THE ROYAL SOCIETY

The labile brain. I. Neuronal transients and nonlinear coupling

Karl J. Friston

Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK
(*k.friston@fil.ion.ucl.ac.uk*)

In this, the first of three papers, the nature of, and motivation for, neuronal transients is described in relation to characterizing brain dynamics. This paper deals with some basic aspects of neuronal dynamics, interactions, coupling and implicit neuronal codes. The second paper develops neuronal transients and nonlinear coupling in the context of dynamic instability and complexity, and suggests that instability or lability is necessary for adaptive self-organization. The final paper addresses the role of neuronal transients through information theory and the emergence of spatio-temporal receptive fields and functional specialization.

By considering the brain as an ensemble of connected dynamic systems one can show that a sufficient description of neuronal dynamics comprises neuronal activity at a particular time and its recent history. This history constitutes a neuronal transient. As such, transients represent a fundamental metric of neuronal interactions and, implicitly, a code employed in the functional integration of brain systems. The nature of transients, expressed conjointly in distinct neuronal populations, reflects the underlying coupling among populations. This coupling may be synchronous (and possibly oscillatory) or asynchronous. A critical distinction between synchronous and asynchronous coupling is that the former is essentially linear and the latter is nonlinear. The nonlinear nature of asynchronous coupling enables the rich, context-sensitive interactions that characterize real brain dynamics, suggesting that it plays a role in functional integration that may be as important as synchronous interactions. The distinction between linear and nonlinear coupling has fundamental implications for the analysis and characterization of neuronal interactions, most of which are predicated on linear (synchronous) coupling (e.g. cross-correlograms and coherence). Using neuromagnetic data it is shown that nonlinear (asynchronous) coupling is, in fact, more abundant and can be more significant than synchronous coupling.

Keywords: neuronal transients; complexity; functional integration; neural codes; selection; self-organization

1. INTRODUCTION

This paper is about the dynamical aspects of brain function. Brain states are inherently labile, with a complexity and transience that renders their invariant characteristics elusive. The position adopted in these articles is that the most fruitful approach to understanding brain dynamics is to study this instability and transience. The aim of this paper is to introduce the notion of neuronal transients and the underlying framework, within which issues such as neuronal coupling, neuronal codes, functional integration, self-organization and the special complexity of brain dynamics can be addressed. The central tenet is that the behaviour of neuronal systems can be viewed as a succession of transient spatio-temporal patterns of activity that mediate adaptive perceptual synthesis and sensorimotor integration. This integration is shaped by the brain's anatomical infrastructure, principally connections, that has been selected to ensure the adaptive nature of the dynamics that ensue. Although rather obvious, this formulation embodies one fundamental point; namely that any proper description of brain dynamics should have an explicit temporal dimension. In other words, measures of brain activity are only meaningful when specified over extended periods of time. Simply appreciating this fact leads to quite compelling insights about brain organization and places some extant concepts in a more general context. The first example, considered in this paper, is that of neuronal codes: when trying to construct a taxonomy of neuronal codes it becomes clear that existing formulations are special cases of a more generic transient coding. This is particularly important in relation to fast dynamic interactions among neuronal populations that are characterized by synchrony. Synchronization has become popular in the past years (e.g. Eckhorn *et al*. 1988; Gray & Singer 1989; Engel *et al*. 1991) and yet may represent only one domain in the possible and, as will be shown, actual universe of interactions. Transient coding subsumes both synchronous and asynchronous interactions and it is the latter which mediate the nonlinear and context-sensitive features of brain dynamics.

The arguments presented in these papers depend, in part, on a mathematical formulation that is developed to reinforce, illustrate and, at times, motivate the ideas

Received 2 September 1998 *Accepted* 8 June 1999

introduced. Important mathematical derivations are provided in the appendices for the interested reader, while only key equations are presented in the main text. In general it is the form of these equations that is important, not their content. Details concerning data acquisition and simulation parameters are provided in the figure legends. This paper is divided into six sections. In § 2 we review the conceptual and mathematical basis of neuronal transients. This section uses a fundamental equivalence, between two mathematical formulations of nonlinear systems, to show that descriptions of brain dynamics, in terms of (i) neuronal transients and (ii) the effective connectivity among interacting brain systems, is complete and sufficient. In § 3, the ensuing framework is used to motivate a taxonomy of putative neuronal codes, the relationships among them and the predictions that arise. In § 4, we review the evidence for neuronal transients in terms of phenomena such as 'dynamic correlations' and nonlinear interactions between brain regions evidenced by asynchronous coupling. This section concludes with a direct test of the transient hypothesis, using data acquired with magnetoencephalography (MEG) that is based on the predictions from § 2. Section 5 addresses the general relationship between asynchronous coupling and nonlinear interactions, leading to a discussion in § 6 of the neurobiological mechanisms (e.g. modulatory effects) that might mediate them.

2. NEURONAL TRANSIENTS

(a) *Neuronal transients and time*

The assertion that teleologically meaningful measures or metrics of brain dynamics have an explicit temporal domain is neither new nor contentious (e.g. Von der Malsburg 1981; Optican & Richmond 1987; Engel *et al*. 1991; Aertsen *et al*. 1994; Freeman & Barrie 1994; Abeles *et al*. 1995; deCharms & Merzenich 1996). A straightforward analysis demonstrates its veracity: the brain is a highly nonlinear, spatially extended system that is unique in relation to other complex systems by virtue of its connectedness. The brain's architecture can be regarded as an ensemble of connections, where the nature and organization of these connections entails the substance of the system. The signals that traverse connections (axons and dendritic cell processes) do so in a finite amount of time. Suppose that one wanted to posit a sufficient metric that described the brain as a dynamical system in terms of neuronal activity. A natural choice would be the state variable $\boldsymbol{x}$ in a state equation

$$\partial \boldsymbol{x}(t)/\partial t = f(\boldsymbol{x}, \boldsymbol{C}), \tag{1}$$

where $\boldsymbol{x}$ is a large vector of activities for each unit in the brain. These activities could be expressed in many ways, for example firing at the initial segment of an axon or local field potentials of neuronal populations. Equation (1) simply says that the change in activity with time $\partial \boldsymbol{x}(t)/\partial t$ is a function of $\boldsymbol{x}$ and $\boldsymbol{C}$, a collection of control parameters corresponding to the underlying, time-invariant, connection strengths (e.g. synaptic efficacy). However, equation (1) would not be sufficient because it may take several, possibly tens of, milliseconds for the activity in one neuron, or population, to propagate to its recipient. So the change in any unit is a function not just of activity elsewhere at time t but at time t and in the recent past. This leads to the equation

$$\boldsymbol{x}(t) = f(\boldsymbol{x}(t-u), \boldsymbol{C}). \tag{2}$$

Equation (2) is, in principle, a sufficient description of brain dynamics and involves the variable $\boldsymbol{x}(t-u)$, which represents activity at all times u preceding the moment in question. $\boldsymbol{x}(t-u)$ is simply a neuronal transient (albeit a very global one). The degree of transience depends on how far back in time it is necessary to go to fully capture the brain's dynamics. In less abstract terms, if we wanted to determine the behaviour of a cell in the primary visual cortex (V1), then we would need to know the activity of all connected cells in the immediate vicinity (say within the same cortical column) over the last millisecond or so. We would also need to know the activity in distant sites, like the lateral geniculate nucleus (LGN) and higher cortical areas that send afferents, some ten or more milliseconds ago. In short, we need the recent history of all inputs.

Transients can be expressed in terms of firing rates (e.g. chaotic oscillations; Freeman & Barrie 1994) or individual spikes (e.g. synfire chains; Abeles *et al*. 1994, 1995). In what follows, we will assume that the relevant measurements pertain to those behaviours of cells that can influence other cells. There is a fundamental reason for this, which will become apparent below. The analysis above is not just a mathematical abstraction, it has very real implications at a number of levels: for example, the emergence of fast oscillatory interactions among simulated neuronal populations depends on the time-delays implicit in axonal transmission and the time constants of postsynaptic responses. Another slightly more subtle aspect of this formulation is that changes in synaptic efficacy, such as short-term potentiation or depression, take some time to be mediated by intracellular mechanisms. This means that the interaction between $\boldsymbol{x}(t-u)$ and $\boldsymbol{C}$, that models these activity-dependent effects in equation (2), again depends on the relevant history of activity.

(b) *Different levels of description*

An alternative perspective, on the necessity of going back in time to acquire a sufficient description of neuronal dynamics, is buried in the phrase above; 'a sufficient metric that describes the brain as a dynamical system in terms of neuronal activity'. This perspective is a little abstract, but provides a strong basis for neuronal transients. By restricting ourselves to measuring neuronal activity, there are a vast number of critical variables that are being ignored (e.g. the electro- and biochemical state of every cell process in the brain). If we knew every one of them then equation (1) might be a tenable model, constituting a microscopic level of description that would be entirely sufficient and complete. However, because we do not have access to this complete ensemble of 'hidden' variables, we are apparently unable to ever describe brain dynamics properly. This is not necessarily the case.

(i) *A fundamental equivalence*

Assume that every neuron in the brain is modelled by some immensely complicated nonlinear dynamical system of the sort described by equation (1), where the state

variables range from depolarization at every point in the dendritic tree to the phosphorylation status of every relevant enzyme. The input to this system corresponds to afferent activity and the output to firing at the cell's initial segment. Notice that both the input and the output are homologous in that they both measure that aspect of a system's (cell's) behaviour that can influence another system. Under these assumptions it can be shown that the output is a function of the recent history of its inputs. Furthermore this relationship can be expressed as a Volterra series of the inputs (see Appendix A and § 2(c)). The critical thing here is that we never need to know the underlying and 'hidden' variables that describe the details of each cell's electrochemical and biochemical status, we only need to know the history of its inputs, which, of course, are the outputs of other cells (including the one in question). This leads to a conceptual model of the brain as a collection of dynamical systems (e.g. cells or populations of cells), each of which is represented as an input–state–output model, where the state remains, for us, forever hidden. However, the inputs and outputs are accessible and are causally related where, in this special case of massively connected systems, the output of one system constitutes the input to another. A complete description therefore comprises the nature of these relationships (the Volterra series) and the neuronal transients (past history of all inputs). This constitutes a mesoscopic level of description, which allows a certain degree of 'black-boxness' as long as there is no loss of information or precision in specifying the interactions among the black boxes (cells or populations).

The equivalence, in terms of specifying the behaviour of a neuronal system, between microscopic and mesoscopic levels of description is critical and one that is central to this paper and neuronal transients. In short, the equivalence means that all the information inherent in the unobservable microscopic variables that determine the response of a neuronal system is embodied in the history of its observable inputs and outputs. This means that neuronal transients are a sufficient description of a system which eschew the measurement of hidden variables when predicting responses. Although the microscopic level of description may be more causally interpretable, from the point of view of response prediction, neuronal transients are an equivalent representation.

We have focused above on the distinction between microscopic and mesoscopic levels of description. The macroscopic level is reserved for approaches, exemplified by synergistics (Haken 1983), that try to characterize the spatio-temporal evolution of brain dynamics in terms of a small number of macroscopic order parameters (see Kelso (1995) for an engaging exposition). For example, macroscopic variables can be extracted from large-scale observations, such as magnetoencephalography (MEG), using the order parameter concept: order parameters are created and determined by the cooperation of microscopic quantities and yet, at the same time, govern the behaviour of the whole system. We will not deal with these approaches here but interested readers are referred to Jirsa *et al*. (1995) for an example.

(c) *A nonlinear framework*

The fact that a mesoscopic level of description exists suggests that (i) a complete description of dynamics could be cast in terms of neuronal transients; and (ii) a complete model of effective connectivity (i.e. the causal influences that one neuronal system exerts over another) should take the form of a Volterra series. These are quite fundamental conclusions. The primary focus of these papers is the first conclusion: namely, one can either try to measure every aspect of brain function and characterize the dynamics in terms of equation (1), or one can identify the essential inputs and outputs of its components and work explicitly with their recent history, i.e. frame the dynamics in terms of neuronal transients as in equation (2). The former is impossible. The latter is the subject of this paper.

Volterra series offer a very general form for the functions in equation (2) and can be expressed as

$$x_i(t) = \Omega_0[\boldsymbol{x}(t)] + \Omega_1[\boldsymbol{x}(t)] + \ldots + \Omega_n[\boldsymbol{x}(t)] + \ldots, \qquad (3)$$

where $x_i(t)$ is the activity of the ith unit. $\Omega_n[\cdot]$ is the nth order Volterra operator and has associated with it a kernel or smoothing function h that operates on the recent history of the inputs. Volterra series are functional Taylor expansions and are generally thought of as nonlinear convolutions or polynomial expansions with memory (see Appendix A). We have found this nonlinear system identification framework very useful when characterizing neuromagnetic and haemodynamic time-series from functional magnetic resonance imaging (fMRI) and it is used below many times. See also Stevens (1994).

The distinction, and equivalence, between microscopic and mesoscopic levels of description is illustrated in figure 1. Clearly this equivalence cannot be demonstrated for real neuronal systems but can be shown using reasonably realistic synthetic or model systems. Figure 1 contains a schematic that summarizes the equations behind the simulations used in these papers (see Appendix B). Collectively these equations are an example of equation (1) because they deal with all the relevant state variables (depolarization, channel configuration, discharge probability, etc.) and control parameters (synaptic efficacies, time constants, etc.). These equations constitute a microscopic level of description enabling one to predict the evolution of the system given only its state at one point in time. This system of differential equations has an entirely equivalent description in terms of the Volterra kernels that mediate between inputs and the responses (the first few are shown in figure 1*b*). These kernels can be applied to incoming transients, to give the responses, without knowing the underlying state variables. All that is needed is the input over a period of time.

For completeness, it can be noted that the Volterra formulation, based on the recent history of a system's inputs, is conceptually related to temporal embedding used to 'reconstruct' the dynamics of a system given only one variable. Temporal embedding involves using the current value of the state variable and a succession of values at a number of discrete times in the past. See Muller-Gerking *et al*. (1996) for a useful discussion of this approach in relation to the nonlinear characterization of neuronal time-series.

