THE ROYAL SOCIETY

The labile brain. III. Transients and spatio-temporal receptive fields

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 paper we consider an approach to neuronal transients that is predicated on the information they contain. This perspective is provided by information theory, in particular the principle of maximum information transfer. It is illustrated here in application to visually evoked neuronal transients. The receptive fields that ensue concur with those observed in the real brain, predicting, almost exactly, functional segregation of the sort seen in the visual system. This information theoretical perspective can be reconciled with a selectionist stance by noting that a high mutual information among neuronal systems and the environment has, itself, adaptive value and will be subject to selective pressure, at any level one cares to consider.

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

1. INTRODUCTION

This paper is concerned with the information conveyed by a neuronal transient and the implications for the temporal structure of neuronal processing (e.g. perceptual synthesis) and unit responses (e.g. spatio-temporal receptive fields). In § 2, we consider the constraints on, and implications of, distributing information over time in neuronal transients, while § 3 demonstrates the predictive validity of the transient hypothesis by showing how functional segregation in extrastriate cortex (the spatio-temporal receptive fields of units in secondary visual area V2) emerges spontaneously when the principle of maximum information transfer is applied to neuronal transients.

2. INFORMATION AND NEURONAL TRANSIENTS

If the diversity of transients depends on nonlinear or asynchronous coupling, it follows that this coupling is fundamental because it is the genesis of information that is embodied in the dynamics of integrated neuronal populations. This suggests there must be a proper balance between synchronous and asynchronous coupling. The information theoretical analysis presented in this section leads to an obvious but interesting view of neuronal transients that points to some characteristic time-scales for neuronal processing that depend on the coupling among neuronal populations.

(a) *An uncertainty principle for the brain*

The uncertainty principle states that there is an inherent trade-off between the certainty with which one can specify the energy (or momentum) of a small particle and the time (position) at which it was observed. This follows from the fact that the energy (or momentum) is related to the frequency of the 'wave' describing the particle. Clearly one cannot know both the exact frequency and the exact point in time that a frequency is expressed. In a similar vein, one cannot know the exact form of a neuronal transient and the exact time that it occurred (it takes a finite amount of time for its form to become apparent). This trade-off can be expressed more formally in terms of the entropy. The entropy is the average information about which neuronal transient has occurred. Under Gaussian assumptions (Jones 1979) for a time-window of τ observations (i.e. a temporal uncertainty of τ) the entropy of a transient sampled at these times is

$$H\{x(t)\} = \log(2\pi e^{\tau}\det\{R\})/2, \quad (1)$$

where R is the $(\tau \times \tau)$ autocorrelation matrix of the neuronal process $x(t)$. As the temporal uncertainty increases, the average information obtained by actually observing the transient increases. In short, we can either know which particular transient is being currently expressed or when it is expressed but not both at the same time. Clearly the relationship implied by equation (1) will be subject to the constraints of the neuronal system in question, imposed by the form of R. These constraints are in turn determined by the Volterra kernels, or effective connectivity, that mediate the dynamics. Figure 1 shows the average information obtained by using estimates of R based on the neuromagnetic data used in Friston (paper 1, this issue) sampled every 4 ms. The critical thing to note is that there is an almost linear relationship between the duration of a transient and the average information one obtains on knowing its form. This is an important point and simply states that the information in the history of some neuronal dynamics increases in proportion to the depth of that history. A time-frequency analysis of a single

Received 2 September 1998 *Accepted* 8 June 1999

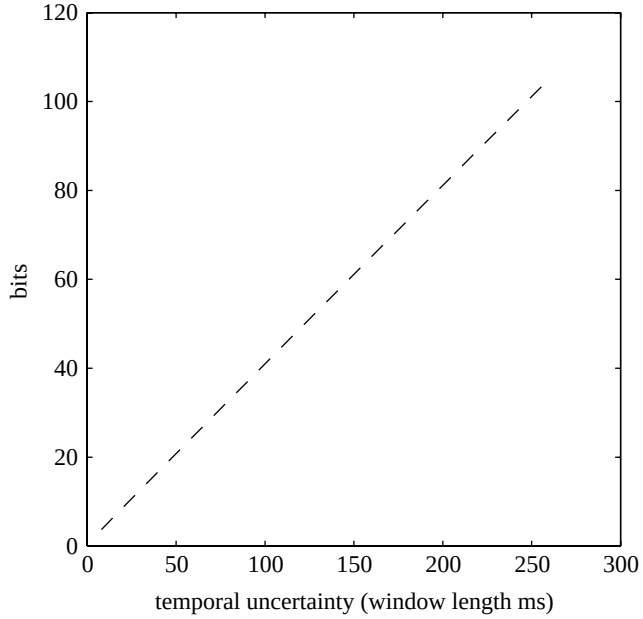

Figure 1. Entropy of a single MEG time-series expressed as a function of window length. The entropies were based on the appropriate correlation matrices estimated from a 2^{14} ms epoch of MEG data (that from the prefrontal region shown in figure 3 of Friston, paper 1, this issue) sampled every 4 ms.

time-series emulates a multiplexing over frequencies. The longer the transient the greater the number of (low) frequencies that can be estimated.

(b) *Short or long transients?*

If there is more information available in a long transient, relative to a short one, is this a sufficient motivation for the brain to use long transients? The answer to this question lies in the nature of the 'motivation' and how the brain can 'use a transient'. The 'motivation' reduces to selective pressure at a neuronal or evolutionary time-scale and the 'use' of a transient is operationally defined by the Volterra kernels that mediate between a transient input to a neuronal population and the ensuing response. The useful duration of a transient is determined by the temporal extent of the Volterra kernels or effective connectivity. If these kernels are temporally protracted (i.e. can sample inputs from the distant past) then the information inherent in longer transients will available for shaping the population's response. This in turn will lead to richer, more diverse responses that are more sensitive to the temporal context in which they occur. Is this generally adaptive? From an evolutionary point of view, not necessarily.

Consider small adaptive neuronal systems such as the nervous systems of insects. Assuming that the small transmission delays, implied by the physical size of insect brains, renders the temporal extent of the Volterra kernels comparatively small, then the information that can be sampled from any transient will be limited, as will the corresponding repertoire of context-sensitive neuronal responses. Consider now larger animals, such as man, where the temporal extent of the kernels may exceed, say, 500 ms. Here the responses of any neuronal population will be predicated on a much more information-rich history of inputs and therefore have the potential to be more adaptive, but at a price. The price relates to the speed at which inputs are transformed into outputs. For the insect, a new transient is available, say, every 10 ms or so, whereas for systems with kernels that cover 500 ms, a transient is only refreshed a couple of times a second. From the point of view of the insect, the man will respond in an incomprehensively complex way but intolerably slowly. Conversely from the man's point of view the world (of insects) will rush past very fast but in a simple and predictable fashion. Which is most adaptive? Clearly both are adaptive. The more important point here is that there is a trade-off between the complexity and context sensitivity of neuronal responses and the characteristic time constants of these responses.

Figure 2. One of the eight natural scenes used to identify the spatio-temporal receptive fields according to the principles described in the main text.

The anecdotal example above used small and large brains but there are likely to be many analogous examples within one nervous system (e.g. reflexes versus cognitive operations). These arguments speak to the notion that any neuronal system can be characterized in terms of the temporal extent of its underlying Volterra kernels. Long kernels will engender more complex dynamics and will extract more information from afferent transients. The price paid for this is that the neuronal moment is suspended in time, rendering any particular instant inaccessible. This inaccessibility is due to the fact that the neuronal response to any instantaneous event is inevitably conflated with the history that precedes it. In other words, neuronal systems with long Volterra kernels can never 'represent' instants in time because the neuronal representation of these instants always reflects the context in which they occur. In short, for complex systems the 'moment' that is represented is necessarily inflated to preclude a representation of the sensorium that retains an instantaneous temporal acuity.

