can characterize dynamics in terms of transients: in short it is shown that any (event-referenced) cross-covariance matrix $\boldsymbol{x}_i^T\,\boldsymbol{x}_j$, embodying dynamic modulations of coherent firing, can be expressed in terms of correlated transients. The generality of the arguments in the previous section can be established using SVD. Consider the matrix equivalent of equation 2.

$$\boldsymbol{x}_i^T\,\boldsymbol{x}_j = \boldsymbol{a}_i^T\,\boldsymbol{a}_j\,.\,\boldsymbol{t}_i^T\,\boldsymbol{t}_j + \boldsymbol{e}_i^T\,\boldsymbol{e}_j, \tag{4}$$

$x_i$represents a (mean corrected) matrix of firing rate data from neuron i, with one row for each epoch, and one column for each time bin. $\boldsymbol{a}_i$ is a column vector of coefficients representing the relative expressions of $\boldsymbol{t}_i$ in each epoch. $\boldsymbol{t}_i$ is a row vector describing the transient and $\boldsymbol{e}_i^T\,\boldsymbol{e}_j$ is the cross-covariance matrix observed in the absence of any stimuli. By noting the existence of the singular value decomposition of $\boldsymbol{x}_i^T\,\boldsymbol{x}_j - \boldsymbol{e}_i^T\,\boldsymbol{e}_j$,

$$\boldsymbol{x}_i^T\,\boldsymbol{x}_j - \boldsymbol{e}_i^T\,\boldsymbol{e}_j = l^1\boldsymbol{t}_i^{1T}\,\boldsymbol{t}^1{}_j + l^2\boldsymbol{t}_i^{2T}\,\boldsymbol{t}_j^2 + l^3\boldsymbol{t}_i^{3T}\,\boldsymbol{t}_j^3 + \ldots, \tag{5}$$

one observes that any cross-covariance structure $\boldsymbol{x}_i^T\,\boldsymbol{x}_j$, corrected for non-specific components $\boldsymbol{e}_i^T\,\boldsymbol{e}_j$, can be expressed as the sum of covariances resulting from the expression of paired transients ($\boldsymbol{t}_i^k$ and $\boldsymbol{t}_j^k$). The expression of these transient ($\boldsymbol{a}_i^k$ and $\boldsymbol{a}_j^k$) covaries according to the singular values l^k where $\boldsymbol{a}_i^{kT}\,\boldsymbol{a}_j^k = l^k$ and $\boldsymbol{a}_i^k = (\boldsymbol{x}_i - \boldsymbol{e}_i)\,.\,\boldsymbol{t}_i^k$. In this more general model any observed neural transient is described by a linear combination of the $\boldsymbol{t}_i^k$ (or $\boldsymbol{t}_j^k$). This can be regarded as a multivariate extension of the single transient model in the previous section. SVD is similar to principal components analysis (PCA) in the sense that it finds linear combinations of the components (time bins) of the observation (i.e. singular vectors) that have the greatest covariance. PCA finds the linear combination (i.e. eigenvectors or principal components) that has the greatest covariance with its self (i.e. variance). Indeed if one applied SVD to data matrices from the same neuron this would be identical to a PCA, however in this application SVD is being used to find the linear combination of time bins from one neuron that has the greatest covariance with a second linear combination from the other neuron. SVD is a ubiquitous mathematical device that will found in most high level software packages. Some simulation results are used below to suggest that SVD could be used to characterize the form and expression of transients embedded in empirical data.

4. SIMULATIONS

In this section an analysis of simulated spike trains is presented to show that the transient model is a sufficient explanation for the results described in Vaadia *et al.* (1995) and secondly to illustrate the potential power of SVD in characterizing these transients.