Arguments like those above suggest that the 'neuronal moment' lasts for tens if not hundreds of milliseconds and that the instantaneous behaviour of neuronal units cannot be divorced from their immediate temporal context. This

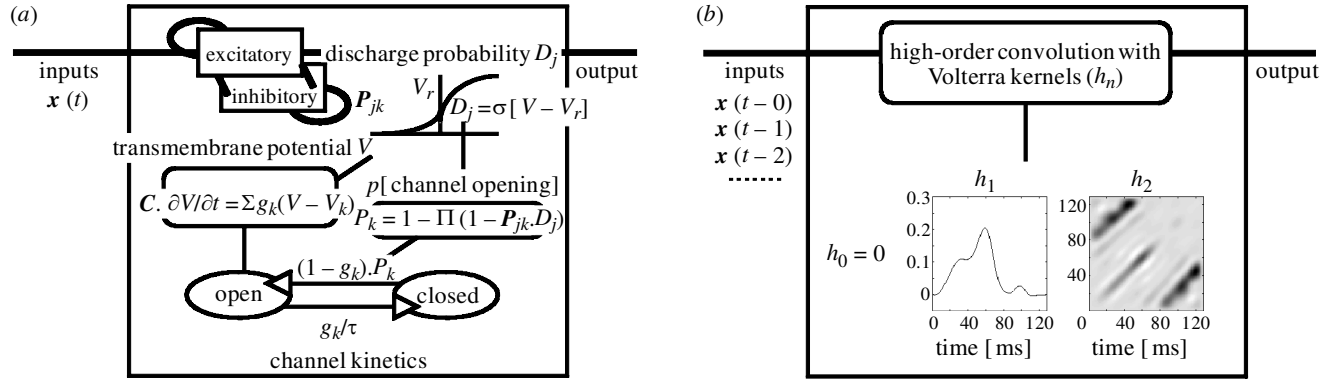

Figure 1. Schematic illustrating the distinction between (*a*) a microscopic level of description of a neuronal system (e.g cell or population) where all the hidden variables are known, enabling the output to be causally related to the instantaneous input. and (*b*) a more 'black-box' mesoscopic level in which it is only necessary to know the recent history of the inputs to determine the outputs. The left panel details the operational equations of simulated neuronal populations used in subsequent sections and the right panel gives an example of the Volterra kernels that characterize the ensuing input–output relationships. The simulated populations are described in detail in Appendix B and comprise two subpopulations (one excitatory and one inhibitory). These populations are described in terms of the mean transmembrane potential (V) and the probability that constituent units of subpopulation j will fire (D_j) in a deterministic way. The connectivity between them is described in terms of P_{jk} the probability that a discharge event in j will open a postsynaptic channel in subpopulation k. The probability of channel opening P_k, is computed by considering all potential inputs, including extrinsic inputs. The probability of channel opening enters into a first-order kinetic model of channel configuration for all k channel types. The expected proportion of open channels g_k in turn mediates changes in transmembrane potential though conductance changes and the depolarization relative to the equilibrium potential V_k for that channel. Finally the discharge probability is computed as a sigmoid function of V and the effective reversal potential V_r. The same input–output behaviour can be emulated by convolving the recent history of the inputs with a series of Volterra kernels of increasing order. In the example shown zeroth-, first- and second-order kernels are shown for the AMPA simulations depicted in figure 4. In this instance the input was taken to be the injected current and the output corresponded to the simulated local field potential. These kernels were estimated using least squares after expressing them in terms of basis functions (eighth-order discrete sine set over 128 ms) as described in Appendix E.

is not startling and is similar to noting that 'population codes' are necessarily distributed over many units. Neuronal transients take this one step further and stipulate that 'transient codes' are necessarily distributed over time. Looked at in this way, neural transients are a natural extension of the trend to characterize brain dynamics in relation to the context in which they occur. Neuronal transients represent an attempt to generalize the notion of population dynamics into the temporal domain.

(d) *Effective connectivity and Volterra kernels*

The second conclusion above (a complete model of effective connectivity should take the form of a Volterra series) implies that a complete characterization of effective connectivity, among neuronal systems, can be framed in terms of the Volterra kernels associated with the transformations of, and interactions among, inputs that yield the outputs.

(i) *Effective connectivity*

Functional integration is usually inferred using correlations among measurements of neuronal activity in different brain systems. In imaging neuroscience, the term 'functional connectivity' denotes the simple presence of these correlations (Friston 1995*a*). However, correlations can arise in a variety of ways. For example, in multi-unit electrode recordings they can result from stimulus-locked transients evoked by a common input, or reflect stimulus-induced oscillations mediated by synaptic connections (Gerstein & Perkel 1969; Gerstein *et al*. 1989). Integration within a distributed system is better understood in terms of 'effective connectivity'. Effective connectivity refers explicitly to 'the influence that one neural system exerts over another, either at a synaptic (i.e. synaptic efficacy) or population level' (Friston 1995*a*). It has been proposed (Aertsen & Preißl 1991) that 'the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons'. This speaks to two important points: (i) effective connectivity is dynamic; and (ii) it depends on a model of the interactions.

If effective connectivity is the influence that one neural system exerts over another, it should be possible, given the effective connectivity and the afferent activity, to predict the response of a recipient population. This is precisely what Volterra kernels do. Any model of effective connectivity will be a special case of a Volterra series and any measure of effective connectivity can be reduced to a set of Volterra kernels. An important aspect of effective connectivity is its context sensitivity. Effective connectivity is simply the 'effect' that an input has on the output of a target system. This effect will be sensitive to the history of the inputs (and outputs) and, of course, the microscopic state and causal architecture intrinsic to the target population. This intrinsic dynamical structure is embodied in the Volterra kernels and the current state of the target population enters thought the history of the outputs, that can re-enter as inputs. In short, Volterra kernels are synonymous with effective connectivity because they characterize the measurable 'effect' that an input has on its target. The use

of Volterra kernels in characterizing effective connectivity will be dealt with elsewhere.

3. NEURONAL CODES

(a) *Different sorts of code*

The conjecture that functional integration may be mediated by the mutual induction and maintenance of stereotyped spatio-temporal patterns of activity (i.e. neuronal transients) in distinct neuronal populations was presented in Friston (1995*b*, 1997*a*). Functional integration refers here to the concerted interactions among neuronal populations that mediate perceptual binding, sensorimotor integration and cognition. It pertains to the mechanisms of, and constraints under which, the dynamics of one population influence those of another. It has been suggested by many, that integration among neuronal populations uses transient dynamics that represent a temporal 'code'. A compelling proposal is that population responses, encoding a percept, become organized in time, through reciprocal interactions, to discharge in synchrony (Milner 1974; Von der Malsburg 1985; Singer 1994). The use of the term 'encoding' here speaks directly to the notion of codes.

A 'code' is used here to mean a measurement or metric of neuronal activity that captures teleologically meaningful transactions among different parts of the brain. No attempt is made to discern the meaning or content of a putative code. All that is assumed is that a code or metric must necessarily show some dependency when used to assess two interacting neuronal populations or brain areas. In other words, a neuronal code is a metric that reveals interactions among neuronal systems by enabling some prediction of the activity in one population given the activity in another. Clearly, from § 2, neuronal transients represent the most generic form of code. Given that neuronal transients have a number of attributes (e.g. their duration, mean level of firing, predominant frequency, etc.), any one of these attributes is a contender for a more parsimonious code. Although the term code is not being used to denote anything that 'codes' for something in the environment, it could be used to define some aspect of a sensory evoked transient that had a high mutual information with a sensory parameter that was manipulated experimentally (e.g. Optican & Richmond 1987; Tovee *et al*. 1993).

Given the above definition, the problem of identifying possible codes reduces to establishing which metrics are mutually predictive or statistically dependent when applied to two connected neuronal systems. This is quite an important point and leads to a very clear formulation of what can and cannot constitute a code and the different sorts of codes that might be considered. Furthermore using this operational definition, the problem of finding the right code(s) reduces to identifying the form of the Volterra operators in equation (3), for if we know these we can predict exactly what will ensue in any unit, given the dynamics elsewhere. Conversely, it follows that the different forms that equation (3) can take should specify the various codes likely to be encountered. We will return to this point in a subsequent section.

To discuss the nature of neuronal transients, in relation to codes, a taxonomy is now introduced. The most general form of coding is considered to be transient coding. All other codes are special cases, or special cases of special cases. The most obvious special case of a transient is when that transient shrinks to an instant in time. The associated codes will be referred to as instantaneous codes that subsume temporal coding and rate coding, depending on the nature of the metric employed. A more important distinction is whether two transients in two neuronal systems are synchronous or asynchronous, leading to the notion of synchronous codes and asynchronous codes. This synchronization may in turn be oscillatory or not, leading to oscillatory codes or non-oscillatory codes. Therefore oscillatory codes are a special case of synchronous codes that are themselves special cases of transient codes. This hierarchial decomposition is shown schematically in figure 2. This taxonomy is now reviewed in more detail, working from the special cases to the more general.

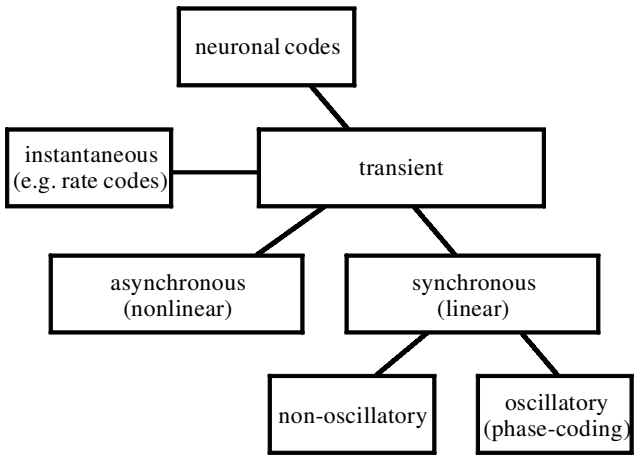

Figure 2. A simple taxonomy of neuronal codes, where commonly appreciated forms of encoding are seen as special cases of each other.

(i) *Instantaneous codes: temporal and rate coding*

The distinction between temporal coding and rate coding (see Shadlen & Newsome 1995; de Ruyter van Steveninck *et al*. 1997) centres on whether the precise timing of individual spikes is sufficient to facilitate meaningful neuronal interactions. In temporal coding, the exact time at which an individual spike occurs is the important measure and the spike-train is considered as a point process. The term temporal coding is used here in this restricted sense, as opposed to designating codes that have a temporal domain (e.g. Von der Malsburg 1985; Singer 1994). There is a critical distinction between instantaneous temporal codes and those that invoke some temporal patterning of spikes over time. The former include, for example, the instantaneous phase relationship between a spike and some reference oscillation. Although, in simple systems, knowledge of this phase will allow some prediction of responses in a target unit, it would not be sufficient for more nonlinear systems where one would need to know the history of phase modulation. The second sort of temporal code is distributed over time and represents a transient code. There are clear examples of these codes that have predictive validity, for example, the primary cortical representation of sounds by the

coordination of action potential timing (deCharms & Merzenich 1996). These codes depend on the relative timing of action potentials and implicitly, by appealing to an extended temporal frame of reference, fall into the class of transient codes. A very good example of this is provided by the work of de Ruyter van Steveninck *et al*. (1997), who show that the temporal patterning of spike-trains, elicited in fly motion-sensitive neurons by natural stimuli, carry twice the amount of information as an equivalent (Poisson) rate code.

Instantaneous rate coding considers spike-trains as stochastic processes whose first-order moments (i.e. mean activity) provide a metric with which neuronal interactions are enacted. These moments may be in terms of spikes themselves or other compound events (e.g. the average rate of bursting; Bair *et al*. 1994). The essential aspect of rate coding is that a complete metric would be the average firing rates of all the system's components at one instant in time. Interactions based on rate coding are usually assessed in terms of cross-correlations and many models of associative plasticity are predicated on these correlated firing rates (e.g. Hebb 1949). In this paper, instantaneous rate codes are considered insufficient as proper descriptions of neuronal interactions because, in the absence of 'hidden' microscopic variables, they are not useful. This is because they predict nothing about a cell, or population, response unless one knows the microscopic state of that cell or population.

(ii) *Synchronous codes: oscillatory and non-oscillatory codes*

The proposal most pertinent to these forms of coding is that population responses, participating in the encoding of a percept, become organized in time through reciprocal interactions so that they come to discharge in synchrony (Von der Malsburg 1985; Singer 1994) with regular periodic bursting. Frequency-specific interactions and synchronization are used synonymously in this paper. It should be noted that synchronization does not necessarily imply oscillations. However, synchronized activity is usually inferred operationally by oscillations implied by the periodic modulation of cross-correlograms of separable spike-trains (e.g. Eckhorn *et al*. 1988; Gray & Singer 1989) or measures of coherence in multichannel electrical and neuromagnetic time-series (e.g. Llinas *et al*. 1994). The underlying mechanism of these frequency-specific interactions is usually attributed to phase-locking among neuronal populations (e.g. Sporns *et al*. 1992; Aertsen & Preißl 1991). The key aspect of these metrics is that they refer to the extended temporal structure of synchronized firing patterns, either in terms of spiking (e.g. synfire chains; Abeles *et al*. 1994; Lumer *et al*. 1997) or oscillations in the ensuing population dynamics (e.g. Singer 1994).

One important aspect, that distinguishes oscillatory from non-oscillatory codes, is that the former can embody consistent phase relationships that may play a mechanistic role in the ontology of adaptive dynamics (e.g. Tononi *et al*. 1992). This has been proposed for theta rhythms in the hippocampus and more recently for gamma rhythms (e.g. Burgess *et al*. 1994; see Jefferys *et al*. (1996) for further discussion).

Many aspects of functional integration and feature linking in the brain are thought to be mediated by synchronized dynamics among neuronal populations (Singer 1994). Synchronization reflects the direct, reciprocal exchange of signals between two populations, whereby the activity in one population influences the second, such that the dynamics become entrained and mutually reinforcing. In this way the binding of different features of an object may be accomplished, in the temporal domain, through the transient synchronization of oscillatory responses (Von der Malsburg 1981). This 'dynamical linking' defines their short-lived functional association. Physiological evidence is compatible with this theory (e.g. Engel *et al*. 1991): synchronization of oscillatory responses occurs within, as well as among, visual areas, for example between homologous areas of the left and right hemispheres and between areas at different levels of the visuomotor pathway (Engel *et al*. 1991; Roelfsema *et al*. 1997). Synchronization in the visual cortex appears to depend on stimulus properties, such as continuity, orientation and motion coherence. Synchronization may therefore provide a mechanism for the binding of distributed features and contribute to the segmentation of visual scenes. More generally, synchronization may provide a powerful mechanism for establishing dynamic cell assemblies that are characterized by the phase and frequency of their coherent oscillations.

The problem with these suggestions is that there is nothing essentially dynamic about oscillatory interactions. As argued by Erb & Aertsen (1992), 'the question might not be so much how the brain functions by virtue of oscillations, as most researchers working on cortical oscillations seem to assume, but rather how it manages to do so in spite of them'. In order to establish dynamic cell assemblies, it is necessary to create and destroy synchronized couplings. It is precisely these dynamic aspects that render synchronization *per se* relatively uninteresting but speak to changes in synchrony (e.g Desmedt & Tomberg 1994) and the transitions between synchrony and asynchrony as the more pertinent phenomenon.