Although it is not easy to relate these arguments to perception, they do suggest that what we perceive may be temporally divorced from what we sample with our sensory receptors and that perceptual synthesis may necessarily involve a loss of temporal precision. The compelling experiments of Moutoussis & Zeki (1997) on perceptual asynchrony in vision speak exactly to this temporal dislocation, where different attributes of the

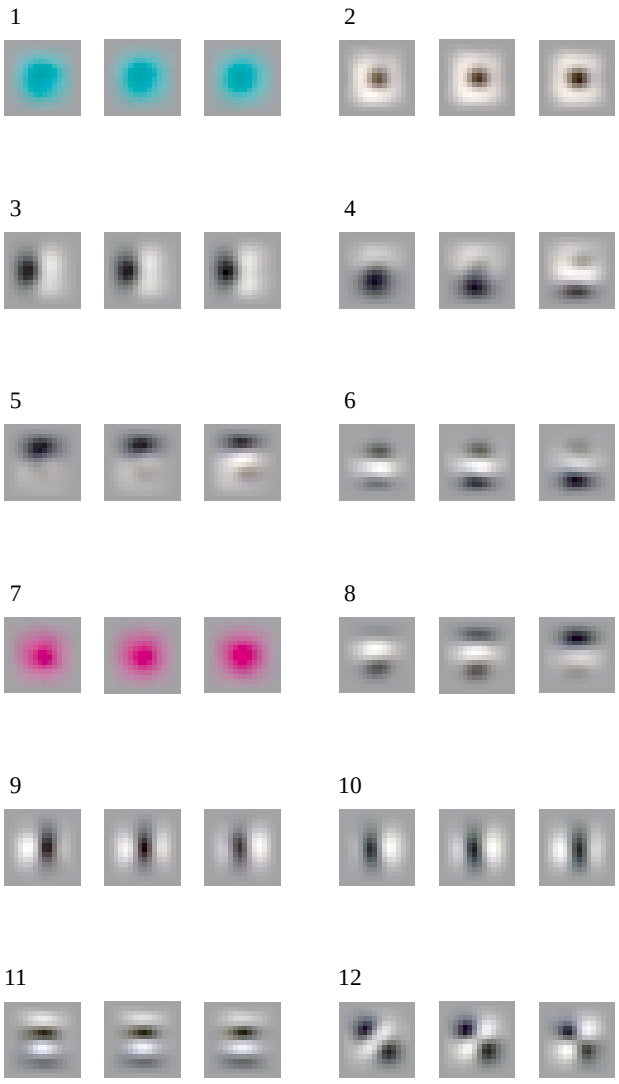

Figure 3. The spatio-temporal receptive fields obtained from one of the 50 ICA analyses described in the main text. The ith receptive field corresponds to the kernels $h_{ij}(u)$ of the spatio-temporal 'un-mixing matrix' defining the independent components. j indexes the spatial location and wavelength and u time. These receptive fields are arranged according to their spatial location at three time-points in the recent history of the retinal transients they sample (i.e. $h_{ij}(60\,\text{ms})$, $h_{ij}(80\,\text{ms})$ and $h_{ij}(90\,\text{ms})$). Kernel coefficients were normalized such that $\max(\text{abs}(h_{ij}(u))) = 0.5$ and 0.5 was added to each coefficient. The resulting values for each of the three wavelengths were used to specify the colour, in terms of red, green and blue at each location in the receptive field. This display format accommodates negative connection strengths to a particular wavelength, at a particular location and time and renders zero connectivity an intermediate grey.

visual scene, presented at the same instant, become temporally dispersed at a perceptual level.

(c) *Empirical estimates*

What are the likely time-frames involved for sensory systems? The entropy (figure 1) can be thought of as an upper bound on the information about the environment available in a transient (see below). The actual information in the responses of a neuronal population would depend on the Volterra kernels that effect a nonlinear transformation of this input. Say we had to differentiate between 32 different visual objects. We would then need $\log_2(32) = 5$ bits of information. This would only be available after sampling a transient for at least 10 ms. Interestingly this is the same conclusion reached by Tovee *et al.* (1993), who used information theory to analyse the spike-trains elicited by several faces, in different locations, in the temporal cortex of rhesus monkeys. In these experiments most of the information pertaining to a 500 ms transient (the first principal component of many trials) was available in the first 20–50 ms of activity. de Ruyter van Steveninck *et al.* (1997) reached similar conclusions using a different approach to measuring entropy, based on a discrete event space of firing sequences. They estimate that the average information in a 30 ms window, with a time-resolution of 3 ms, was about 5 bits. The convergence between these analyses of spike-trains and our magnetoencephalography (MEG) analysis should not be overinterpreted because entropy is not scale invariant (note the entropies in figure 1 were based on correlation matrices) and spatially integrated neuromagnetic signals cannot be compared easily to spike-trains. Furthermore, there are some special issues to be considered when trying to characterize the information in spike-trains. An excellent discussion can be found in Rolls & Treves (1997).

This section observed that transients contain more information than instantaneous codes. By virtue of the fact that the transient neuronal responses of any neuronal population are constructed by a nonlinear (Volterra) convolution of its inputs, these responses will reflect the history of neuronal activity elsewhere. This precludes any representation of an instant in time that is not incorporated into its immediate history. The form of the Volterra kernels, mediating the influence one population exerts over another, will determine the degree of this temporal embedding. The principles that underpin the 'best' kernels remain to be elucidated. However, there is one situation in which the optimum kernels may be defined and that is in early sensory cortices. Here there should be the highest degree of predictability of the evoked transients, given the sensory inputs causing them. In § 3, we pursue this information-theoretical approach to transients and show that some remarkable predictions can be made by simply considering what is the best way for the brain to extract information from visually evoked transients in early visual processing.

3. TRANSIENTS AND FUNCTIONAL SEGREGATION

(a) *V2 as a functional segregator*

This section concerns the principles that underlie functional specialization in visual cortex and, in particular, how the principle of maximum information transfer, paired with the notion of neuronal transients, predicts some fundamental features of segregation in early visual processing. Functional specialization depends on extrinsic and intrinsic connections within and among cortical units, populations and subareas, whose convergent and divergent architecture underlie the segregation of features in the visual field (Zeki 1990). This segregation is reflected in the emergence of distinct spatio-temporal receptive fields of units at various stages of the visual pathways. In this section we will focus on V2 as the final common stage in the segregation of retinal input. In what follows we will use

a framework of functional segregation that is extremely well synthesized and described in Zeki (1993).

One of the most fundamental features of segregation in the visual brain is a successive bifurcation of visual processing pathways that is apparent at a number of levels. In terms of projections from the retina to the lateral geniculate nuclei (LGN), there is a distinction between the magnocellular and the parvocellular pathways, projecting to the lower two and upper four layers of the LGN, respectively. The magnocellular pathway originates in the M ganglion cells of the retina and is relayed though the LGN to layer 4B of V1 and on to the thick stripes of V2. These M pathways can be regarded as undergoing a second bifurcation, sending efferents to the motion sensitive area V5 (the motion pathway) and V3 (dynamic form). The parvocellular pathway has its origin in the P ganglion cells and ultimately divides to give a colour pathway and a form pathway based on colour. From the P layers of LGN the pathways are relayed to layers 2 and 3 of V1 where they feed the blobs (colour pathway) and interblobs (form from colour). These two subdivisions are relayed to V4 through the thin and interstripe structures of V2, respectively.