(*a*) *Cross correlations produced by neural transients*

Two processes $x_i(t)$ and $x_j(t)$ were simulated using equation (1) and an assumed form for the transients. $x_i(t)$ and $x_j(t)$ were used to determine the probability that the simulated neurons would fire. Spike-train processes were simulated for 221 epochs. Each epoch lasted 3500 ms with a total of 3649 spikes for neuron i and 9862 spikes for neuron j. These values were chosen to reproduce the parameters described in Vaadia *et al.* (1995). For simplicity the same transient was used for both neurons and corresponds to $t_i(t)$ above (see figure 1*a*). This transient represents a phasic evolution in the propensity to fire following the stimulus event. $x_i(t)$ and $x_j(t)$ were taken to reflect the relative firing propensity of neurons i and j. These processes (see figure 1*b*) were the sum of: (i) The transient multiplied by coefficients (a_i and a_j) selected from a normal bivariate distribution with zero mean and correlation 0.7 (i.e. $\mathrm{Cov}(a_i, a_j) = 0.7$ *and* $E(a_i) = E(a_j) = 0$). (ii) Residual processes obtained by convolving two random Gaussian innovations with a Gaussian kernel of full width at half maximum of 256 ms and taking a linear combination to introduce non-specific correlations between neurons i and j. These components correspond to $e_i(t)$ and $e_j(t)$ in the previous section. This sum was constrained to lie in the range (0, 1) using a suitably scaled error function. Simulated spike-trains were obtained by modulating the probability of firing according the above curves (see figure 1*c*).

The simulated spike trains were time-averaged in bins of 70 ms and subject to JPSTH analysis as described in Vaadia *et al.* (1995). The results of this analysis are seen in figure 2 and should be compared with figure 2*a* in Vaadia *et al.* (1995). The data corresponding to a conventional peri-stimulus time histogram (PSTH) shows little modulation (side panels). The main panel is an image representation of the cross-correlation matrix (referred to the stimulus event at 1000 ms) and reproduces the generic features reported in Vaadia *et al.*; namely a highly structured cross-correlation matrix with negligible modulation of mean firing rates.

(*b*) *Singular value decomposition*

The ability of SVD to 'recover' the underlying transients was demonstrated by applying SVD to the simulated data of the previous section. The estimated