(iii) *Transient coding: synchronous and asynchronous codes*

An alternative perspective on neuronal codes is provided by work on dynamic correlations (Aertsen *et al*. 1994) as exemplified in Vaadia *et al*. (1995). 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. One implication of this work is that a complete model of neuronal interactions has to accommodate dynamic changes in correlations, modulated on time-scales of 100–1000 ms. A simple explanation for these dynamic correlations has been suggested (Friston 1995*b*): it was pointed out that the coexpression of neuronal transients in different parts of the brain could account for dynamic correlations (see § 4). This transient hypothesis suggests that interactions are mediated by the expression and induction of reproducible, highly structured spatio-temporal dynamics that endure over several hundred milliseconds. As in synchronization coding, the dynamics have an explicit temporal dimension but there is no special dependence on oscillations or synchrony. In particular, the frequency structure of a transient in one part of the brain may be very different from that in another. In synchronous interactions the frequency

structures of both will be the same (whether they are oscillatory or not).

If the transient model is correct then important transactions among cortical areas will be overlooked by techniques that are predicated on rate coding (correlations, covariance patterns, spatial modes, etc.) or synchronization models (e.g. coherence analysis and cross-correlograms). Clearly the critical issue here is whether one can find evidence for asynchronous interactions that would render the transient level of the taxonomy a useful one (see figure 2). Such evidence would speak to the importance of neuronal transients and place synchronization in a proper context. In § 4, we review the indirect evidence for neuronal transients and then provide direct evidence by addressing the relative contribution of synchronous and asynchronous coupling to interactions between brain areas.

4. THE EVIDENCE FOR NEURONAL TRANSIENTS

We are all familiar with neuronal transients in the form of evoked transients in electrophysiology. However, the critical thing is whether transients in two neuronal populations mediate their own induction. For example, at the level of multi-unit micro-electrode recordings, correlations can result from stimulus-locked transients, evoked by a common afferent input, or reflect stimulus-induced interactions—phasic coupling of neuronal assemblies, mediated by synaptic connections (Gerstein & Perkel 1969; Gerstein *et al*. 1989). The question here is whether these interactions can be asynchronous? One important indication that stimulus-induced interactions are not necessarily synchronous comes from dynamic correlations.

(a) *Dynamic correlations and neuronal transients*

As mentioned above, Vaadia *et al*. (1995) presented compelling results concerning neuronal interactions in monkey cortex, enabling them to make two fundamental points: (i) it is possible that cortical function is mediated by the dynamic modulation of coherent firing among neurons; and (ii) that these time-dependent changes in correlations can emerge without modulation of firing rates. One implication is that a better metric of neuronal interactions could be framed in terms of dynamic changes in correlations. This possibility touches on the distinction between temporal coding and rate coding as described in the § 3. This distinction, and the related debate (e.g. 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*. (1995) 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. In Shadlen & Newsome (1995), 'precise' timing means synchronization within 1–5 ms. In contrast Vaadia *et al*. (1995) demonstrate looser coherence over a period of about 100 ms (using 70 ms time bins). A simple explanation for this temporally modulated coherence or dynamic correlation is provided by the notion of neuronal transients.

Imagine that two neurons respond to an event with a similar transient (a short-lived, stereotyped, time-dependent change in the propensity to fire). 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 of the epoch and then fade away, therein emulating a dynamic modulation of coherence. In other words, a transient modulation of covariance can be equivalently formulated as a covariance in the expression of transients. The generality of this equivalence can be established using singular value decomposition (SVD). Dynamic correlations are inferred on the basis of the cross-correlation matrix of the trial by trial activity as a function of peristimulus time. This matrix is referred to as the joint peristimulus time histogram (JPSTH) and implicitly discounts correlations due to stimulus-locked transients by dealing with correlations over trials (as opposed to time following the stimulus or event). Let $\boldsymbol{x}_i$ be a matrix whose rows contain the activities recorded in unit i over a succession of time bins following the stimulus, with one row for each trial. Similarly for $\boldsymbol{x}_j$. After these matrices have been normalized, the cross-correlation matrix is given by $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$ where T denotes transposition. By noting the existence of the singular value decomposition

$$\boldsymbol{x}_I^{\mathrm{T}}\boldsymbol{x}_j = \lambda_1\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{v}_1 + \lambda_2\boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{v}_2 + \lambda_3\boldsymbol{u}_3^{\mathrm{T}}\boldsymbol{v}_3 + \ldots, \tag{4}$$

one observes that any cross-covariance structure $\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$ can be expressed as the sum of covariances due to the expression of paired transients (the singular vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$). The expression of these transients covaries according to the singular values λ_k. In this model, any observed neuronal transient in unit i is described by a linear combination of the $\boldsymbol{u}_k$ (or $\boldsymbol{v}_i$ in unit j).

This is simply a mathematical device to show that dynamic changes in coherence are equivalent to the coherent expression of neural transients. In itself it is not important, in the sense that dynamic correlations are just as valid a characterization as neuronal transients and indeed may provide more intuitive insights into how this phenomenon is mediated at a mechanistic level (e.g. Riehle *et al*. 1997). What is important is that the existence of dynamic correlations implies the existence of transients that exist after stimulus-locked effects have been discounted. The next step is to find definitive evidence that transients underpin asynchronous coupling, or equivalently that coupled transients in two neuronal populations have a different form or frequency structure. The essential issue, that remains to be addressed, is whether a transient in one brain system, that mediates the expression of another transient elsewhere, has the same or a different temporal patterning of activity. The importance of this distinction will become clear below.

(b) *Synchrony, asynchrony and spectral density*

Synchronized, fast dynamic interactions among neuronal populations represent a possible mechanism for functional integration (e.g. perceptual binding) in the brain, but focusing on synchrony precludes a proper consideration of asynchronous interactions that may have an equally important and possibly distinct role. In this section, the importance of synchronization is evaluated in

relation to the more general notion of neural transients that allow for both synchronous and asynchronous interactions. Transients suggest that neuronal interactions are mediated by the mutual induction of stereotyped spatio-temporal patterns of activity among different populations. If the temporal structures of these transients are distinct and unique to each population, then the prevalence of certain frequencies in one cortical area should predict the expression of different frequencies in another. In contrast, synchronization models posit a coupled expression of the same frequencies. Correlations among different frequencies therefore provide a basis for discriminating between synchronous and asynchronous coupling.

Consider time-series from two neuronal populations or cortical areas. The synchrony model suggests that the expression of a particular frequency (e.g. 40 Hz) in one time-series will be coupled with the expression of the same (40 Hz) frequency in the other (irrespective of the exact phase relationship of the transients or whether they are oscillatory or not). In other words, the modulation of this frequency in one area can be explained or predicted by its modulation in the second. Conversely, asynchronous coupling suggests that the power at a reference frequency, say 40 Hz, can be predicted by the spectral density in the second time-series at some frequencies other than 40 Hz. These predictions can be tested empirically using standard time-frequency and regression analyses as exemplified below. These analyses are an extension of those presented in Friston (1997*a*) and confirm that both synchronous and asynchronous coupling are seen in real neuronal interactions. They use MEG data, obtained from normal subjects while performing self-paced finger movements. These data were kindly provided by Klaus Martin Stephan and Andy Ioannides.

(c) *Definitive evidence for asynchronous coupling*

After Laplacian transformation of multichannel magnetoencephalographic data, two time-series were selected. The first was an anterior time-series over the central prefrontal region and the second was a posterior parietal time-series, both slightly displaced to the left. These locations were chosen because they had been implicated in a conventional analysis of responses evoked by finger movements (Friston *et al*. 1996). Figure 3 shows an example of these data in the time domain $x(t)$ and in the frequency domain $g(\omega,t)$ following a time-frequency analysis. See Appendix C for a description of time-frequency analyses and their relation to wavelet transformations. In brief, they give the frequency structure of a time-series, over a short period, as a function of time. The time-frequency analysis shows the dynamic changes in spectral density between 8 and 64 Hz over about 16 s. The cross-correlation matrix of the parietal and prefrontal time-frequency data is shown in figure 3*b*. There is anecdotal evidence here for both synchronous and asynchronous coupling. Synchronous coupling, based on the co-modulation of the same frequencies, is manifest as hot spots along, or near, the leading diagonal of the cross-correlation matrix (e.g. around 20 Hz). More interesting are correlations between high frequencies in one time-series and low frequencies in another. In particular, note that the frequency modulation at about 34 Hz in the parietal region (second time-series) could be explained by several frequencies in the prefrontal region. The most profound correlations are with lower frequencies in the first time-series (26 Hz), but there are also correlations with higher frequencies (54 Hz) and some correlations with prefrontal frequencies around 34 Hz itself. The problem here is that we cannot say whether there is a true asynchronous coupling or whether there is simple synchronous coupling at 34 Hz with other higher and lower frequencies being expressed, in a correlated fashion, within the prefrontal region. These within-region correlations can arise from broad-band coherence (Bressler *et al*. 1993) or harmonics of periodic transients (see Jurgens *et al*. 1995). In other words, synchronous coupling at 34 Hz might be quite sufficient to explain the apparent correlations between 34 Hz in the parietal region and other frequencies in the prefrontal region. To address this issue we have to move beyond cross-correlations and make statistical inferences that allow for synchronous coupling over a range of frequencies within the prefrontal area. This is effected by treating the problem as a regression analysis and asking whether the modulation of a particular frequency in the parietal region can be explained in terms of the modulation of frequencies in the prefrontal region, starting with the model

$$g_2(\omega_0,t) = \sum \beta(\omega) \times g_1(\omega,t), \tag{5}$$

where $g_2(\omega,t)$ and $g_1(\omega,t)$ are the spectral densities from the parietal and prefrontal time-series and ω_0 is the frequency in question (e.g. 34 Hz). $\beta(\omega)$ are the parameters that have to be estimated. To allow for couplings among frequencies that arise from the correlated expression of frequencies within the areas, the predictor or explanatory variables $g_1(\omega,t)$ are decomposed into synchronous and asynchronous predictors. These correspond to the expression of the reference frequency in the prefrontal region $g_1(\omega_0,t)$ and the expression of all the remaining frequencies orthogonalized with respect to the first predictor $g_1(\omega,t)^*$ (see Appendix D). By orthogonalizing the predictors in this way we can partition the total variance in parietal frequency modulation into those components that can be explained by synchronous and asynchronous coupling, respectively,

$$g_2(\omega_0,t) = \beta(\omega_0) \times g_1(\omega_0,t) + \sum \beta(\omega)^* \times g_1(\omega,t)^*. \tag{6}$$

Furthermore by treating one of the predictors as a confound we can test, statistically, for the contribution of the other (i.e. either synchronous or asynchronous) using standard inferential techniques, in this instance the *F*-ratio based on a multiple regression analysis of serially correlated data (see Appendix D).

To recap for those less familiar with regression techniques, we take a reference frequency in one time-series (e.g. the parietal region) and try to predict it using the expression of all frequencies in the other (e.g. prefrontal region). To ensure that we do not confuse asynchronous and synchronous interactions due to broad-band coherence and the like, correlations with the reference frequency, within the prefrontal region, are removed from the predictors.

An example of the results of this sort of analysis are shown in figure 3*c*. Figure 3*c*(i) shows the proportion of variance that can be attributed to either synchronous (broken line) or asynchronous coupling (solid line) as a function of frequency (ω_0). In other words, the proportion of variance in parietal spectral density that can be predicted on the basis of changes in the same frequency in the prefrontal region (broken line) or on the basis of other frequencies (solid line). Figure 3*c*(ii) portrays the significance of these predictions in terms of the associated *p*-values and shows that both synchronous and asynchronous coupling are significant at 34 Hz (i.e. the middle peak in figure 3*c*(i),(ii)).

In contrast the high correlations between 48 Hz in the second time-series and 26 Hz in the first is well away from the leading diagonal in the cross-correlation matrix, with little evidence of correlations at either of these frequencies alone. The regression analysis confirms that, at this frequency, asynchronous coupling prevails. The only significant coupling is asynchronous (right peak in figure 3*c*(i),(ii)) and suggests that the expression of 48 Hz gamma activity in the parietal region is mediated, in part, by asynchronous interactions directly, or vicariously, with the prefrontal cortex. Note that a common modulating source, influencing both the parietal and prefrontal regions, cannot be invoked as an explanation for this sort coupling because this effect would be expressed synchronously.

(d) *Summary*

The above example was provided to illustrate both mixed and asynchronous coupling and to introduce the concept that simply observing correlations between different frequency modulations is not sufficient to infer asynchronous coupling. Broad-band coherence in the context of oscillations leads naturally to cross-frequency coupling and, more importantly, non-oscillatory but synchronous interactions will, as a matter of course, introduce them by virtue of the tight correlations between different frequencies within each area (e.g. Jurgens *et al*. 1995). By discounting these within time-series correlations, using the orthogonalization above, one can reveal any underlying asynchronous coupling. Results of this sort are fairly typical (we have replicated them using different subjects and tasks) and provide definitive evidence for asynchronous coupling. Generally our analyses show both synchronous and asynchronous effects, where the latter are typically greater in terms of the proportion of variance explained. As in the example presented here, it is usual to find both sorts of coupling expressed in the same data.

In conclusion, using an analysis of the statistical dependence between spectral densities measured at different points in the brain, the existence of asynchronous coupling can be readily confirmed. It is pleasing that such a simple analysis should lead to such an important conclusion. These results are consistent with transient coding and imply that correlations (rate coding) and coherence (synchrony coding) are neither complete nor sufficient characterizations of neuronal interactions and suggest that higher-order, more general interactions may be employed by the brain. In the remaining sections, the importance of asynchronous interactions and their mechanistic basis will be addressed using both simulated and real neuronal time-series.

5. COUPLING AND CODES

(a) *Asynchronous coupling and nonlinear interactions*

Why is asynchronous coupling so important? The reason is that asynchronous interactions embody all the nonlinear interactions implicit in functional integration and it is these that mediate the diversity and context-sensitive nature of neuronal interactions. The nonlinear nature of interactions between cortical brain areas renders the effective connectivity among them inherently dynamic and contextual. Compelling examples of context-sensitive interactions include the attentional modulation of evoked responses in functionally specialized sensory areas (e.g. Treue & Maunsell 1996) and other contextually dependent dynamics (see Phillips & Singer (1997) for an intriguing discussion).

One of the key motivations for distinguishing between synchronous and asynchronous coupling is that the underlying mechanisms are fundamentally different. In brief, it will be suggested that synchronization emerges from the reciprocal exchange of signals between two populations, wherein the activity in one population has a direct or 'driving' effect on the activity of the second. In asynchronous coding, the incoming activity from one population may exert a 'modulatory' influence, not on the activity of units in the second, but on the interactions among them (e.g. effective connectivity). This indirect influence will-lead to changes in the dynamics intrinsic to the second population that could mediate important contextual and nonlinear responses. Before addressing the neural basis of these effects in the next section, the relationship between asynchrony and nonlinear coupling is established and discussed in relation to neuronal codes.