It can be seen that V2 is a critical point of divergence, representing the last stage of the visual hierarchy that retains a full complement of functionally selective cells (although there are also direct connections from V1 to V3, V4, and V5). The physiology of V2 (Hubel & Wiesel 1977; Zeki 1993) shows that V2 contains functionally heterogenous populations of cells, i.e. orientation-selective, direction-selective and wavelength-selective units are all found within its subareas. The thick stripes of V2 receive their input from layer 4B of V1, where orientation and direction cells predominate and mediate motion or dynamic form processing through their connections to V5 and V3, respectively. Not surprisingly, direction-selective cells are concentrated in the thick stripes of V2. The thin stripes of V2 receive their input from the blobs of V1, where the majority of cells are not orientation selective but many are wavelength selective. Finally the interstripes receive input from the interblobs and show orientation- but not wavelength-selective responses (Shipp & Zeki 1985; De Yoe & Van Essen 1985; Hubel & Livingstone 1987). In summary, thick stripes contain orientation- and direction- but not wavelength-selective units. Thin stripes contain wavelength- but not orientation- and direction-selective units and the interstripes contain orientation- but not direction- or wavelength-selective units. Clearly this is a gross simplification but a useful one and leads to a clear trichotomy of selective spatio-temporal responses in V2.

Could this unique parsing of orientation, direction and wavelength selectivity have been predicted on the basis of theorizing alone? It could have been at a heuristic level: if one considers the brain as an inferential machine (see Dayan *et al*. 1995), a system that is trying to capture, represent or model the underlying causes in the sensorium, then an elemental visual event can have, among others, three causes. It could be caused by light with a particular wavelength composition, reflected from a visual feature that may or may not be moving and that may or may not be orientated. Clearly the wavelength composition is not determined by the motion or spatial form of the event, leading to cells that extract this cause (i.e. wavelength selective but not orientation or direction selective). The spatial structure of a small patch of retinal input is not necessarily dependent on its wavelength or motion (leading to orientation- but not wavelength- or direction-selective cells) and finally the motion of the patch is not a function of its colour but is necessarily dependent on some spatial structure that is moving (leading to orientation and direction selectivity in the absence of wavelength selectivity).

The above analysis depends on the assumption that the brain is an inferential device that tries to extract the underlying causes of the input it receives as efficiently as possible. In what follows we make this line of reasoning more precise by framing it in terms of information theory. The critical aspect, from the point of view of this paper, is that to make any meaningful inferences about events, particularly those involving motion, one needs to consider the information embodied in neuronal transients—in this instance the transients evoked at a retinal level by visual events. In terms of unit responses in V2, this translates into an analysis of the predicted spatio-temporal receptive fields and the underlying Volterra kernels used to construct the responses.

In what follows it will be shown that, by combining neuronal transients and the principle of maximum information transfer, not only does response selectivity emerge spontaneously, but the segregation of selectivity described above is emulated exactly, leading to predictions about receptive field properties that are borne out by electrophysiological and neuroanatomical studies of V2 (Hubel & Wiesel 1977; Shipp & Zeki 1985; De Yoe & Van Essen 1985; Hubel & Livingstone 1987). First we will discuss the principle of maximum information transfer and its relationship to efficient coding and redundancy. We will then consider neuronal transients and their implications for the dynamical aspects of receptive fields. In particular, transients are used to motivate a characterization of spatio-temporal receptive fields, which includes the time domain. This characterization is provided by the Volterra kernels, or effective connections, that specify a unit's responses to its inputs. By applying the principle of maximum information transfer, in a way that explicitly accommodates the time dimension, we can determine an 'optimum' set of receptive fields (i.e. Volterra kernels). The spatio-temporal fields that ensue can then be characterized, in terms of their selectivity, to see if they fall into the groups suggested by the empirical evidence above.

(b) *Efficiency, redundancy and the principle of information maximization*

The principle of maximum information transfer (e.g. Linsker 1988; Atick & Redlich 1990; Bell & Sejnowski 1995) has proved extremely powerful in predicting some of the basic receptive field properties of cells involved in early visual processing (e.g. Olshausen & Field 1996). This principle represents a formal statement of the common-sense notion that neuronal dynamics in sensory systems should reflect, efficiently, what is going on in the environment (Barlow 1961). Whether this principle holds at higher levels of sensorimotor integration and cognition remains an open question. However, it is clear that

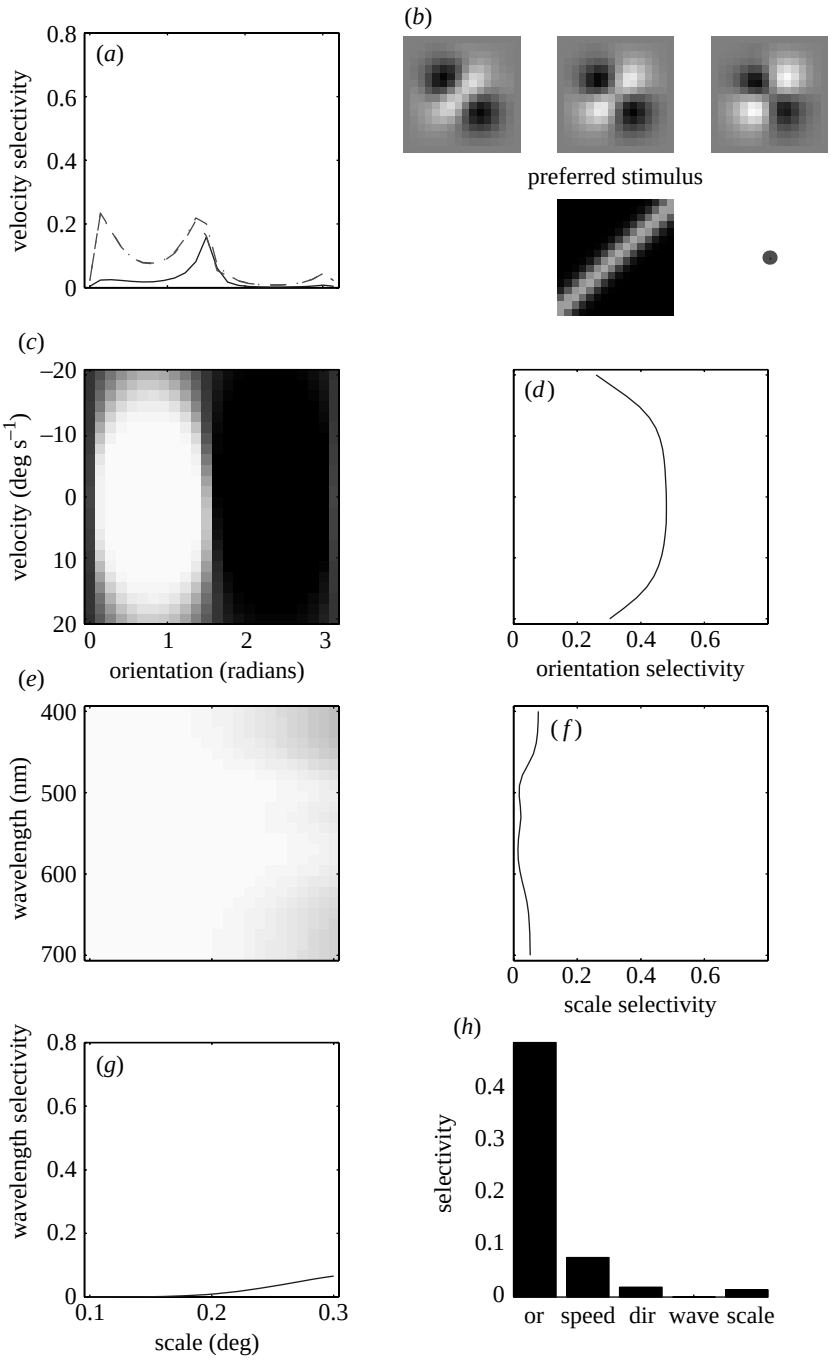

Figure 4. Tuning curve and selectivity analysis of an orientation selective receptive field. (*b*) The upper-three inserts depict the receptive field in question and conform to the display format adopted in figure 3. The lower insert shows the orientation of the preferred stimulus (that eliciting the greatest response). The preferred velocity is indicated by the small pointer to the right (non-existent in this case because the preferred velocity was zero). The two images (*c*, *e*) correspond to arrays of tuning curves (obtained by computing the response to simulated bar stimuli of different orientations, wavelengths, scale, eccentricity, etc.). (*c*) depicts velocity tuning as a function of orientation (or vice versa) and (*e*) wavelength tuning as a function of scale (or vice versa). By taking the maximal difference in evoked responses, the selectivity for each attribute was computed as a function of the other (shown in the four graphs (*a*, *d*, *f*, *g*) aligned with the two images). Velocity selectivity (*a*) is decomposed into speed (with responses averaged over both directions, dashed line) and direction (averaged over speeds, dot–dash line). (*h*) The selectivity profile summarizes these data, showing the response differential in relation to the maximum response elicited. In this instance the receptive field shows clear orientation selectivity, and only orientation selectivity, responding to bars at about 45° at all wavelengths, scales and over a broad range of speeds.