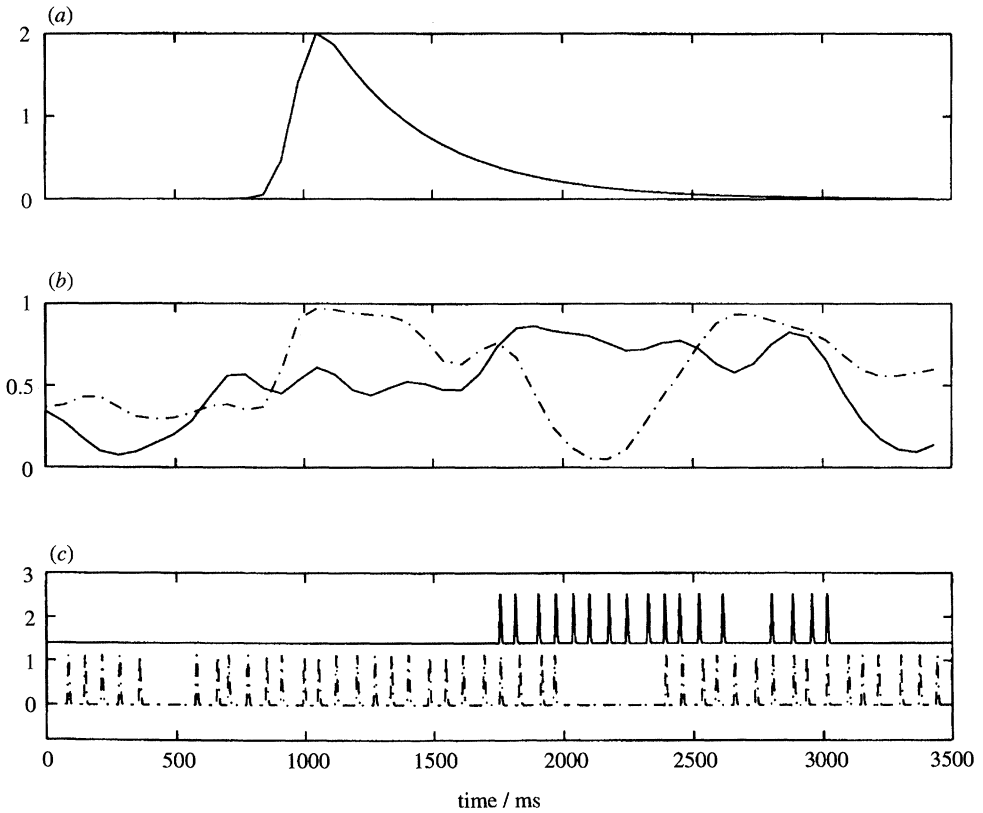

Figure 1. Details of the simulated spike-trains. Spike-train processes were simulated for 221 epochs. An example from one epoch is shown. (*a*) The transient used in the simulations, corresponding to $t_i(t)$ in the main text. (*b*) Processes $x_i(t)$ and $x_j(t)$: the relative propensity to fire for neurons i and j (solid and broken lines respectively). These processes are the sum of: (i) the transient in (*a*) multiplied by coefficients (a_i and a_j in the main text); and (ii) residual processes obtained by convolving random Gaussian innovations with a Gaussian kernel. These components correspond to $e_i(t)$ and $e_j(t)$ in the main text. This sum was constrained to lie in the range (0, 1) using a suitably scaled error function. (*c*) Simulated spike-trains obtained by modulating the probability of firing according the above curves.

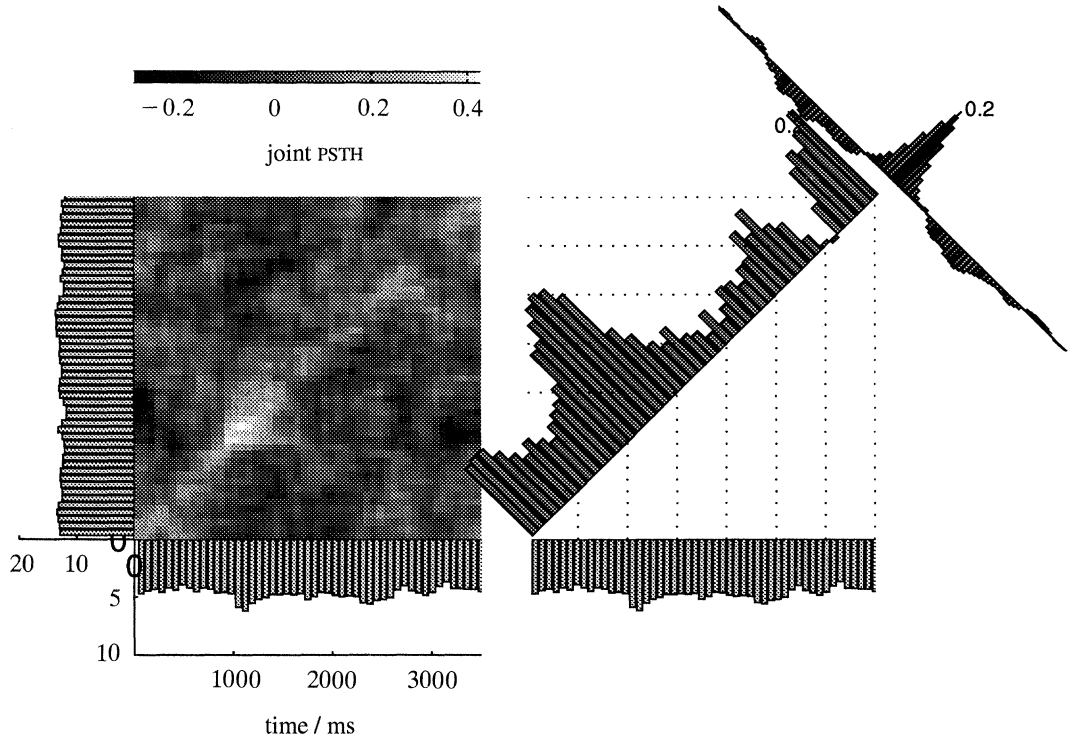

Figure 2. Joint peri-stimulus time histogram (JPSTH) based on the simulated data described in figure 1. The simulated spike trains were time-averaged in bins of 70 ms and subject to JPSTH analysis. The results of this analysis are seen above and should be compared with figure 2*a* in Vaadia *et al.* Left: mean firing rate (per second) averaged over epochs (side panels). These data correspond to a conventional peri-stimulus time histogram (PSTH) and show little modulation when referred to the stimulus events. The main panel is an image representation of the cross-correlation matrix (referred to the stimulus event at 1000 ms). Right: coincidence-time histogram (main diagonal) showing the time-dependent nature of the correlations (referred to the stimulus event) and conventional time-averaged cross-correlogram (upper right).