Perhaps the easiest way to see that synchronized interactions are linear is to consider that the dynamics of any particular neuronal population can be modelled in terms of a Volterra series of the inputs from others. If this expansion includes only the first-order terms then the Fourier transform of the first-order Volterra kernel completely specifies the relationship between the spectral density of input and output in a way that precludes interactions among frequencies, or indeed inputs (as shown below). The very presence of significant coupling between frequencies, above and beyond covariances between the same frequencies, implies a first-order approximation is insufficient and, by definition, second- and higher-order nonlinear terms are required. In short, asynchronous coupling reflects the nonlinear component of neuronal interactions and as such is vital for a proper characterization of functional integration.

To see more explicitly why asynchronous interactions are so intimately related to nonlinear effects consider equation (3) where, for simplicity, we focus on the effects of unit j on unit i, discounting the remaining units.

$$\boldsymbol{x}_i(t) = \Omega_0[\boldsymbol{x}_j(t)] + \Omega_1[\boldsymbol{x}_j(t)] + \Omega_2[\boldsymbol{x}_j(t)] + \ldots, \tag{7}$$

where $\boldsymbol{x}_i(t)$ is the activity of one unit or population and $\boldsymbol{x}_j(t)$ another. By discounting the constant and high-order terms we end up with a simple convolution model of neuronal interactions:

$$\boldsymbol{x}_i(t) = \Omega_1[\boldsymbol{x}_j(t)] = h_1 \otimes \boldsymbol{x}_j(t), \tag{8}$$

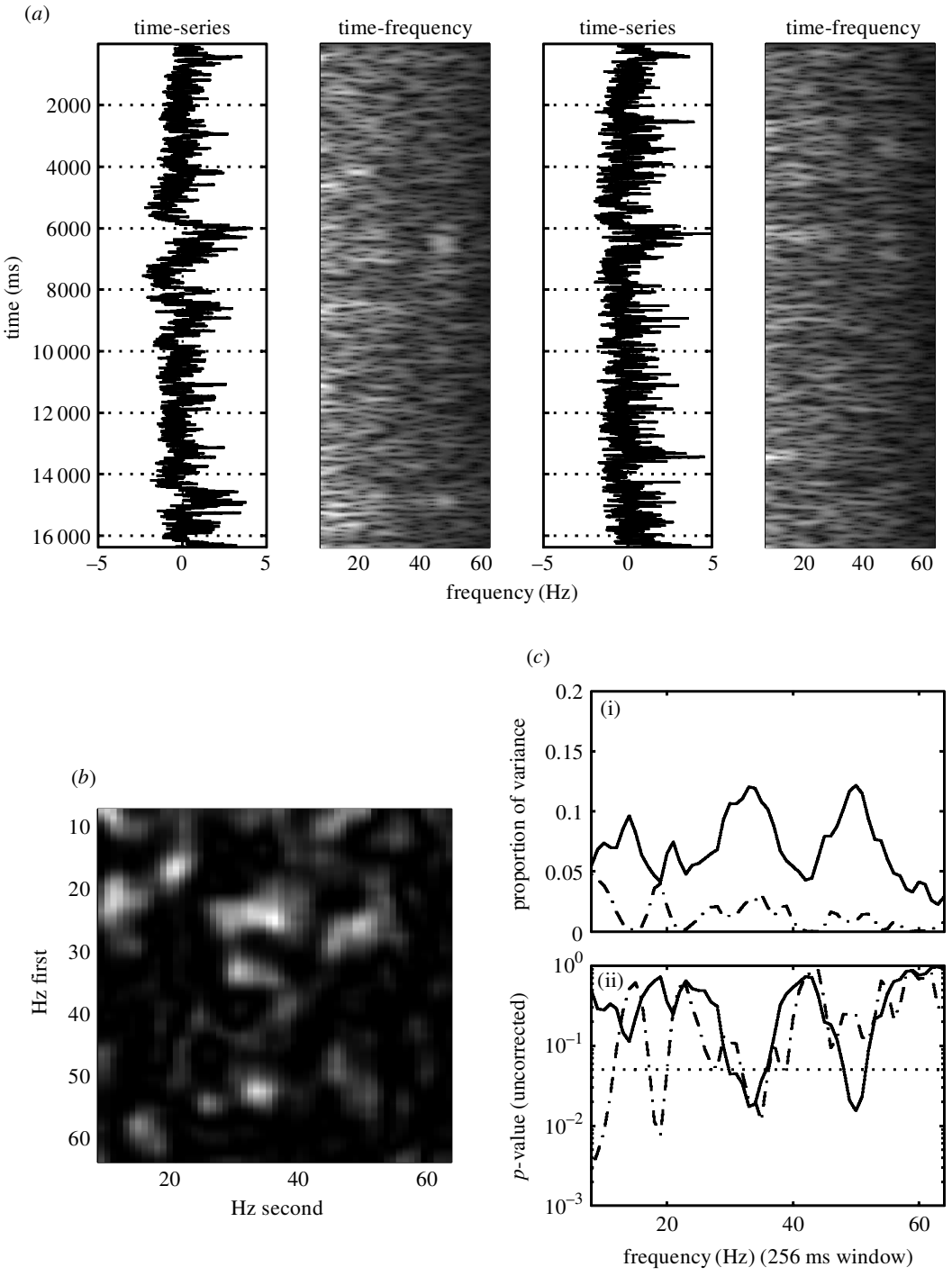

Figure 3. Time-frequency analysis of MEG time-series from two remote cortical regions designed to characterize the relative contribution of synchronous and asynchronous coupling; in terms of correlated changes in spectral density within and among frequencies respectively. Neuromagnetic data were acquired from normal subjects using a KENIKRON 37 channel MEG system at 1 ms intervals for periods of up to 2 min. During this time subjects were asked to make volitional joystick movements either in random directions, or to the left, every 2 s or so. Epochs of data comprising 2^{14} ms were extracted. ECG artefacts were removed by linear regression and the data were transformed using a V_3 transformation (i.e. Laplacian derivative (Ioannides *et al.* 1990)) to minimize spatial dependencies among the data. Paired epochs were taken from a left prefrontal and left parietal region that were subsequently bandpass filtered (1–128 Hz). The data in this figure come from a normal male performing leftwards movements. (*a*) The two times-series $x(t)$ (plots) and their corresponding time-frequency profiles $g(\omega,t)$ (images). The first time-series comes from the left prefrontal region roughly over the anterior cingulate and SMA. The second comes from the left superior parietal region. The data have been normalized to zero mean and unit standard deviation. The frequencies analysed were 8 Hz to 64 Hz in 1 Hz steps. (*b*) This is a simple characterization of the coupling among frequencies in the two regions and represents the cross-correlation matrix of the time-frequencies $g(\omega,t)$. In this display format the correlation coefficients have been squared. (*c*) These are the results of the linear regression analysis that partitions the amount of modulation in the second (parietal) time-series into components that can be attributed to synchronous (broken lines) and asynchronous (solid lines) contributions from the first (prefrontal) time-series (see main text and Appendix D). (i) The relative contribution in terms of the proportion of variance explained, and (ii) in terms of the significance using a semi-log plot of the corresponding p-values, both as

where $\otimes$ denotes convolution and h_1 is the first-order Volterra kernel. This can be expressed in the frequency domain as

$$g_i(\omega,t) = |l(\omega)|^2 \times g_j(\omega,t), \quad (9)$$

where $l(\omega)$ is known as the transfer function and is simply the Fourier transform of the first-order kernel h_1. This equality says that the expression of any frequency in unit i is predicted exactly by the expression of the same frequency in unit j (after some scaling by the transfer function). This is exactly how synchronous interactions have been characterized and furthermore is identical to the statistical model employed to test for synchronous interactions above. In equation (6), the parameters $\beta(\omega_0)$ are essentially estimates of $|l(\omega_0)|^2$ in equation (9). From this perspective the tests for asynchronous interactions in § 4 can be viewed as an implicit test of nonlinear interactions, while discounting a linear model as a sufficient explanation for the observed coupling. See Erb & Aertsen (1992) for an example of transfer functions that obtain after the equations, defining a simulated neuronal population, have been simplified to render them linear.

In summary, the critical distinction between synchronous and asynchronous coupling is the difference between linear and nonlinear interactions among units or populations. Synchrony implies linearity. The term 'generalized synchrony' has been introduced to include nonlinear interdependencies (see Schiff *et al.* 1996). Generalized synchrony therefore subsumes synchronous and asynchronous coupling. A very elegant method for making inferences about generalized synchrony is described in Schiff *et al.* (1996). This approach is particularly interesting from our point of view because it implicitly uses the recent history of the dynamics through the use of temporal embedding to reconstruct the attractors analysed. However, unlike our approach based on a Volterra series formulation, it does not explicitly partition the coupling into synchronous and asynchronous components.

(b) *The taxonomy revisited*

Relating synchrony and asynchrony directly to the Volterra series formulation leads to a more formal and principled taxonomy of putative neuronal codes. Recall that the first level of the taxonomy distinguishes between transient codes and instantaneous codes. In terms of the Volterra model the latter are a special case of equation (7), when all the Volterra kernels shrink to a point in time. In this limiting case the activity in one unit is simply a nonlinear function of the instantaneous activity in the other unit (i.e. a polynomial expansion).

$$\boldsymbol{x}_i(t) = h_0 + h_1\,\boldsymbol{x}_j(t) + h_2\,\boldsymbol{x}_j(t)^2 + \ldots . \quad (10)$$

All other cases enter under the rubric of transient codes. These can be similarly decomposed into those that include nonlinear terms (asynchronous) and those that do not (synchronous). The final level of decomposition is of synchronous interactions into oscillatory and non-oscillatory. The former is a special case where the transfer function $l(\omega)$ shrinks down on to one particular frequency. (For completeness it should be noted that oscillatory codes expressed as kernels of any order, that exist predominantly at one frequency may exist; G. Green, personal communication).

In this framework, it can be seen that the most important distinction, that emerges after discounting special or limiting cases, is that between asynchronous and synchronous coupling and the implicit contribution of nonlinear interactions. The presence of coupling among different frequencies, demonstrated in § 4, speaks to the prevalence of strong nonlinearities in the functional integration of neuronal populations. The nature of these nonlinearities is the focus of the rest of this paper.

6. THE NEURAL BASIS OF ASYNCHRONOUS INTERACTIONS

In Friston (1997*b*) it was suggested that, from a neurobiological perspective, the distinction between asynchronous and synchronous interactions could be viewed in the following way. Synchronization emerges from the reciprocal exchange of signals between two populations, where each 'drives' the other, such that the dynamics become entrained and mutually reinforcing. In asynchronous coding the afferents from one population exert a 'modulatory' influence, not on the activity of the second, but on the interactions among them (e.g. effective connectivity or synaptic efficacy) leading to a change in the dynamics intrinsic to the second population. In this model, there is no necessary synchrony between the intrinsic dynamics that ensue and the temporal pattern of modulatory input. An example of this may be the facilitation of high-frequency gamma oscillations among nearby columns in visual cortex by transient modulatory input from the pulvinar. Here the expression of low-frequency transients in the pulvinar will be correlated with the expression of high-frequency transients in visual cortex. To test this hypothesis one would need to demonstrate that asynchronous coupling emerges when extrinsic

Figure 3. (*Cont.*) functions of frequency in the parietal region. The dotted line in the latter corresponds to $p = 0.05$ (uncorrected for the frequencies analysed). This particular example was chosen because it illustrates all three sorts of coupling (synchronous, asynchronous and mixed). From inspection of the cross-correlation matrix it is evident that power in the beta range (20 Hz) in the second time-series is correlated with similar frequency modulation in the first, albeit at a slightly lower frequency. The resulting correlations appear just off the leading diagonal (broken line) on the upper left. The graphs on the right show that the proportion of variance explained by synchronous and asynchronous coupling is roughly the same and, in terms of significance, synchrony supervenes. In contrast the high correlations, between 48 Hz in the second time-series and 26 Hz in the first, are well away from the leading diagonal, with little evidence of correlations within either of these frequencies. The regression analysis confirms that, at this frequency, asynchronous coupling prevails. The variation at about 34 Hz in the parietal region could be explained by several frequencies in the prefrontal region. A formal analysis shows that both synchronous and asynchronous coupling coexist at this frequency (i.e. the middle peak in the graphs on the right).

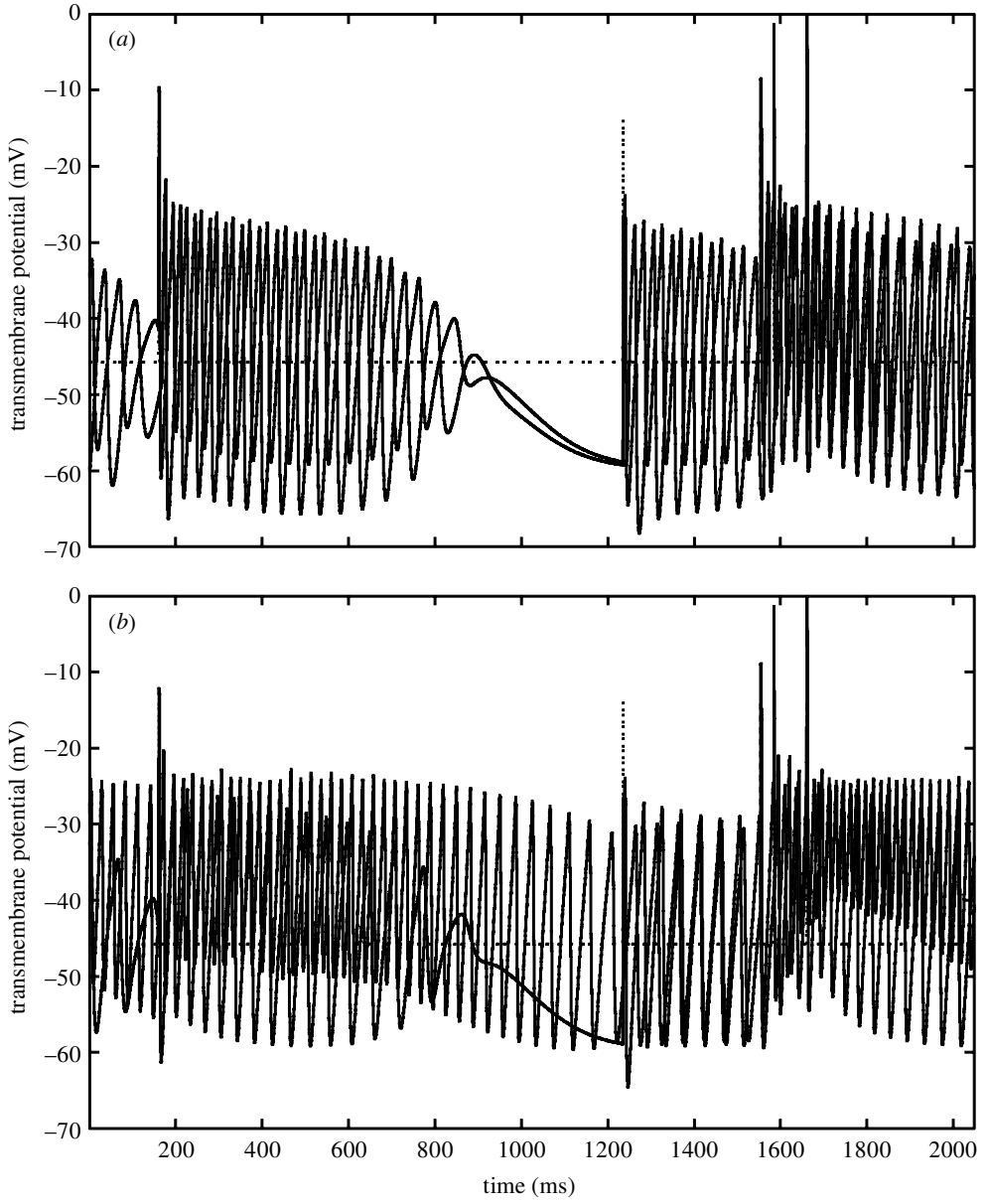

Figure 4. Simulated local field potentials (LFP) of two coupled populations using two different sorts of postsynaptic responses (AMPA and NMDA-like) to extrinsic inputs, to the second population from the first. These data were simulated using the model described in figure 1 and Appendix B. The dotted line shows the depolarization effected by sporadic injections of current into the first population. The key thing to note is that under AMPA-like or driving connections the second population is synchronously entrained by the first (*a*), whereas, when the connections are modulatory or voltage dependent (NMDA), the effects are much more subtle and resemble a frequency modulation (*b*). For the AMPA simulation self-excitatory AMPA connections of 0.15 and 0.06 were used for the first and second populations, respectively, with an AMPA connection between them of 0.06. For the NMDA simulation the self-excitatory connection was increased to 0.14 in the second population and the AMPA connection between the populations was changed to NMDA-like with a strength of 0.6.

connections are changed from driving connections to modulatory connections. Clearly this cannot be done in the real brain. However, we can use computational techniques to create a biologically realistic model of interacting populations and test this hypothesis directly.