adaptive responses at any level necessitate a high degree of mutual information between the dynamics of visual cortex and changes in the visual world as sampled at the retina. In the present context the principle of maximum information transfer suggests that the receptive fields of visual neurons should be configured in a way that maximizes the mutual information between the neuronal activity that they engender and the sensory inputs on which they are contingent. This maximization is usually considered in the light of some sensible constraints, for example, the presence of noise in the sensory input (Atick & Redlich 1990) or dimension reduction (Oja 1989),

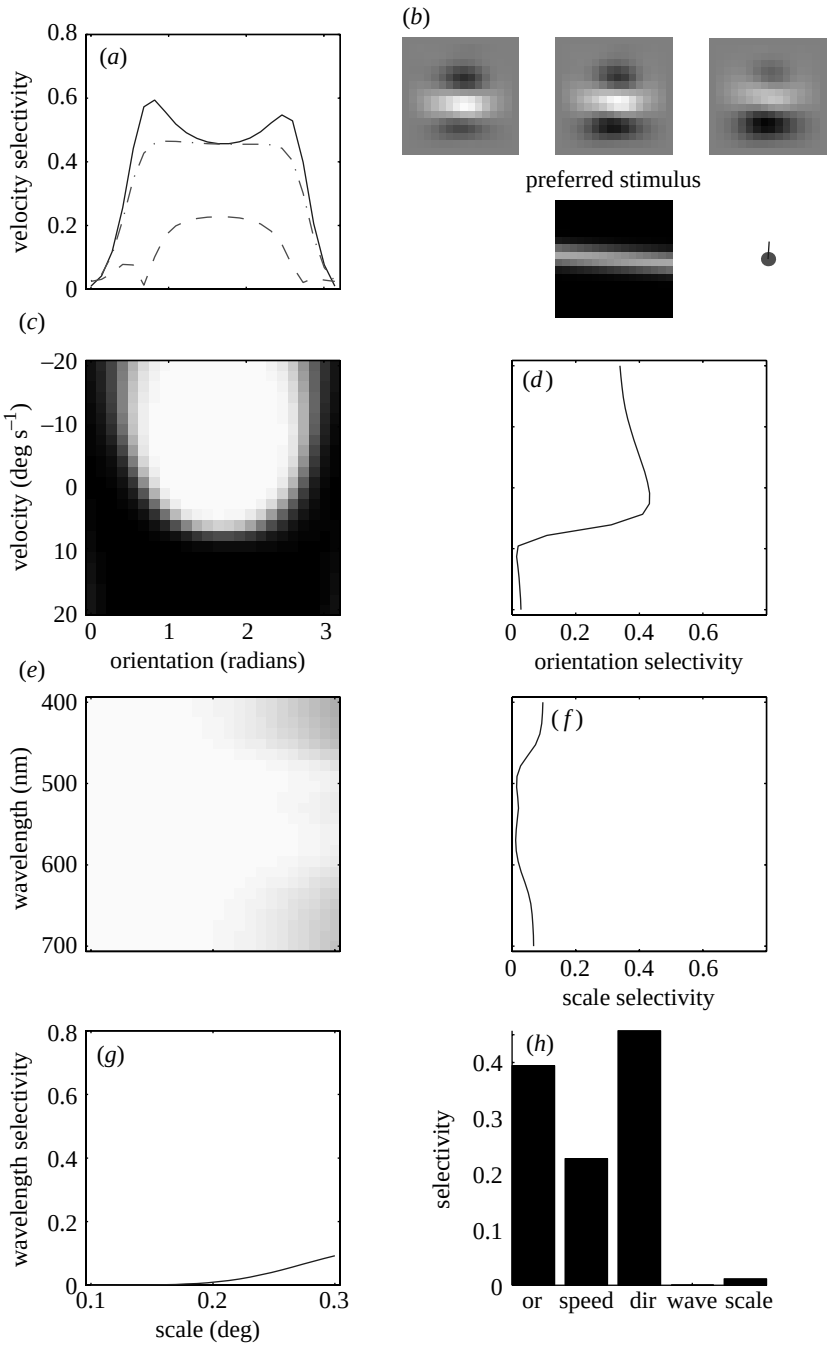

Figure 5. Tuning curve and selectivity analysis of a direction selective receptive field. For the format of this figure, see the legend to figure 3. This 'cell' prefers upwards moving horizontal bars and evidences a substantial amount of speed, direction and orientation selectivity but is relatively indifferent to wavelength or scale.

implicit in the fact that there are a smaller number of divergent outputs from a neuronal population than convergent afferents (Friston *et al.* 1992).

This principle is closely related to the idea of efficient coding. It is sometimes difficult to see the close relationship among all the various perspectives taken on (and terms used) by different authors. Generally speaking the principles of maximum information transfer, sparse coding, redundancy minimization and efficient coding are all variations on the same theme. We will spend some time trying to relate these perspectives and show that the only thing that really distinguishes among them is the nature of the constraints under which the most information is extracted. For a deterministic system, in other words, one in which noise can be disregarded, the mutual information between the input and the output reduces to the average information or entropy of the output. Consider again the Volterra series as a model for the dependency of activity in a population of units in visual cortex (o) on activity in a retinotopically corresponding population in the retina (x):

$$o_i(t) = \Omega_0[x(t)] + \Omega_1[x(t)] + \ldots + \Omega_n[x(t)] + \ldots . \quad (2)$$

For any given input x, we want to maximize the mutual information between x and the output o. The mutual information is given by

$$I\{o,x\} = H\{o\} - H\{o|x\}, \quad (3)$$

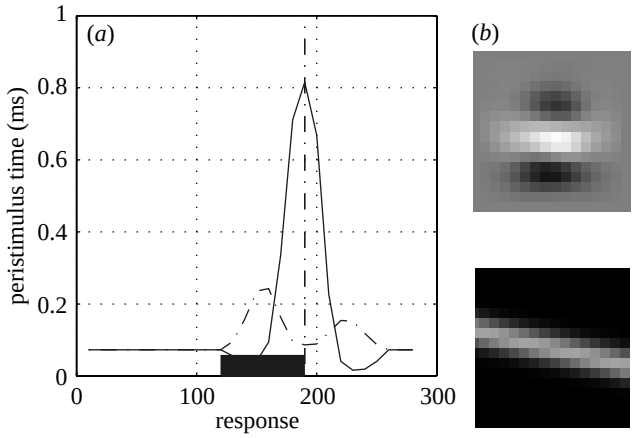

Figure 6. The responses of the direction selective 'cell' depicted in figure 5 to identical bar stimuli (lower insert in (*b*)) moving in opposite directions. The attenuated response in the null direction (dashed line) can be attributed to the nonlinearities implicit in equation (5).