transients were taken to be the first singular vectors of $\boldsymbol{x}_i^T \boldsymbol{x}_j - \boldsymbol{e}_i^T \boldsymbol{e}_j$, where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ correspond to mean corrected matrices of binned firing rates with one row for each epoch and a column for each bin. $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ represent the firing rates obtained in the absence of simulated transients (in an experimental situation these data would be taken from epochs that did not include the stimulus event). Note the similarity between the actual (see figure 1*a*) and the first singular vectors or estimated form for the transients (see figure 3*a*). Unlike the JPSTH analysis, the SVD technique described here explicitly discounts non-specific correlations (i.e. $\boldsymbol{e}_i^T \boldsymbol{e}_j$). The importance of these non-specific effects can be noted in the JPSTHs in Vaadia *et al.* that show marked correlations before the stimulus arrives.

5. DISCUSSION

The biological mechanisms underlying the variable expression of transients could, as suggested (Vaadia *et al.* 1995), be explained in terms of population dynamics. Another explanation might relate to the modulation of prefrontal neurons by ascending neuromodulatory afferents. The phasic and transient responses of dopaminergic and cholinergic neurons to behaviourally salient stimuli (e.g. those predicting appetitive reward) have time courses on the order of 100–200 ms (for examples, see DeLong *et al.* 1983; Richardson & DeLong 1986; Ljungberg *et al.* 1992). This modulatory afferentation could contribute to the observed patterns of correlations. The relative contribution of dopaminergic and other modulatory inputs is clearly open to pharmacological study.

An important contribution made by Vaadia *et al.* was that the dynamic modulation of correlations was specific to the behavioural context (compare their figure 2*a* and 2*b*). In the present framework this corresponds to the expression of behaviourally specific or stimulus specific transients.

It is important to realize that the transients obtained with SVD are not necessarily the true underlying transients. This is because SVD finds a series of orthogonal (uncorrelated) transients and yet multiple and coincident transients expressed by neurons could be correlated both over epochs and over time within an epoch. It can be said that the underlying transients can be expressed in terms of a linear combination of the transients identified by SVD. If there is only one transient (i.e. one singular value is substantially larger than the rest) then the real and estimated transients could be assumed to be the same. A limitation of SVD is that it can only be applied to two spike trains and in this sense it is not as useful as alternative multivariate approaches.

One problem with SVD is that it does not provide for statistical inference about which transients are significant. For example, if the data were very noisy the spectrum of singular values would suggest that many transients were required to model the observed (event-