(a) *Interactions between simulated populations*

Two populations were simulated using the model described in Appendix B. This model simulated entire neuronal populations in a deterministic or analogue fashion based loosely on known neurophysiological mechanisms. In particular, we modelled three sorts of synapse, fast inhibitory (GABA), fast excitatory (AMPA) and slower voltage-dependent synapses (NMDA). Connections intrinsic to each population used only GABA- and AMPA-like synapses. Simulated glutaminergic extrinsic connections between the two populations used either driving AMPA-like synapses or modulatory NMDA-like synapses. In these and the remaining simulations, transmission delays for extrinsic connections were fixed at 8 ms. By using realistic time constants the characteristic oscillatory dynamics of each population were expressed in the gamma range.

The results of coupling two populations with unidirectional AMPA-like connections are shown in figure 4*a* in terms of the simulated local field potentials (LFP).

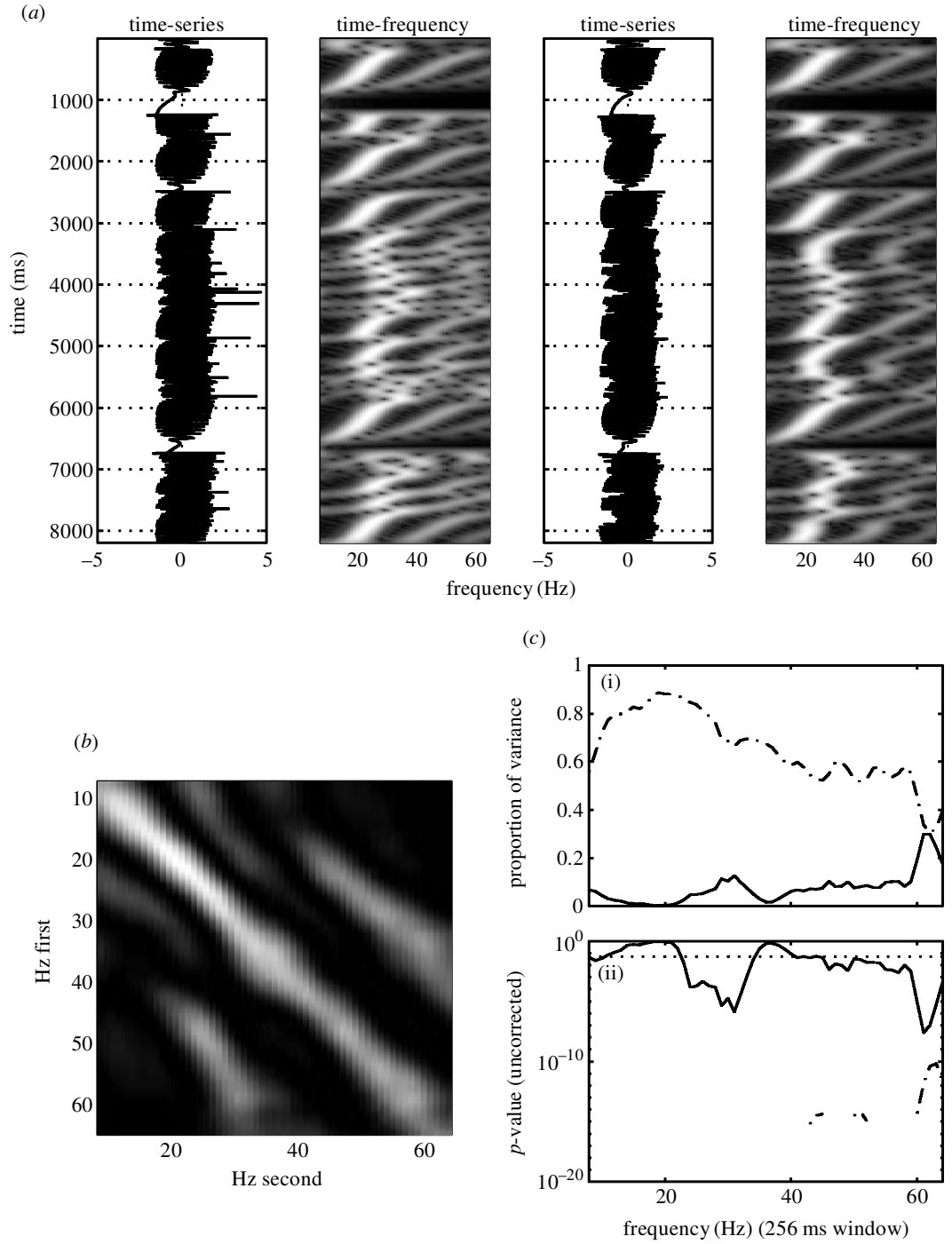

Figure 5. Time-frequency and coupling analyses for the LFPs of the simulations employing AMPA-like connections. The format and underlying analyses of this figure are identical to figure 3. The key thing to note is that the cross-correlations are almost symmetrical, suggesting synchrony and extensive broad-band coherence. Indeed most *p*-values for synchronous (linear) coupling were too small to compute.

Occasional transients in the driving population were simulated by injecting a depolarizing current, of the same magnitude, at random intervals (dotted line). The tight synchronized coupling that ensues is evident. This example highlights the point that near-linear coupling can arise even in the context of loosely coupled, highly nonlinear neuronal oscillators of the sort modelled here. It should be noted that the connection strengths had to be carefully chosen to produce this synchronous entraining. Driving connections do not necessarily engender synchronized dynamics, a point that we will return to later. Contrast these entrained dynamics under driving connections with those that emerge when the connection is modulatory or NMDA-like (figure 4*b*). Here there is no synchrony and, as predicted, fast transients of an oscillatory nature are facilitated by the afferent input from the first population. This is a nice example of asynchronous coupling that is underpinned by nonlinear modulatory interactions between neuronal populations. The nature of the coupling can be characterized more directly using the time-frequency analysis (identical in every detail) applied to the neuromagnetic data of the previous section.

Figure 5 shows the analysis of the AMPA simulation and demonstrates very clear broad-band coherence with

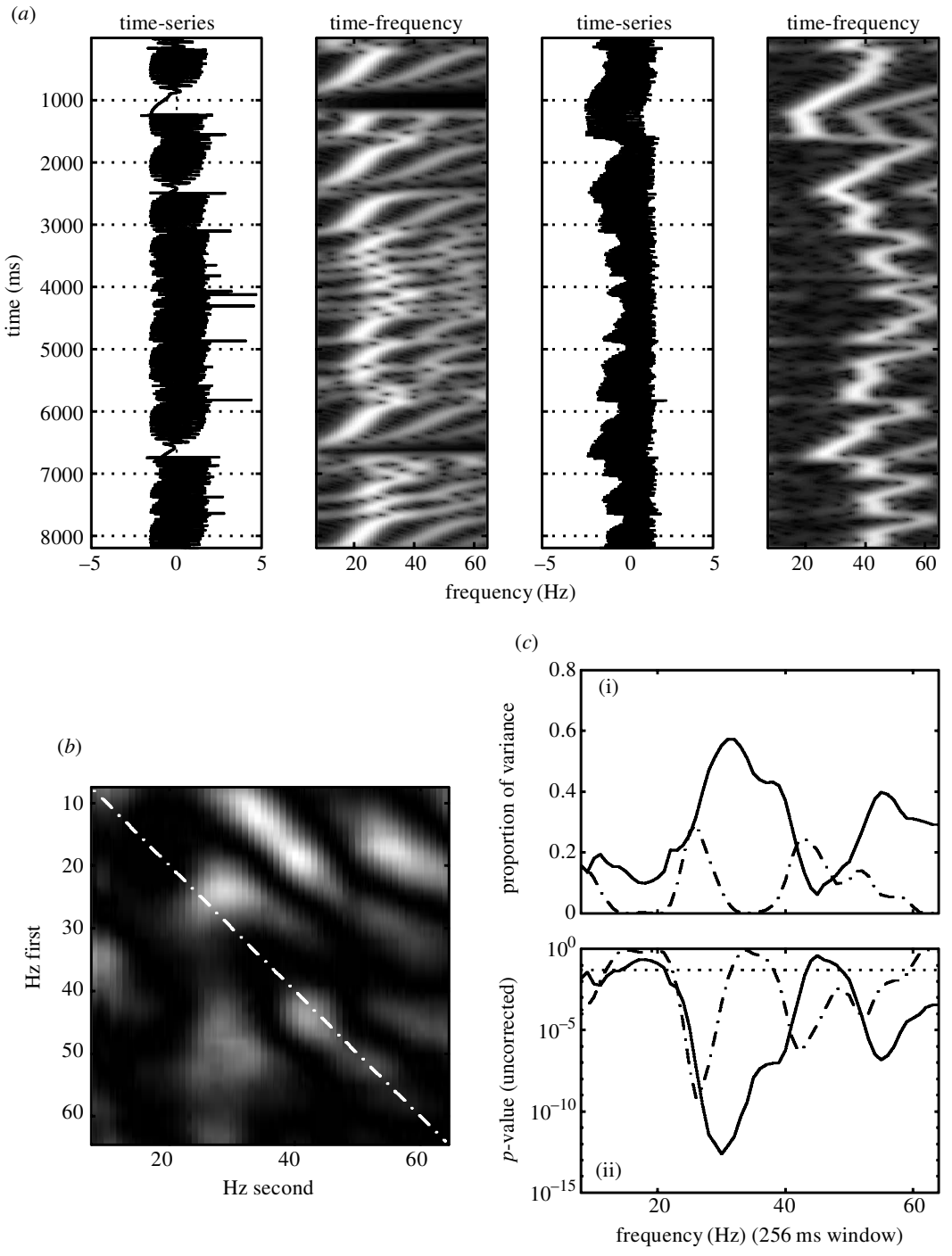

Figure 6. As for figure 5 but now for simulations employing voltage-dependent NMDA-like connections. In contradistinction to figure 5, the coupling here includes some profoundly asynchronous (nonlinear) components involving frequencies in the gamma range implicated in the analyses of real (MEG) data shown in figure 3. In particular, note the asymmetrical cross-correlation matrix and the presence of asynchronous and mixed coupling implicit in the *p*-value plots on the lower right.

most of the cross-correlations among different frequencies lying symmetrically about the leading diagonal. Synchrony accounts for most of the coupling, both in terms of the variance in frequency modulation (figure 5*c*(i)) and in terms of significance (figure 5*c*(ii)). Note that at some frequencies the synchronous coupling was so significant that the *p*-values were too small to compute. These results can now be compared to the equivalent analysis of the NMDA simulation (figure 6).

In contradistinction, the cross-correlation matrix looks much more like that obtained with the MEG data in figure 3. Both in terms of the variance, and inference, asynchronous coupling supervenes at most frequencies but, as in the real data, mixed coupling is also evident. These results can be taken as a heuristic conformation of the hypothesis that modulatory, in this case voltage-dependent, interactions are sufficiently nonlinear to account for the emergence of asynchronous dynamics.

In summary, asynchronous coupling is synonymous with nonlinear coupling. Nonlinear coupling can be framed in terms of the modulation of intrinsic interactions, within a cortical area or neuronal population, by

extrinsic input offered by afferents from other parts of the brain. This mechanism predicts that the modulation of fast (e.g. gamma) activity in one cortical area can be predicted by a (nonlinear function of) activity in another area. This form of coupling is very different from coherence or other metrics of synchronous or linear coupling and concerns the relationship between the first-order dynamics in one area and the second-order dynamics (spectral density) expressed in another. In terms of the above NMDA simulation, transient depolarization in the modulating population causes a short-lived increased input to the second. These afferents impinge on voltage-sensitive NMDA-like synapses with time constants (in the model) of 100 ms. These synapses open and slowly close again, remaining open long after an afferent volley. Because of their voltage-sensitive nature, this input will have no effect on the dynamics intrinsic to the second population unless there is already a substantial degree of depolarization. If there is, then, through self-excitation and inhibition, the concomitant opening of fast excitatory and inhibitory channels will generally increase membrane conductance, decrease the effective membrane time constants and lead to fast oscillatory transients. This is what we observe in figure 4*b*.

(b) *Nonlinear interactions and frequency modulation*

The above considerations suggest that modulatory afferents can mediate a change in the qualitative nature of the intrinsic dynamics though a nonlinear voltage-dependent effect that can be thought of in terms of a frequency modulation of the intrinsic dynamics by this input. This motivates a plausible model of the relationship, between the intrinsic dynamics of one population, characterized by its spectral density $g_2(\omega_0,t)$ and the activity in the first population $\boldsymbol{x}_1(t)$, that is mediated by the modulatory effects of the latter. These effects can be modeled with a Volterra series:

$$g_2(\omega_0,t) = \Omega_0[\boldsymbol{x}_1(t)] + \Omega_1[\boldsymbol{x}_1(t)] + \dots . \tag{11}$$

It is interesting to note that this relationship is a more general version of the statistical model used to test for coupling in the previous section, namely equation (5). This is because a time-frequency analysis itself is a simple form of a Volterra series (see Appendix C). The motivation behind this particular form of coupling between two regions is predicated on the mechanistic insights provided by simulations of the sort presented above.