where $H\{o|x\}$ is the conditional entropy or uncertainty in the outputs, given the inputs. For a deterministic system there is no such uncertainty and $H\{o|x\}$ can be discounted (see Bell & Sejnowski 1995). It follows that maximizing the output entropy is the same as maximizing the mutual information. The efficiency of a neuronal system can be considered as the complement of redundancy, the less redundant, the more efficient a system will be. More formally

$$\text{efficiency} \sim I = H\{o\} - \Sigma H\{o_i\}, \tag{4}$$

(cf. Gawne & Richmond 1993) where o_i are the constituent units in the output population. I is sometimes referred to as simply the 'information' in a system and is ubiquitous in the independent component analysis and related literature as the objective function that is maximized. Equation (4) says that efficiency is the difference between the joint entropy and the sum of the entropies of the individual units (componential entropies). Intuitively this makes sense if one considers that the variability in activity of any one unit corresponds to its entropy. Therefore an efficient system, embodying a fixed $H\{o\}$, does so with the minimum changes in firing. It also follows that, subject to the constraint that the componential entropies $\Sigma H\{o_i\}$ are the same, increasing the efficiency increases the mutual information between input and output though maximizing $H\{o\}$. Maximizing $H\{o\}$ usually involves removing correlations or mutual predictability among the output units. This is equivalent to ensuring that the output 'selectivities' are as dissimilar as possible. Approaches that seek to maximize the joint entropy of the outputs include principle component analysis (PCA) learning algorithms, which sample the subspace of the inputs that have the highest entropy, and independent component analysis (ICA), which finds nonlinear functions of the inputs that maximize the entropy subject to different but appropriate constraints (see § 3(c); Bell & Sejnowski 1995). In PCA, the componential entropies, or variances of the individual units, are constrained by setting limits on the weights used to linearly transform the inputs (so that they have unit sum of squares). In ICA, the outputs are constrained to lie in some bounded range by the application of a nonlinear squashing function to compounds of the inputs. In both PCA and ICA, the output entropy is maximized, explicitly in ICA, and by ensuring the outputs are orthogonal and account for the largest variance in PCA.

The alternative approach to increasing efficiency is to minimize the componential entropies while ensuring the joint entropy remains high. The latter is assured as long as the outputs can reliably predict the inputs. This minimization is generally associated with sparse encoding of salient features of the inputs. In other words, a unit that only fires infrequently will generally be not firing. Because of this, its state is quite predictable and $H\{o_i\}$ will be small. This approach is illustrated nicely in Olshausen & Field (1996).

In this work, we consider that the 'best' set of receptive fields, associated with a point in retinotopic space, corresponds to a set of nonlinear functions (Volterra operators) of visually evoked retinal dynamics that has the maximum joint entropy. This ensures efficient coding and conforms to the principle of maximum information transfer.

(c) *Neuronal transients and maximizing information transfer*

Perhaps the simplest examples of neuronal transients are the self-limiting dynamics that are elicited by salient events as seen in evoked potential studies. In the current context, the importance of neuronal transients is that the pattern of activity elicited by a visual stimulus in retinal or geniculate units has an explicit temporal domain. This is crucial when considering the responses of individual neurons higher in the visual system. The response of a particular unit, say in V2, is a function not only of the retinal activity at that time, but the recent history of retinal dynamics mediated by polysynaptic relays. This is a consequence of (i) lateral interactions, mediated by intrinsic connectivity, and (ii) recurrent interactions among reciprocally linked populations in the visual pathways, mediated by extrinsic connectivity. The response of a unit at any time will be a highly nonlinear function of inputs from extrinsic afferents from lower areas, lateral inputs from within the unit's area and re-entrant inputs from higher areas. These will be a function of activity patterns at some earlier time. By recursion, it follows that the response to retinal inputs at the current time also includes components that are due to retinal inputs at all previous times. In short, any neuron has a receptive field that embraces not only all the presynaptic afferents it receives, but the activity in those afferents now and in the recent past. Given that the time taken for activity to be propagated along recurrent forward and backward connections could be in the order of tens of milliseconds, the temporal extent of a unit's spatio-temporal receptive field could be as large as 100 ms or more. Therefore, to determine the unit's response on the basis of retinal input, one would need to know the neuronal transient that has just been expressed at a retinal level.

This perspective offered by neuronal transients leads to a picture of selective responses and receptive field configurations that is much less hierarchical than in conventional formulations. Although more complicated

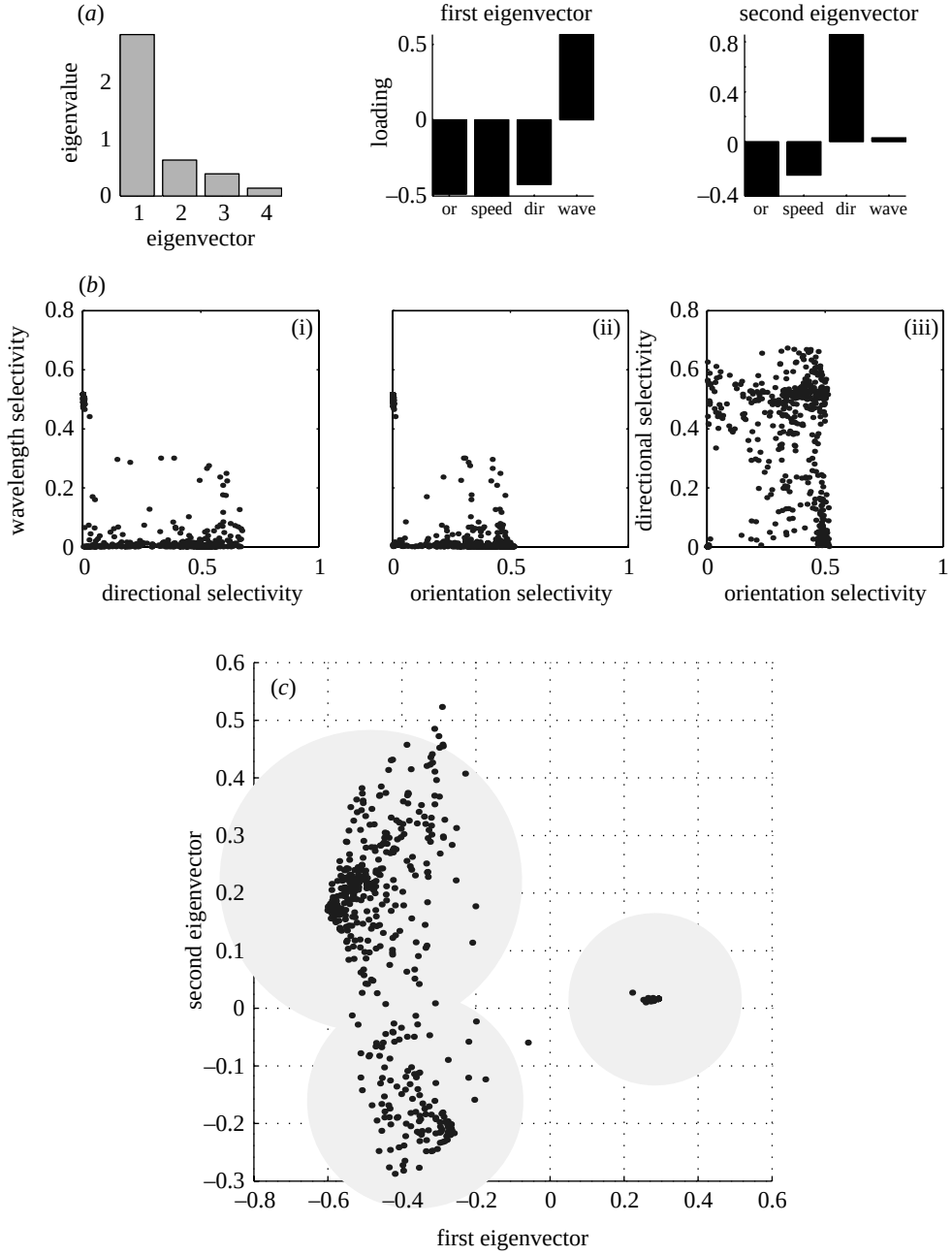

Figure 7. Segregation of selective responses. These results constitute a meta-analysis of the selectivity profiles of all 'cells' over all ICA analyses. The selectivities for orientation (or), speed (speed), direction (dir) and wavelength (wave) were subject to a PCA (after mean correction and Euclidean normalization). The resulting eigenvalue spectrum and first two eigenvectors (i.e. principal components) are shown in (*a*). These suggest that the main axis of segregation is between 'cells' showing wavelength selectivity and those that do not. The second axis of segregation pertains to direction selectivity. The interrelationships between these selectivities are shown directly (*b*) by plotting the three attributes (wavelength, direction and orientation) against each other. The segregation into three selectivity groupings is evident using a principal coordinate analysis (*c*) in which the first two principal component scores of each 'cell' are plotted against each other.