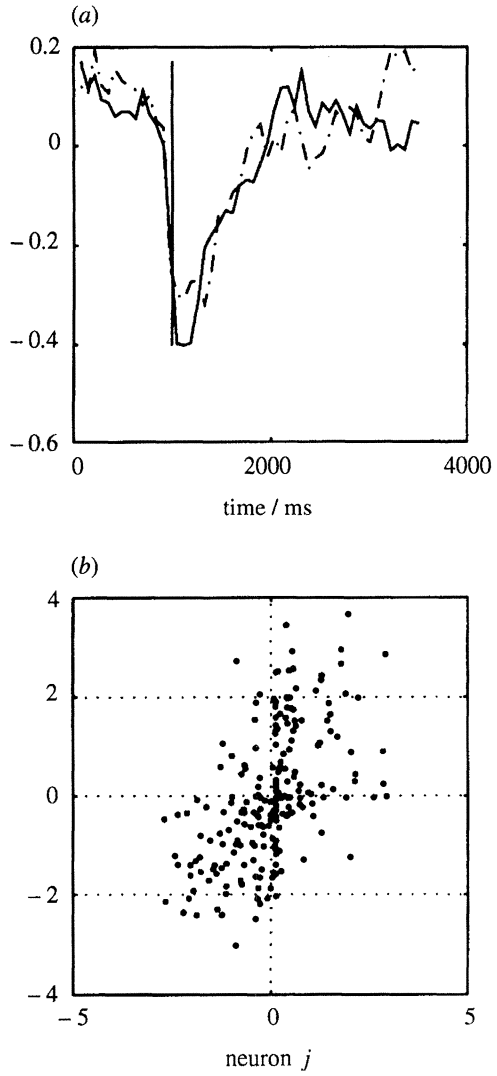

Figure 3. Singular value decomposition (SVD) of time-averaged (over 70 ms bins) firing rates. (*a*) The first singular vectors (estimated transients) following SVD of $\boldsymbol{x}_i^T \boldsymbol{x}_j - \boldsymbol{e}_i^T \boldsymbol{e}_j$. $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ correspond to mean corrected matrices of binned firing rates with one row for each epoch and a column for each bin. $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ represent the equivalent firing rates obtained in the absence of simulated transients. (*b*) The correlated expression of the transients are depicted by plotting $\boldsymbol{a}_i^1$ against $\boldsymbol{a}_j^1$. $\boldsymbol{a}_i^1$ and $\boldsymbol{a}_j^1$ represent the epoch by epoch event-related expression of $\boldsymbol{t}_i^1$ and $\boldsymbol{t}_j^1$ respectively and are given by $\boldsymbol{a}_i^1 = (x_i - \boldsymbol{e}_i).\boldsymbol{t}_i^1$ (similarly for $\boldsymbol{a}_j^1$).

referenced) cross-correlation matrix. However, it would not be possible to infer which of the estimated transients were the result of true neuronal transients and which were simply caused by noise. Clearly one would like an analytic approach that had the strengths of SVD and could be applied to more than two spike trains while allowing for statistical inference. Such an approach will be the subject of collaborative work between our unit and Vaadia *et al.*

In conclusion this paper puts forward an alternative, or indeed complementary, perspective on the rich temporal structure of event-related neuronal interactions. It has been shown that dynamic changes in coherence are equivalent to the coherent expression of neural transients. This may be important for characterizing neuronal interactions, because the expression of these transients may themselves show systematic time-dependent changes; for example over time scales pertinent to learning and plasticity.

I thank Ray Dolan for help during the development of these ideas. K.J.F. is funded by the Wellcome Trust.

REFERENCES

Foldiak, P. 1990 Forming sparse representations by local anti-Hebbian learning. *Biol. Cybern.* **64**(2) 165–170.

Gally, J. A., Montague, P. R., Reeke, G. N. & Edelman, G. M. 1990 The NO hypothesis: possible effects of a short lived, rapidly diffusible signal in the development and function of the nervous system. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3547–3551.

Hornik, K. & Kuan, C. M. 1992 Convergence analysis of local feature extraction algorithms. *Neural Networks* **5**, 229–240.

Kohonen, XX. 1982 Self organised formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69.

Linsker, R. 1988 Self organisation in a perceptual network. *Computer.* (March) 105–117.

Lopez, H. S., Burger, B., Dickstein, R., Desmond, N. L. & Levy, W. B. 1990 Associative synaptic potentiation and depression: quantification of dissociable modifications in the hippocampal dentate gyrus favors a particular class of synaptic modification equations. *Synapse* **5**, 33–47.

Miller, K. D. 1992 Models of activity-dependent neural development. *Neurosciences* **4**, 61–73.

Richardson, R. T. & DeLong, M. R. 1986 Nucleus basalis of Meynert neuronal activity during a delayed response task in monkey. *Brain Res.* **399**, 364–368.

Rubner, J. & Schulten, K. 1990 Development of feature detectors by self organisation: a network model. *Biol. Cybern.* **62**(3) 193–9.

Shadlen, M. N. & Newsome, W. T. 1994 Noise, neural codes and cortical organisation. *Curr. Opin. Neurobiol.* **4**, 569–579.

Vaadia, E., Haalman, I., Abeles, M. *et al.* 1995 Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature, Lond.* **373**, 515–518.

Willshaw, D. J. 1979 How to label nerve cells so that they can interconnect in an ordered fashion. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5176–5178.

Received 2 June 1995; accepted 20 June 1995