Given the dynamics of the two populations from the NMDA simulations, we can now estimate the form and significance of the Volterra kernels in equation (11) to characterize more precisely the nature of the nonlinear coupling. In this case, the Volterra kernels were estimated using ordinary least squares. This involves a dimension reduction and taking second-order approximations of the Volterra series expansion (see Appendix E for details). Inferences about the significance of these kernels are made by treating the least-squares estimation as a general linear model in the context of a multiple regression for serially correlated data (Worsley & Friston 1995) as for the time-frequency analyses. The results of this analysis, for each frequency ω_0, are estimates of the Volterra kernels themselves (figure 7*b*) and their significance, i.e. the probability of obtaining the observed data if the kernels were equal to zero (figure 7*a*). By applying the estimated kernels to the activity in the first population one can visualize the expected and actual frequency modulation at w_0 (figure 7*c*).

It is clear that this approach picks up the modulation of specific frequency components in the second population and furthermore the time constants (i.e. duration or temporal extent of the Volterra kernels) are consistent with the time constants of channel opening in the simulation, in this case the long time constants associated with voltage-dependent mechanisms. Knowing the form of the Volterra kernels allows one to characterize the frequency modulation elicited by any specified input. Figure 8 shows the actual frequency structure of the modulated population in the NMDA simulation and that predicted on the basis of the activity in the first population. Figure 8*d* shows the complicated form of frequency modulation that one would expect with a simple Gaussian input over a few hundred milliseconds.

Of course the critical test here is to apply this analysis to the real data of § 5 and see if similar effects can be demonstrated. They can. A good example is presented in figure 9, showing how a slow, nonlinear function (modelled by the Volterra kernels) of prefrontal activity closely predicts the expression of fast (gamma) frequencies in the parietal region. The implied modulatory mechanisms, which may underpin this effect, are entirely consistent with the anatomy, laminar specificity and functional role attributed to prefrontal efferents (Rockland & Pandya 1979; Selemon & Goldman-Rakic 1988).

7. CONCLUSION

This paper has introduced some basic considerations pertaining to neuronal interactions in the brain, framed in terms of neuronal transients and nonlinear coupling. The key points of the arguments developed in this paper follow.

(i) Starting with the premise that the brain can be represented as an ensemble of connected input–state–output systems (e.g. cellular compartments, cells or populations of cells), there exists an equivalent input–output formulation in terms of a Volterra-series expansion of each system's inputs that produces its outputs (where the outputs to one system constitute the inputs to another).

(ii) The existence of this formulation suggests that the history of inputs, or neuronal transients, and the Volterra kernels are a complete and sufficient specification of brain dynamics. This is the primary motivation for framing dynamics in terms of neuronal transients (and using a Volterra formulation for models of effective connectivity).

(iii) The Volterra formulation provides constraints on the form that neuronal interactions and implicit codes must conform to. There are two limiting cases: (i) when the neuronal transient has a very short history; and (ii) when high-order terms disappear. The first case corresponds to instantaneous codes (e.g. rate codes) and the second to synchronous interactions (e.g. synchrony codes).

(iv) High-order terms in the Volterra model of effective connectivity speak explicitly to nonlinear interactions

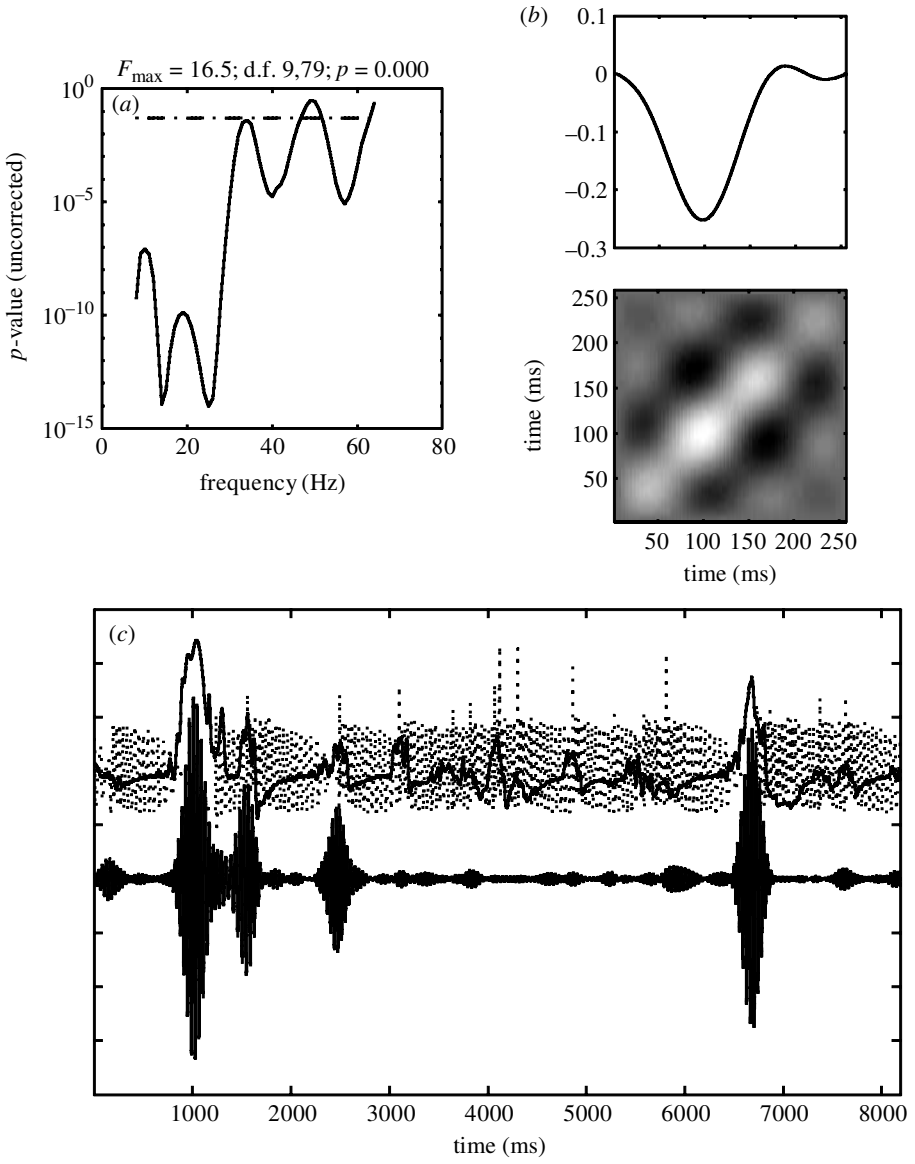

Figure 7. Nonlinear convolution (i.e. Volterra kernel) characterization of the frequency modulation effected by extrinsic modulatory inputs on the fast intrinsic dynamics. The analysis described in Appendix E was applied to the simulated LFPs shown in figure 4 using NMDA-like connections. These are the same time-series analysed in figure 6. In this instance, we have tried to find the Volterra kernels that best model the frequency modulation of the dynamics in the second simulated population given the time-series of the first. (*a*) Plot of the significance of the Kernels as a function of frequency in the modulated (second) population. For the most significant effects (at 25 Hz) the estimated first- and second-order kernels are shown in (*b*). Applying these kernels to the time-series of the first population (dotted lines in (*c*)) one obtains a modulatory variable (solid line) that best predicts the observed frequency modulation (bottom line in (*c*)).

and implicitly to asynchronous coupling. Asynchronous coupling implies coupling among the expression of different frequencies.

(v) Coupling among the expression of different frequencies is easy to demonstrate using neuromagnetic measurements of real brain dynamics. This implies that nonlinear, asynchronous coupling is a prevalent component of functional integration.

(vi) High-order terms in the Volterra model of effective connectivity correspond to modulatory interactions that can be construed as a nonlinear effect of inputs that interact with the dynamics intrinsic to the recipient system. This implies that driving connections may be linear and engender synchronous interactions, whereas modulatory connections, being nonlinear, may cause, and be revealed by, asynchronous coupling.

The latter sections of this paper have shown that asynchronous coupling can account for a significant and substantial component of interactions between brain areas as measured by neuromagnetic signals. Asynchronous coupling of this sort implies nonlinear coupling and both speak to the differential form of neuronal transients that are expressed coincidentally in different brain areas. This observation has been extended by testing the hypothesis that a parsimonious and neurobiologically plausible mechanism of nonlinear coupling employs voltage-dependent synaptic interactions. This led to the prediction that the dynamics in one region can predict the

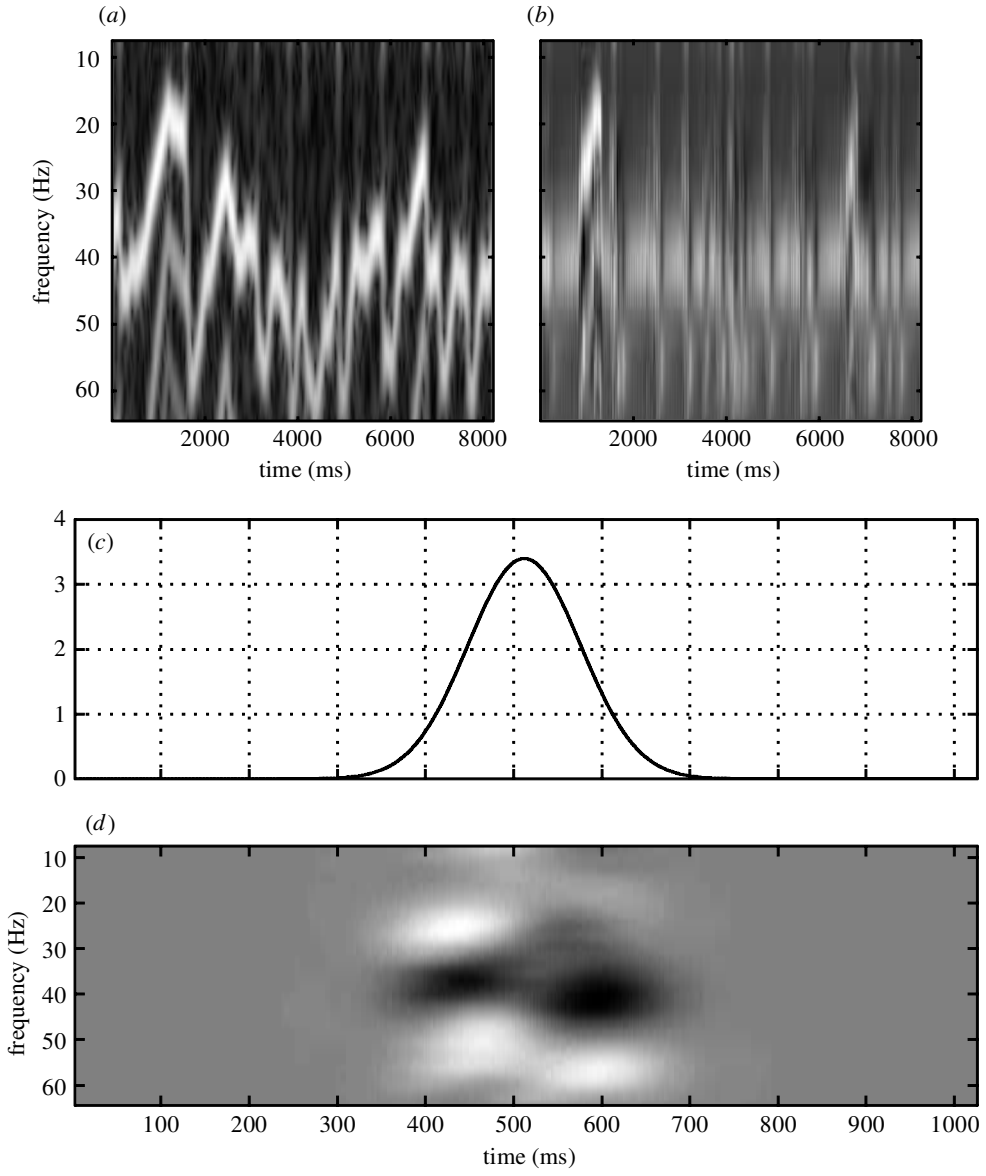

Figure 8. Characterization of the observed and predicted frequency modulation using the estimated Volterra kernels from the analysis summarized in figure 7. (*a*) Actual, and (*b*) predicted, time-frequency profiles of the LFP of the second population. The predicted profile was obtained using the frequency-specific kernel estimates applied to the LFP of the first population. These kernels can be applied to any input, for example the 'synthetic' transient in (*c*). (*d*) The ensuing frequency modulation shows a biphasic suppression around 40 Hz activity with transient increases (with different latencies and time constants) at 25, 50 and 55 Hz.

changes in the frequency structure (a metric of intrinsic dynamics) in another. Not only is this phenomenon easily observed in real data, but in many instances it is extremely significant. In particular, it was shown that a nonlinear function of prefrontal dynamics could account for a significant component of the frequency modulation of parietal dynamics. It should be noted that dynamic changes in spectral density may arise spontaneously from metastable dynamics even in the absence of extrinsic input (see Friston, paper 2, this issue). However, this does not affect the conclusions above because it has been shown that at least some of the (parietal) frequency modulation can be explained by extrinsic (prefrontal) inputs.