receptive fields may be assembled from simpler receptive fields at lower levels, a more dynamical view suggests that unit responses at every level in the early visual pathways have access to information from all other levels, and at previous times, mediated by abundant backwards connections. For example given that the LGN receives more afferents from the cortex than it does from the retina, does it make sense to consider the LGN as a 'lower' visual station than the cortex? In other words, by virtue of the recurrent and embedded loops arising from backwards connections is it appropriate to place any component of the loop as 'higher' in relation to another component? Although an interesting perspective, this view should be moderated by noting that the visual pathways have to extract the causes of changes in the visual field by constructing highly nonlinear functions of visual inputs in accord with equation (2). To construct these functions it has to use a series of nonlinear transformations that are constrained by the neuronal infrastructure available. To get a sufficiently nonlinear transformation it may require several weakly nonlinear stages, implemented at each synaptic relay, or area, in a polysynaptic chain. If this is

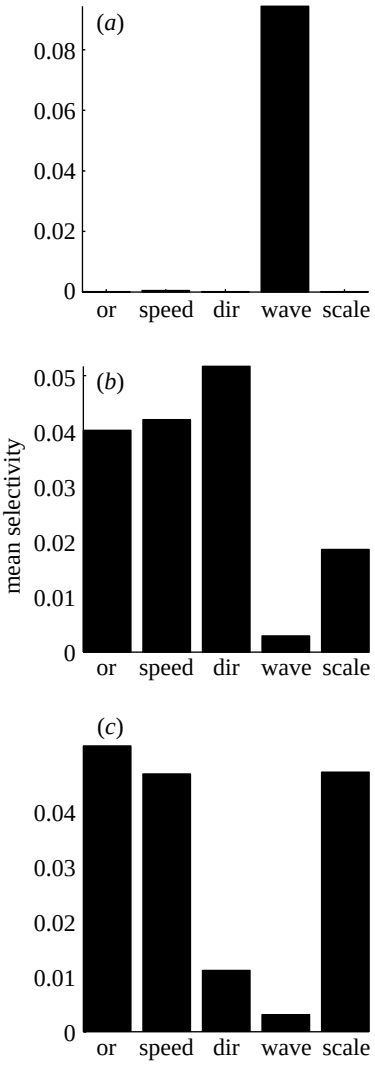

Figure 8. Mean selectivity profiles for the three groups of segregation identified in figure 7. The smallest group (*a*) has, almost exclusively, wavelength selectivity (cf. cells in the thin stripes of V2). The largest group (*b*) is direction-, speed- and orientation-selective (cf. cells in thick stripes). The cells in the intermediate group (*c*) show substantial orientation and scale selectivity (cf. cells in the interstripes).

the case, a hierarchy might be a natural consequence of the fairly stereotyped response properties of neurons themselves. This again highlights the importance of constraints when considering how principles like information maximization might be instantiated in the brain.

In principle, a Volterra expansion of retinal input could accommodate all the lateral and backwards modulatory effects alluded to above if the only cause (i.e. input) of the V2 responses (output) was retinal. In this instance the Volterra kernels are 'standing in' for all the polysynaptic transformations and effects of recurrent loops that mediate V2 responses to retinal changes. These responses, and implicitly the kernels, define the receptive fields. Motivated by the importance of constraints and some recent mathematical advances, let us assume a fairly simple form for this Volterra series equation (2), which describes the nonlinear transformation between retinal inputs and unit responses in retinotopically equivalent points in V2.

$$o_i(t) = \sigma\left\{\sum_j \int_0^\infty h_{ij}(u) \times x_j(t-u)\mathrm{d}u\right\}, \tag{5}$$

where $\sigma\{\cdot\}$ is a nonlinear sigmoid or squashing function (the logistic function) that ensures the outputs lie in the range 0 to 1. The receptive fields of unit i in V2 are determined by the coefficients $h_{ij}(u)$. These can be thought of as the time-dependent effective connectivity between the jth unit in the retina and the ith unit in V2. The time-dependent or dynamic connection strengths define the spatio-temporal receptive field of the simulated V2 units. Strictly speaking, the inputs should include any unit that can exert an influence, directly or through polysynaptic relays. In this paper, we ignore spatial integration and top-down influences of an unspecified nature and just consider the inputs deriving from a small patch of the retina. Equation (5) has high-order terms by virtue of the sigmoid squashing function and is a simple variant of time-delayed neural networks as considered by Wray & Green (1994) in the context of Volterra series. Intuitively it says that the activity of any V2 unit can be modelled as a nonlinear function of inputs, where these inputs are the activities of retinal units over the recent past, convolved or weighted over space and time by some input-specific kernel.

To apply the principle of maximum information transfer we have to find the dynamic connection strengths $h_{ij}(u)$ that define the unit's spatio-temporal receptive field. In other words, we have to find $h_{ij}(u)$ that maximizes $H\{o\}$ where

$$H\{o\} = H\{x\} + \langle \ln(\mathcal{J}(x(t))) \rangle, \tag{6}$$

where $\mathcal{J}$ is the Jacobian associated with equation (5) and is a function of $h_{ij}(u)$. Given that the entropy of the inputs is fixed we have maximize the right-hand term in equation (6). Fortunately this can be achieved with relative ease using ICA. Indeed ICA has been applied in the context of static receptive fields (i.e. ignoring the temporal domain) with compelling results (Bell & Sejnowski 1997; Van Hateren & Van der Schaaf 1998). If one could find the dynamic connections $h_{ij}(u)$, then the corresponding spatio-temporal receptive fields would maximize the mutual information, not between the outputs and the inputs at any one time, but between the outputs and the inputs over the recent past. In other words, the receptive fields are construed as mediating responses to salient visual events as opposed to spatial patterns. By trying to solve this more complicated, but biologically more pertinent, problem, we hypothesized that the ensuing receptive fields would conform closely to those actually observed in the real brain. In particular we would expect to see the selectivity and segregation of selective responses of the sort described above.

In summary, assuming a model like equation (5), the coefficients $h_{ij}(u)$ that maximize output entropy, and implicitly information transfer, can be identified using techniques developed for independent component analysis (ICA) for any retinal input sequence. By scanning natural scenes and transforming the data into simulated retinal responses, one can identify optimum dynamic connections $h_{ij}(u)$ and implicitly the associated response properties or spatio-temporal receptive fields.