The importance of these observations relates both to the mechanisms of functional integration in the brain and to the way that we characterize neuronal interactions. In particular, these results stress the importance of asynchronous interactions that are beyond the scope of synchrony. Although this conclusion is interesting from a theoretical standpoint, in terms of identifying the right metric that is sensitive to the discourse between different brain areas, it also has great practical importance in the sense that many ways of characterizing neuronal time-series are based on synchronization or linear models of neuronal interactions (e.g. cross-correlograms, principal component analysis, singular value decomposition, coherence analyses etc.). An appreciation that nonlinear effects can supervene in terms of their size and significance over linear effects such as coherence (see figure 3) may be important to ensure that we are measuring the right things when trying to characterize functional integration. The theoretical implications are far-reaching because they appeal directly to the

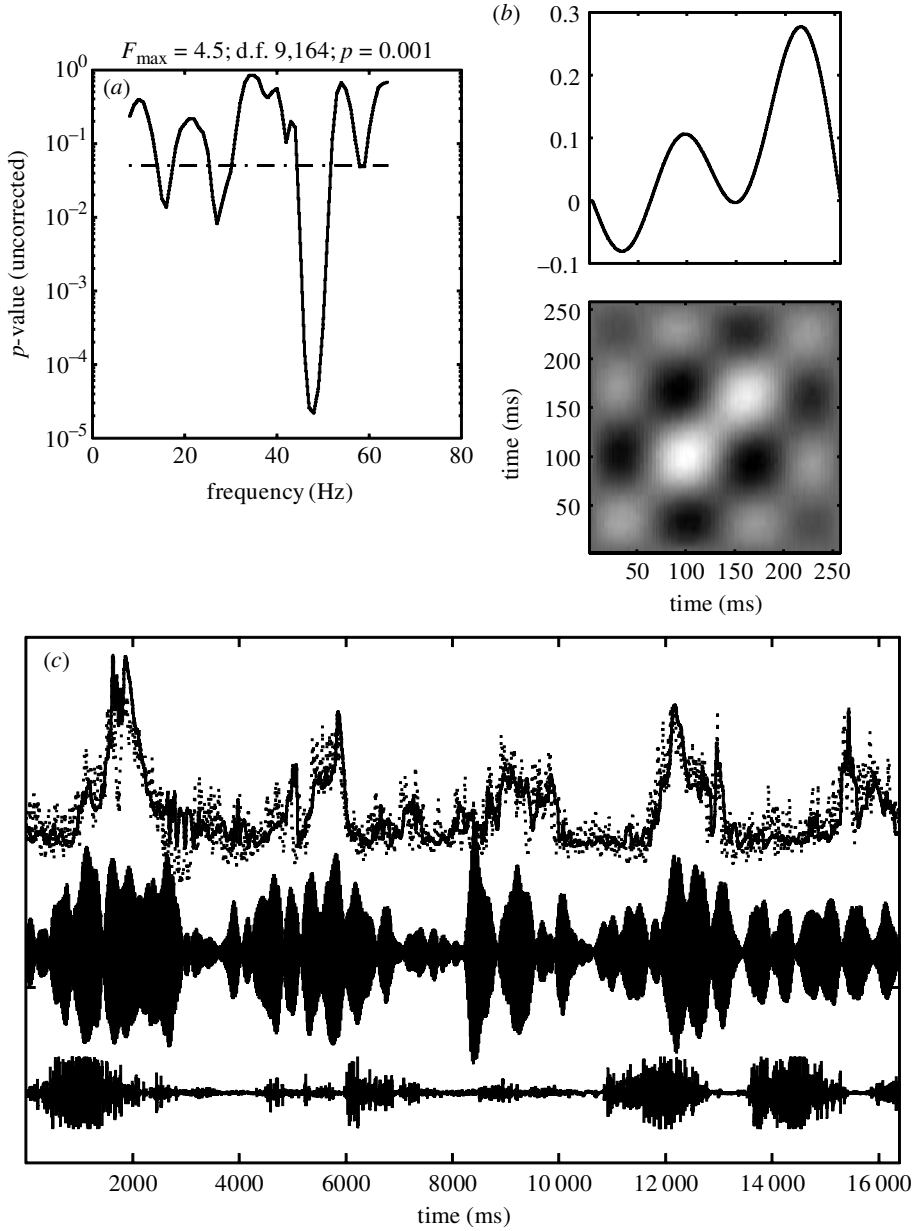

Figure 9. Nonlinear convolution (i.e. Volterra kernel) characterization of the frequency modulation of parietal dynamics that obtains by treating the prefrontal time-series as an extrinsic modulatory input. The format and underlying analyses of this figure are identical to the analysis of simulated LFPs presented in figure 7. (*a*) There is evidence here for profound frequency modulation at 48 Hz that appears to be loosely associated with movement but not completely so. The top traces in (*c*) represent the observed prefrontal time-series before and after nonlinear convolution with the estimated kernels on the upper right. The middle trace is the parietal time-series filtered at 48 Hz and the lower trace is the electromyogram reflecting muscle activity associated with movement. (*b*) Inspection of the first- and second-order kernels suggests that fast gamma (48 Hz) activity in the parietal region is modulated by the transient expression of alpha (8 Hz) activity in the prefrontal region.

context-sensitive nature of neuronal interactions. Modulatory effects are probably central to the mechanisms that mediate attentional changes in receptive field properties and more generally the incorporation of context when constructing a unit's responses to sensory inputs (see Phillips & Singer 1997). One of the reasons that we chose the prefrontal and parietal regions in the MEG analyses was that both these regions are thought to participate in distributed attentional systems (e.g. Posner & Petersen 1990) and indeed our own work with fMRI in human subjects suggests a modulatory role for prefrontal–parietal projections (Büchel & Friston 1997). In paper 2 (Friston, this issue) we relate neuronal transients to dynamical systems theory and complexity to arrive at a view of functional organization in the brain that embraces both linear and nonlinear interactions.

K.J.F. is funded by the Wellcome Trust. I would like to thank Nicki Roffe for help preparing this manuscript and Semir Zeki, Richard Frackowiak, Erik Lumer, Dave Chawla, Christian Büchel, Chris Frith, Cathy Price and the reviewers for their scientific input. Some of this work was first presented orally at the Fifth Dynamical Neuroscience Satellite Symposium of Society for Neuroscience in New Orleans in 1997. I would like to acknowledge the useful feedback and comments from that meeting.

APPENDIX A. NONLINEAR SYSTEM IDENTIFICATION AND VOLTERRA SERIES

Neuronal and neurophysiological dynamics are inherently nonlinear and lend themselves to modelling by nonlinear dynamical systems. However, due to the complexity of biological systems it is difficult to find analytical equations that describe them adequately. An alternative is to adopt a very general model (Wray & Green 1994). A conventional method for representing a nonlinear dynamic system (e.g. a neuron) is an input–state–output model (Manchanda & Green 1998). These models can be classified as either that of a fully nonlinear system (where the inputs can enter nonlinearly) or a linear-analytical system with the form

$$\left.\begin{aligned} \partial \boldsymbol{s}(t)/\partial t &= f_1(\boldsymbol{s}) + f_2(\boldsymbol{s}) \times \boldsymbol{x}_j \\ x_i &= f_3(\boldsymbol{s}) \end{aligned}\right\}, \tag{A1}$$

where $\boldsymbol{s}$ is a vector of states (e.g. the electrochemical state of every cell compartment), f_1, f_2 and f_3 are nonlinear functions, x_j is the input (e.g. afferent activity from unit j) and x_i the output (e.g. efferent activity from unit i). Simple extensions to this description accommodate multiple inputs. Using functional expansions, Fliess *et al.* (1983) have shown how more general nonlinear differential models can be reduced to a linear-analytical form. The Fliess fundamental formula describes the causal relationship between the outputs and the recent history of the inputs. This relationship can be expressed as a Volterra series. Volterra series allow the output to be computed purely on the basis of the past history of the inputs without reference to the state vector. The Volterra series is an extension of the Taylor series representation to cover dynamic systems and has the general form

$$x_i(t) = \Omega_0[\boldsymbol{x}(t)] + \Omega_1[\boldsymbol{x}(t)] + \ldots + \Omega_n[\boldsymbol{x}(t)] + \ldots, \tag{A2}$$

where $\Omega_n[\cdot]$ is the nth order Volterra operator:

$$\Omega_n[\boldsymbol{x}(t)] = \sum_{j_1} \ldots \sum_{j_n} \int_0^\infty \ldots \int_0^\infty h^n_{j_1 \ldots j_n}(u_1, \ldots u_n) \times x_{j_1}(t-u_1) \ldots x_{j_n}(t-u_2)\mathrm{d}u_1 \ldots \mathrm{d}u_n.$$

$\boldsymbol{x}(t) = [x_1(t),\ x_2(t),\ \ldots]$ is the neuronal activity in all the other connected units. $h^n_{j_1 \ldots j_n}$ is the nth order Volterra kernel for units $j_1 \ldots j_n$. It can be shown that these series can represent any analytical time-invariant system. For fully nonlinear systems the above expansion, about the current inputs, can be considered as an approximation that is locally correct. If the inputs enter in a sufficiently nonlinear way the Volterra kernels will themselves change with input (cf. activity-dependent synaptic connections), something that will be developed in paper 2 (Friston, this issue) in terms of instability and complexity. The Volterra series has been described as a 'polynomial series with memory' and is more generally thought of as a high order or 'nonlinear convolution' of the inputs to provide an output. See Bendat (1990) for a fuller discussion.

From the present perspective the Volterra kernels are essential in characterizing the effective connectivity or influences that one neuronal system exerts over another because they represent the causal characteristics of the system in question. Volterra series provide central links to conventional methods of describing input–output behaviour such as the time-frequency analyses used in this paper. See Manchanda & Green (1998) for a fuller discussion of Volterra series in the context of neural networks.

APPENDIX B. THE NEURONAL SIMULATIONS

The simulations used a biologically plausible model of the dynamics of either one or several neuronal populations. The model was of a deterministic or 'analogue' sort (cf. Erb & Aertsen 1992) whose variables pertain to the collective, probabilistic behaviour of subpopulations of neurons. The variables in this model represent the expected transmembrane potentials over units in each subpopulation and the probability of various events underlying changes in that mean. Each population was modelled in terms of an excitatory and inhibitory subpopulation (see Jefferys *et al.* (1996) for an overview of these architectures) whose expected (i.e. mean) transmembrane potentials V_1 and V_2, were governed by the following differential equations.

$$\left.\begin{aligned} C \times \partial V_1/\partial t = {} & g_{1e} \times (V_1 - V_e) + g_{1v} \times (V_1 - V_e) \\ & + g_{1i} \times (V_1 - V_i) + g_l \times (V_1 - V_l) \\ C \times \partial V_2/\partial t = {} & g_{2e} \times (V_2 - V_e) + g_{2i} \times (V_2 - V_i) \\ & + g_l \times (V_2 - V_l) \end{aligned}\right\}, \tag{B1}$$

where C is the membrane capacitance (taken to be 1 μF) and g_{1e}, g_{1v} and g_{1i} are the expected proportion of excitatory (AMPA-like and NMDA-like) and inhibitory (GABA-like) channels open at any one time over all excitatory units in the population. Similarly for g_{2e} and g_{2i} in the inhibitory population. g_l is a leakage conductance. V_e, V_i and V_l are the equilibrium potentials for the various channels and resting conditions respectively. Channel configurations were modelled using a two-compartment, first-order model, in which any channel could be open or closed. For any given channel type k:

$$\partial g_k/\partial t = (1 - g_k) \times P_k - g_k/\tau_k, \tag{B2}$$

where P_k is the probability of channel opening and τ_k is the time constant for that channel type (to model classical neuromodulatory effects τ_{AMPA} was replaced by $\tau_{\mathrm{AMPA}}/(1-M)$, where M was 0.8 unless otherwise specified). P_k was determined by the probability of channel opening in response to one or more presynaptic inputs. This is simply one minus the probability it would not open:

$$P_k = 1 - \Pi(1 - P_{jk} \times D_j(t - u_{jk})), \tag{B3}$$

where P_{jk} is the conditional probability that a discharge event in subpopulation j would cause the channel to open and $D_j(t)$ is probability of such an event. P_{jk} represents the mean synaptic efficacy for afferents from subpopulation j and can be thought of as a connection strength. u_{jk} is the associated transmission delay. The final expression closes the loop and relates the discharge probability to the expected transmembrane potential in equation (B1).

$$D_j = \sigma_\eta\{V_j - V_r\}, \tag{B4}$$

where $\sigma_\eta\{\cdot\}$ is a sigmoid function $\sigma_\eta\{x\} = 1/(1 + \exp(-x/\eta))$. V_r can be likened to the reversal potential. For NMDA-like channels the channel opening probability was voltage dependent:

$$P_{jv} = P^*_{jv} \times \sigma_8\{V_j - V_v\}, \tag{B5}$$

where P^*_{jv} is the conditional probability of opening given a presynaptic input from subpopulation j when the postsynaptic membrane is fully depolarized.

Each simulation comprised a 4096 iteration burn in followed by 4096 iterations to see the dynamics that ensued. Each iteration corresponds to 1 ms. V_1 was taken as an index of the simulated local field potential. Voltage-dependent connections were only used between the excitatory subpopulations of distinct populations. The intrinsic excitatory–inhibitory, inhibitory–excitatory and inhibitory–inhibitory connections P_{jk} where all fixed at 0.6, 0.4 and 0.2, respectively. Intrinsic excitatory–excitatory connections were manipulated to control the degree of spontaneous oscillation in the absence of other inputs. Extrinsic connections were excitatory–excitatory connections, employing either AMPA or NMDA synapses and were specified depending on the architecture of the system being modelled. All extrinsic transmission delays were 8 ms. Remaining model parameters were

$V_e = 0$ mV, $g_l = 1/60$ ms,
$V_i = -100$ mV, $\tau_e = 6$ ms,
$V_l = -60$ mV, $\tau_v = 100$ ms,
$V_r = -20$ mV, $\tau_i = 10$ ms,
$V_v = -10$ mV.

Note that sodium or potassium channels are not explicitly modelled here. The nonlinear dependency of discharge probability on membrane potential is implicit in equation (B4).

APPENDIX C. TIME-FREQUENCY ANALYSIS AND WAVELET TRANSFORMATIONS

The time-frequency analyses in this paper employed standard windowed Fourier transform techniques with a 256 ms Hanning window. MEG frequencies analysed ranged from 8 to 64 Hz. Using a continuous time formulation, for any given time-series $x(t)$, the [time-dependent] spectral density $g(\omega,t)$ can be estimated as

$$\left.\begin{aligned} g(\omega,t) &= |s(\omega,t)|^2, \\ \text{where } s(\omega,t) &= x(t) \otimes \{\omega(t) \times \exp(-j\omega t)\} \\ &= \int w(u) \times \exp(-j\omega t) \times x(t-u)\,\mathrm{d}u \end{aligned}\right\}, \tag{C1}$$

ω is 2π times the frequency in question and $j = \sqrt{-1}$. Here $\otimes$ denotes convolution and $w(t)$ is some suitable windowing function of length l. A Hanning function (a bell-shaped function) $w(t) = (1 - \cos(2\pi t/(l+1)))/2]$ with $l = 256$ iterations or milliseconds was used. This windowed Fourier transform approach to time-frequency analysis is interesting, in the present context, because it is a special case equation (A2), i.e. $g(\omega,t)$ obtains from applying a second-order Volterra operator to $x(t)$, where the corresponding kernel is

$$h(u_1,u_2) = w(u_1) \times w(u_2) \times \exp(-j\omega u_1) \times \exp(-j\omega u_2). \tag{C2}$$

(a) *Relationship to wavelet analyses*

This simple time-frequency analysis is closely related to wavelet analyses, particularly those using Morlet wavelets. In wavelet analyses the energy at a particular scale is given by

$$g(\omega,t) = |x(t) \otimes w(t,\omega)|^2, \tag{C3}$$

where $w(t,\omega)$ represents a family of wavelets. For example the complex Morlet wavelet is

$$w(t,\omega) = A \times \exp(-u^2/2\sigma^2) \times \exp(-j\omega t), \tag{C4}$$

where the wavelet family is defined by a constant κ (typically about 6) such that $\omega \times \sigma = \kappa$. Intuitively an analysis using Morlet wavelets is the same as a time-frequency analysis in equation (C1), but where the windowing function becomes narrower with increasing frequency (i.e. $w(t) = A \times \exp(-u^2/2\sigma^2)$). The relative advantages of Morlet transforms, in terms of time-frequency resolution, are discussed in Tallon-Baudry *et al.* (1997).