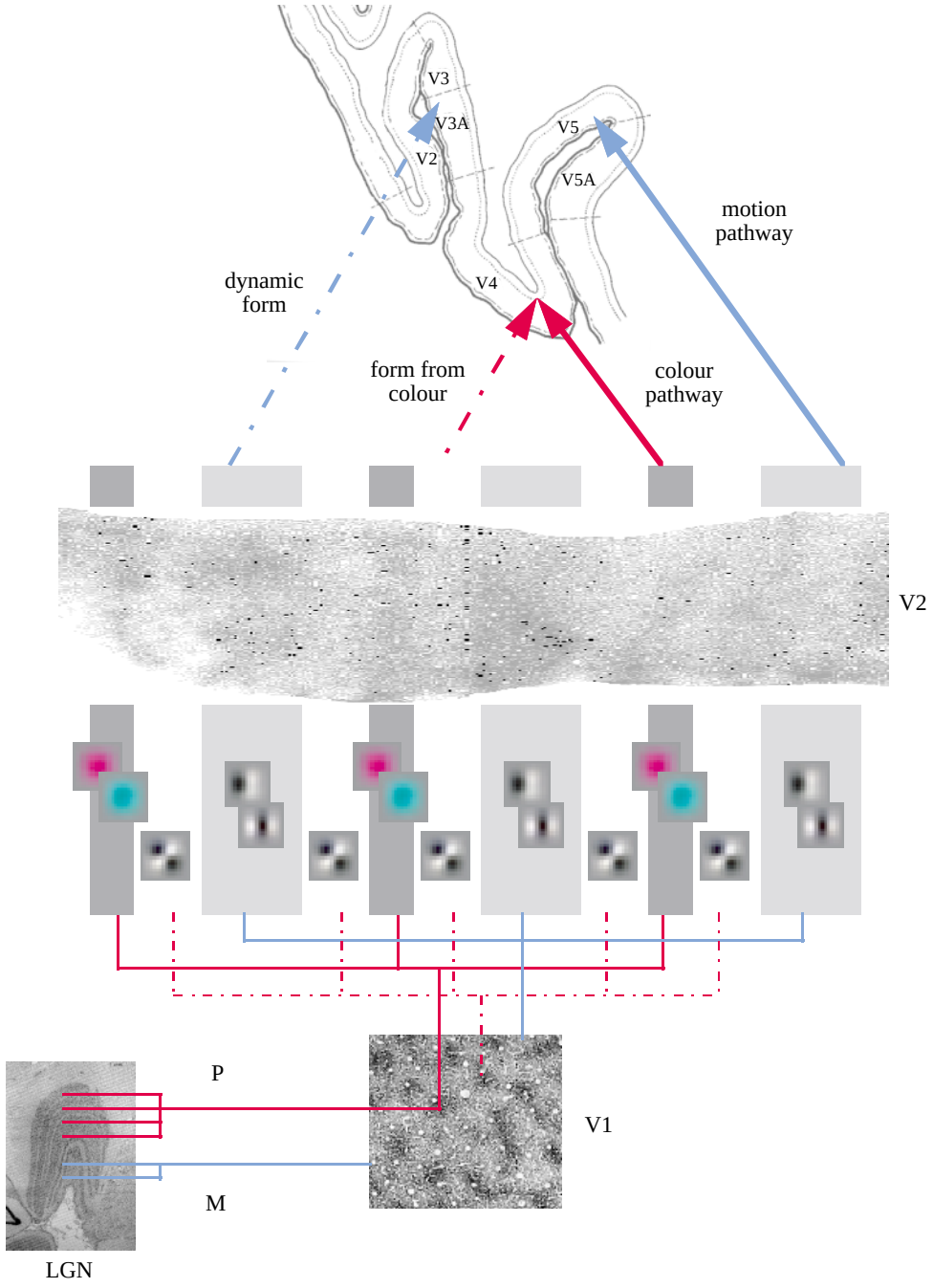

Figure 9. Schematic adapted from Zeki (1993) summarizing the anatomy and functional segregation of processing pathways and the relationship of receptive fields to the stripe structures in V2. LGN, lateral geniculate nucleus; P, parvocellular pathway; M, magnocellular pathway. The receptive fields come from the analysis presented in figure 3.

One can then perform simulated experiments, using conventional bar or grating stimuli, to characterize the selectivity of these receptive fields and compare the ensuing profiles with those observed empirically. An example of this approach is presented below.

(d) *Simulations using natural images*

In the simulations reported here, retinal inputs were simulated by sampling natural coloured images with a 16×16 pixel array moved in little 'sweeps' over the images at an average rate of one pixel per iteration (an example of one of these scenes is given in figure 2). In these simulations, a pixel corresponds to about 0.1° and one iteration to 10 ms, giving an average velocity of about $10^\circ\,s^{-1}$. The resulting input vectors comprised 16 voxels $\times$ 16 voxels $\times$ 16 time-steps for each primary colour ($1.6^\circ \times 1.6^\circ \times 160$ ms $\times 3$ wavelengths). To emulate retinal responses the tristimulus values obtained from the red, green and blue image components were transformed to a retinal cone colour coordinate system according to Pratt (1978) and log

transformed (a small constant of 0.05 was added prior to transformation to avoid logs of zero). This input to the simulation can be thought of as a time-series of instantaneous activity profiles, evoked by moving natural images, in three (colour channel-specific) sets of retinotopically organized photoreceptors.

Each input was reduced using 192 spatio-temporal basis functions (a three-dimensional discrete sine set ($4 \times 4 \times 3$), windowed in the spatial dimensions with a Hanning function). The coefficients $h_{ij}(u)$ underlying the hypothesized receptive fields of 12 units in V2 were determined using ICA to maximize the entropy according to equation (6). Because the algorithm used only returns outputs that lie in the subspace spanned by the initial weight matrix, we used an initial weight matrix that corresponded to the first 12 principal components of the spatio-temporal inputs. This number typically accounts for about 80% of the variance in the simulated retinal dynamics. We used eight different natural scenes sampled with 4096 sweeps of random direction (uniform) and velocity (Gaussian) for each ICA analysis. This was repeated 50 times to ensure stability of the results. The choice of 12 units was motivated by noting that this was the maximum number that gave unequivocally stable results, in terms of the ensuing receptive fields.

The results of a typical analysis (one of the 50) are shown in figure 3, where, for each of the 12 units, the connection strengths are plotted, at three points in time, in the appropriate retinotopic position and colour. It is immediately obvious that these receptive fields fall into two classes. One class shows marked wavelength selectivity (units 1 and 7) whereas the other does not. The blue hue of receptive field 1 reflects the prevalence of sensitivity to short wavelength inputs and not to long (green and red) wavelengths and corresponds roughly to a blue–yellow axis. The red-purple hue of receptive field 7 indicates a sensitivity to long but not intermediate (green) wavelengths, i.e. a red–green axis.

This coloured-grey dichotomy over receptive fields is the first indication that the distinction between wavelength selective responses and non-selective responses is an emergent phenomena. It is quite remarkable that some units are wavelength selective and others are not. If we had selected the coefficients $h_{ij}(u)$ at random then the probability of getting even one uniformly grey receptive field (i.e. no wavelength selectivity) would have been exceedingly small.

Note furthermore that, with the exception of receptive field 2, all the grey fields show some orientated structure, in contradistinction to the coloured fields. This suggests that units showing no wavelength selectivity may well be orientation selective. Some of the fields seem to be static (e.g. field 11), whereas others evidence dynamic changes in the receptive field that might belie motion or direction selectivity (e.g. field 6).

Our predictions were that the wavelength-selective cells would not show orientation or direction selectivity (cf. thin stripes in V2) and conversely wavelength-insensitive units may (thick stripes) or may not (inter-stripes). To test this hypothesis explicitly, we characterized the receptive field properties of our simulated V2 cells by presenting moving, monochromatic bar stimuli and measuring the simulated responses. Receptive field selectivity was assessed in the following way. Each stimulus comprised a bar with a Gaussian profile that was characterized by five parameters: (i) the orientation of the bar (from 0 to π radians), (ii) the velocity of the bar (from -20 to $20^{\circ}\,s^{-1}$); (iii) the monochromatic wavelength employed (from 400 to 700 nm); (iv) the scale or width of the bar (0.1 to 0.3°); and (v) its eccentricity, defined as the displacement from centre at the midpoint of presentation (from -0.4 to 0.4°). The responses of each simulated unit to stimuli of 160 ms duration were computed according to equation (5) using the estimates of $h_{ij}(u)$ from each ICA analysis. The response was taken to be that immediately following the stimulus presentation. Responses were evaluated over all possible stimulus-event configurations resulting in a five-dimensional array of responses. The mode or maximum of this response profile represents the preferred stimulus for the unit in question. Figure 4 shows some typical results, in this instance from receptive field 12 in figure 3. Figure 4*b* shows the receptive field and the preferred stimulus. Figure 4*c* shows the response profile over orientation and velocity at the preferred wavelength, scale and eccentricity. This image can be thought of as a collection of orientation tuning curves (obtained at different velocities) or equivalently, as velocity tuning curves at different orientations. It is clear that maximal responses are obtained with static stimuli at about 45° orientation. From this profile we can compute an orientation selectivity at each velocity (figure 4*d*). Selectivity was simply defined as the difference between the maximum and minium responses. Clearly this selectivity is high and relatively insensitive to the stimulus's velocity. Similarly we computed the velocity selectivity as a function of orientation (figure 4*a*). Velocity has two components, the speed and direction. For the purposes of further analysis we decomposed velocity selectivity into speed selectivity (maximal differences in responses averaged over both directions) and direction selectivity (maximal differences in responses averaged over all speeds). This cell clearly shows only moderate direction selectivity, even at the preferred orientation. The lower image shows the response profile over wavelength and scale. In this instance the responses are almost uniform suggesting very little wavelength (figure 4*g*) or scale (figure 4*f*) selectivity. In short, this receptive field shows orientation selectivity but not wavelength or direction selectivity. This is apparent if one plots the selectivity for each attribute when using the preferred stimulus parameters for this cell (figure 4*h*). This selectivity profile was computed for each cell in all the ICA analyses. Figure 5 shows the results for receptive field 6 in figure 3. This cell shows a moderate amount of orientation selectivity and is very direction selective. An alternative demonstration of this selectivity is presented in figure 6, where the dynamic responses from equation (5) are plotted as a function of time for two bar stimuli that were identical other than in their direction of motion. It can be seen that there is a vigorous response following stimulation with the preferred direction. However, there is a greatly attenuated response for the same stimulus moving in the null direction. It should be noted that these two stimuli were presented for the same duration, with the