APPENDIX D. DETECTING NONLINEAR INTERACTIONS IN TERMS OF ASYNCHRONOUS COUPLING

Nonlinear coupling between two dynamical systems is often difficult to detect. There are two basic approaches, which reflect the search for nonlinearities in general (see Muller-Gerking (1996) for an excellent discussion). The first involves reconstructing some suitable state space and characterizing the dynamics using the ensuing trajectories (e.g. Schiff *et al.* 1996). The second is to test for the presence of nonlinearities using standard inferential techniques to look at mutual predictability after discounting linear coupling. The time-frequency analysis presented in this paper is an example of the latter and is based on the observation that significant correlations among different frequencies (after removing those that can be explained by the same frequency) can only be mediated by nonlinear coupling.

The time-dependent spectral density at ω of the time-series $x_1(t)$ and $x_2(t)$ are estimated as above, with $g_1(\omega,t) = |s_1(\omega,t)|^2$. Similarly for $x_2(t)$ (see Appendix C). To test for synchronous (linear) and asynchronous (nonlinear) coupling at a reference frequency ω_0, $g_1(\omega_0,t)$ is designated as the response variable, with explanatory variables $g_2(\omega_0,t)$ and $g_2(\omega,t)^*$ in a standard multilinear regression, extended to account for serially correlated variables (Worsley & Friston 1995). $g_2(\omega,t)^*$ represents modulation at all frequencies after the modulation at ω_0 has been covaried out.

$$g_2(\omega,t)^* = g_2(\omega,t) - g_2(\omega_0,t) \times \mathrm{pinv}\{g_2(\omega_0,t)\} \times g_2(\omega,t), \tag{D1}$$

pinv$\{\cdot\}$ is the pseudo-inverse. This renders the synchronous $g_2(\omega_0,t)$ and asynchronous $g_2(\omega,t)^*$ explanatory variables orthogonal. To test for asynchronous coupling the explanatory variables comprise $g_2(\omega,t)^*$ while $g_2(\omega_0,t)$ is treated as a confound. To test for synchronous coupling

$g_2(\omega_0,t)$ is treated as the regressor of interest and $g_2(\omega,t)^*$ the confounds. The ensuing multiple regressions give the appropriate sums of squares (synchronous and asynchronous), F-values and associated p-values as a function of ω_0. In practice, the spectral densities are subject to a quadratic root transform as a preprocessing step, to ensure the residuals are approximately multivariate Gaussian (Friston 1997*a*).

The inferences above have to be corrected for serial correlations in the residuals, necessarily introduced during the convolution implicit in the time-frequency analysis (Worsley & Friston 1995). Under the null hypothesis these can be modelled by convolution with $w(t)^2$, the square of the window employed. This would be exactly right if we used the spectral densities themselves and is approximately right using the quadratic root transform. This approximation leads to a test that is no longer exact but is still valid (i.e. slightly conservative).

APPENDIX E. ESTIMATING VOLTERRA KERNELS

The problem of estimating Volterra kernels is not a trivial one. We have adopted a standard least-squares approach. This has the advantage of providing for statistical inference using the general linear model. To do this one must first linearize the problem. Consider the second-order approximation of equation (A2) with finite 'memory' T, where the input is a single time-series $x(t)$ and the output is denoted by $y(t)$.

$$y(t) \approx \Omega_0[x(t)] + \Omega_1[x(t)] + \Omega_2[x(t)]. \tag{E1}$$

The second step in making the estimates of h^0, h^1 and h^2 more tractable is to expand the kernels in terms of a small number P of temporal basis functions $b_i(u)$. This allows us to estimate the coefficients of this expansion using standard least squares:

$$\left.\begin{aligned} &\text{let } h^0 = g^0 \\ &h^1(u_1) = \sum_{i=1}^{p} g_i^1 b_i(u_1) \\ &h^2(u_1,u_2) = \sum_{i=1}^{p}\sum_{j=1}^{p} g_{ij}^2 b_i(u_1) \times b_j(u_2) \end{aligned}\right\}. \tag{E2}$$

Now define a new set of explanatory variables $z_i(t)$ that represent the original time-series $x(t)$ convolved with the ith basis function. Substituting these expressions into equation (E1) and including an explicit error term ε gives

$$y(t) = g^0 + \sum_{i=1}^{p} g_i^1 z_i(t) + \sum_{i=1}^{p}\sum_{j=1}^{p} g_{ij}^2 z_i(t) \times z_j(t) + \varepsilon. \tag{E2}$$

This is simply a general linear model with response variable $y(t)$, the observed time-series, and explanatory variables *1*, $z_i(t)$ and $z_i(t) \times z_j(t)$. These explanatory variables (convolved time-series of the original explanatory variables) constitute the columns of the design matrix. The unknown parameters are g^0, g^1 and g^2 from which the kernel coefficients h^0, h^1 and h^2 are derived, using equation (E2). Having reformulated the problem in this way we can now use standard analysis procedures developed for serially correlated time-series that employ the general linear model (Worsley & Friston 1995). These procedures provide parameter estimates (i.e. estimates of the basis function coefficients and implicitly the kernels themselves) and statistical inferences about the significance of the kernels, or more precisely the effect that they mediate. The issues of serial correlations in the residuals are identical to those described in Appendix D.

REFERENCES

Abeles, M., Prut, Y., Bergman, H. & Vaadia, E. 1994 Synchronisation in neuronal transmission and its importance for information processing. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 39–50. Berlin: Springer.

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidmann, E., Tishby, N. & Vaadia, E. 1995 Cortical activity flips among quasi-stationary states. *Proc. Natl Acad. Sci. USA* **92**, 8616–8620.

Aertsen, A. & Preißl, H. 1991 Dynamics of activity and connectivity in physiological neuronal networks. In *Non linear dynamics and neuronal networks* (ed. H. G. Schuster), pp. 281–302. New York: VCH Publishers.

Aertsen, A., Erb, M. & Palm, G. 1994 Dynamics of functional coupling in the cerebral cortex: an attempt at a model-based interpretation. *Physica* D **75**, 103–128.

Bair, W., Koch, C., Newsome, W. & Britten, K. 1994 Relating temporal properties of spike trains from area MT neurons to the behaviour of the monkey. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 221–250. Berlin: Springer.

Bendat, J. S. 1990 *Nonlinear system analysis and identification from random data*. New York: Wiley.

Bressler, S. L., Coppola, R. & Nakamura, R. 1993 Episodic multi-regional cortical coherence at multiple frequencies during visual task performance. *Nature* **366**, 153–156.

Büchel, C. & Friston, K. J. 1997 Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebr. Cortex* **7**, 768–778.

Burgess, N., Recce, M. & O'Keefe, J. 1994 A model of hippocampal function. *Neural Network* **7**, 1065–1081.

deCharms, R. C. & Merzenich, M. M. 1996 Primary cortical representation of sounds by the coordination of action potential timing. *Nature* **381**, 610–613.

de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberie, R. & Bialek, W. 1997 Reproducibility and variability in neural spike trains. *Science* **275**, 1085–1088.

Desmedt, J. E. & Tomberg, C. 1994 Transient phase-locking of 40 Hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. *Neurosci. Lett.* **168**, 126–129.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. 1988 Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analysis in the cat. *Biol. Cybern.* **60**, 121–130.

Engel, A. K., Konig, P. & Singer, W. 1991 Direct physiological evidence for scene segmentation by temporal coding. *Proc. Natl Acad. Sci. USA* **88**, 9136–9140.

Erb, M. & Aertsen, A. 1992 Dynamics of activity in biology-oriented neural network models: stability analysis at low firing rates. In *Information processing in the cortex. Experiments and theory* (ed. A. Aertsen & V. Braitenberg), pp. 201–223. Berlin: Springer.

Fliess, M., Lamnabhi, M. & Lamnabhi-Lagarrigue, F. 1983 An algebraic approach to nonlinear functional expansions. *IEEE Trans. Circuits Syst.* **30**, 554–570.

Freeman, W. & Barrie, J. 1994 Chaotic oscillations and the genesis of meaning in cerebral cortex. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 13–18. Berlin: Springer.

Fries, P., Roelfsema, P. R., Engel, A., Konig, P. & Singer, W. 1997 Synchronization of oscillatory responses in visual cortex correlates with perception in inter-ocular rivalry. *Proc. Natl Acad. Sci. USA* **94**, 12 699–12 704.

Friston, K. J. 1995*a* Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* **2**, 56–78.

Friston, K. J. 1995*b* Neuronal transients. *Proc. R. Soc. Lond.* B **261**, 401–405.

Friston, K. J. 1997*a* Another neural code? *NeuroImage* **5**, 213–220.

Friston, K. J. 1997*b* Transients metastability and neuronal dynamics. *NeuroImage* **5**, 164–171.

Friston, K. J., Stephan, K. M., Heather, J. D., Frith, C. D., Ioannides, A. A., Liu, L. C., Rugg, M. D., Vieth, J., Keber, H., Hunter, K. & Frackowiak, R. S. J. 1996 A multivariate analysis of evoked responses in EEG and MEG data. *NeuroImage* **3**, 167–174.

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

Gerstein, G. L., Bedenbaugh, P. & Aertsen, A. M. H. J. 1989 Neuronal assemblies *IEEE Trans. Boimed. Engng* **36**, 4–14.

Gray, C. M. & Singer, W. 1989 Stimulus specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86**, 1698–1702.

Haken, H. 1983 *Synergistics: an introduction*, 3rd edn. Berlin: Springer.

Hebb, D. O. 1949 *The organization of behaviour.* New York: Wiley.

Ioannides, A. A., Hasson, R. & Miseldine, G. J. 1990 Model-dependent noise elimination and distributed source solutions for the biomagnetic inverse problem. *SPIE* **1351**, 471.

Jefferys, J. G. R., Traub, R. D. & Whittington, M. A. 1996 Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208.

Jirsa, V. K., Friedrich, R. & Haken, H. 1995 Reconstruction of the spatio-temporal dynamics of a human magnetoencephalogram. *Physica* D **89**, 100–122.

Jurgens, E., Rosler, F., Hennighausen, E. & Heil, M. 1995 Stimulus induced gamma oscillations: harmonics of alpha activity? *NeuroReport* **6**, 813–816.

Kelso, J. A. S. 1995 *Dynamic patterns: the self-organisation of brain and behaviour.* Cambridge, MA: MIT Press.

Llinas, R., Ribary, U., Joliot, M. & Wang, X.-J. 1994 Content and context in temporal thalamocortical binding. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 251–272. Berlin: Springer.

Lumer, E. D., Edelman, G. M. & Tononi, G. 1997 Neural dynamics in a model of the thalamocortical system. II. The role of neural synchrony tested through perturbations of spike timing. *Cerebr. Cortex* **7**, 228–236.

Manchanda, S. & Green, G. G. R. 1999 The input–output behaviour of dynamic neural networks. *Physica* D. (Submitted.)

Milner, P. M. 1974 A model for visual shape recognition. *Psychol. Rev.* **81**, 521–535.

Muller-Gerking, J., Martinerie, J., Neuenschwander, S., Pezard, L., Rebault, B. & Varela, F. J. 1996 Detecting non-linearities in neuro-electric signals: a study of synchronous local field potentials. *Physica* D **94**, 65–81.

Optican, L. & Richmond, B. J. 1987 Temporal coding of two-dimensional patterns by single units in primate inferior cortex. II. Information theoretic analysis. *J. Neurophysiol.* **57**, 132–146.

Phillips, W. A. & Singer, W. 1997 In search of common foundations for cortical computation. *Behav. Brain Sci.* **4**, 657–683.

Posner, M. I. & Petersen, S. E. 1990 The attention system of the human brain. *A. Rev. Neurosci.* **13**, 25–42.

Riehle, A., Grun, S., Diesmann, M. & Aertsen, A. 1997 Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950–1953.

Rockland, K. S. & Pandya, D. N. 1979 Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res.* **179**, 3–20.

Roelfsema, P. R., Konig, P., Engel, A. K., Sireteanu, R. & Singer, W. 1994 Reduced synchronization in the visual cortex of cats with strabismic amblyopia. *Eur. J. Neurosci.* **6**, 1645–1655.

Schiff, S. J., So, P., Chang, T., Burke, R. E. & Sauer, T. 1996 Detecting dynamical interdependence and generalized synchrony though mutual prediction in a neuronal ensemble. *Phys. Rev.* E **54**, 6708–6724.

Selemon, L. D. & Goldman-Rakic, P. S. 1988 Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *J. Neurosci.* **8**, 4049–4068.

Shadlen, M. N. & Newsome, W. T. 1995 Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**, 569–579.

Singer, W. 1994 Time as coding space in neocortical processing: a hypothesis. In *Temporal coding in the brain* (ed. G. Buzsaki, R. Llinas, W. Singer, A. Berthoz & T. Christen), pp. 51–80. Berlin: Springer.

Sporns, O., Gally, J. A., Reeke, G. N. & Edelman, G. M. 1989 Reentrant signalling among simulated neuronal groups leads to coherence in their oscillatory activity *Proc. Natl Acad. Sci. USA* **86**, 7265–7269.

Stevens, C. F. 1994 What form should a cortical theory take? In *Large-scale neuronal theories of the brain* (ed. C. Kock & J. L. Davies), pp. 239–255. Cambridge, MA: MIT Press.

Tononi, G., Sporns, O. & Edelman, G. M. 1992 Re-entry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebr. Cortex* **2**, 310–335.

Tovee, M. J., Rolls, E. T., Treves, A. & Bellis, R. P. 1993 Information encoding and the response of single neurons in the primate temporal visual cortex. *J. Neurophysiol.* **70**, 640–654.

Treue, S. & Maunsell, H. R. 1996 Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. & Aertsen, A. 1995 Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**, 515–518.

Von der Malsburg, C. 1981 The correlation theory of the brain. Internal report. Max-Planck-Institute for Biophysical Chemistry, Gottingen, Germany.

Von der Malsburg, C. 1985 Nervous structures with dynamical links. *Ber Bunsenges. Phys. Chem.* **89**, 703–710.

Worsley, K. J. & Friston, K. J. 1995 Analysis of fMRI time-series revisited—again. *NeuroImage* **2**, 173–181.

Wray, J. & Green, G. G. R. 1994 Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neuronal network. *Biol. Cybern.* **71**, 187–195.