same luminance and covered the same points in retinotopic space. The only difference was the order or direction in which the retinal inputs were stimulated and yet there is a profound difference in the evoked transient.

(e) *Functional segregation*

It now remains to show that the selectivity of the simulated spatio-temporal receptive fields segregate as one might predict on the basis of electrophysiological studies. To characterize this segregation the selectivity profiles of each unit, from all the ICA analyses, were pooled, normalized and subject to a PCA (figure 7*a*). The first principal component or eigenvector showed that the main difference among selectivities was a wavelength versus non-wavelength selectivity. This fits pleasingly with the fundamental dichotomy suggested by the response profiles of cells in the parvocellular and magnocellular pathways. The second principal component suggested that the next most important distinction is between those cells that show direction-selective responses and those that do not.

The interrelationships, among the selectivities for different attributes, are shown by plotting them against each other in figure 7*b*. It is clear that directionally selective cells are not wavelength selective and vice versa (figure 7*b*(i)), similarly for orientation and wavelength selectivity (figure 7*b*(ii)). However, many orientation-selective cells are, not surprisingly, directionally sensitive. The underlying grouping or segregation of selective responses is revealed more clearly by plotting each unit's scores on the first and second principal components against each other. This is known as principal coordinates analysis. In this space it can be seen that units fall roughly into one of three groups, denoted by the circles in figure 7*c*.

The mean selectivity of units within each of these groups is shown in figure 8 and conforms exactly to what was predicted above. Namely a small group of cells that show wavelength selectivity but minimal direction or orientation selectivity. This group corresponds, in our conceptual model, to the units one might typically find in the thin stripes of V2. The largest group shows direction and orientation selectivity but little wavelength selectivity and represents the sorts of response properties found in the thick stripes of V2. An intermediate-sized group, corresponding to the interstripes, shows pronounced orientation selectivity but little direction selectivity and minimal wavelength selectivity. It is pleasing that the size of each group corresponds roughly to the size of the stripe structures actually observed in V2.

Figure 9 is a schematic, based on Zeki (1993), which depicts the relationship between the receptive fields, predicted by neuronal transients and information theory, and the functional architecture of visual processing that is predicated on a synthesis of electrophysiological and anatomical evidence (e.g. Shipp & Zeki 1985; De Yoe & Van Essen 1985; Hubel & Livingstone 1987).

(f) *Temporal convergence and divergence*

The above analysis identified nonlinear transformations of neuronal transients that maximize the mutual information between an input that is temporally extended and the information at a single point in time. In this way, spatio-temporal receptive fields can be considered as mediating a convergence of temporal information, in this case coercing 160 ms worth of information into an instant of time. However, transient dynamics in V2 have a temporally extended domain and will themselves be subject to this sort of compression. This begs the question 'Is it sufficient to maximize the entropy of V2 dynamics at one point in time or should the entropy of V2 transients themselves be maximized?' This question relates to the pioneering work of Optican & Richmond (1987) and the powerful inferences (de Ruyter van Steveninck *et al.* 1997) that have been made through analysing the information in stimulus-locked spike-trains over extended periods of time.

4. CONCLUSION

The main points made in this paper can be summarized as follows.

(i) The upper limit of information contained in a neuronal transient is proportional to its length.
(ii) The length of a neuronal transient depends on the temporal extent of the Volterra kernels that mediate the response of a population to its inputs.
(iii) The existence of a Volterra series formulation of coupled neuronal populations places constraints on the temporal acuity of neuronal responses in that they are necessarily conflated with the recent history of activity in the brain. Temporally extended kernels confer greater context sensitivity but preclude the 'pure' representation of a instantaneous event.
(iv) If Volterra kernels are temporally extended then units in early visual cortex should have a pronounced spatio-temporal structure in their receptive fields. A test of this hypothesis obtains by applying the principle of maximum information transfer to estimate the optimum kernels, to confirm that they emulate the selectivity seen in the real brain.
(v) Applying the information theoretic principles to simulated retinal transients yields kernels (simulated spatio-temporal receptive fields) whose selectivity profiles resemble almost exactly those seen in the real brain.

This and Friston (paper 1 and paper 2, this issue) have presented the case for neuronal transients as a metric of brain dynamics and Volterra kernels as a characterization of the effective connectivity among neuronal populations that mediate them. Empirically, we have seen that asynchronous coupling between anterior and posterior brain areas can be extremely significant. This form of asynchronous coupling, which involves correlations among different frequencies, follows naturally from the coexpression of asynchronous transients in the two brain areas. The importance of neuronal transients, as a general framework for characterizing neuronal interactions, is that they embrace both synchronous and asynchronous coupling. Synchronization, as implied by temporal codes framed in terms of oscillations and phase-locking, or indeed non-oscillatory synchronized firing, can be thought of as a special case of transient coding. The reason it is important to consider asynchronous interactions is that they imply nonlinear coupling and it is this sort of integration that

provides for the diverse and context-sensitive expression of transients. As demonstrated, this nonlinear coupling can supervene in terms of its magnitude and significance in relation to linear or synchronized interactions.

The successive expression of diverse transients is related to dynamic correlations and more directly to dynamic instability. Dynamic instability may be crucial for adaptive brain function from two perspectives. The first is from the point of view of neuronal selection and self-organizing systems. If selective mechanisms underpin the emergence of adaptive neuronal responses, then dynamic instability is, itself, necessarily adaptive. This is because dynamic instability is the source of diversity on which selection acts, and is therefore subject to selective pressure. The second perspective is provided by information theory, in particular the principle of information maximization. By applying the principle of maximum information transfer to neuronal transients, receptive fields emerge that are reminiscent of those found in the real brain. A contribution of this component was to extend information maximization approaches to the temporal domain. Implicit in this extension is the idea that extrinsic and intrinsic connections have been selected, both on an evolutionary and somatic time-scale, such that they extract the most information from the sensorium. By virtue of the fact that this information pertains to an instant in time, this can be seen as a temporal convergence or 'compression' of information over time, or as a dilation of the neuronal moment in which representations of an 'instant' are lost forever. This may represent a fundamental aspect of neuronal dynamics and a perspective on temporal integration in the brain.

Finally it should be asked 'why all this is important?' Perhaps the most general and useful answer is that if one is trying to relate behavioural, psychophysical or other measures of brain function to the underlying neurophysiology then it is important to use the 'right' neurophysiological measures. The conclusions from these papers point to neuronal transients, if they can be measured.

K.J.F. is funded by the Wellcome Trust. I would like to thank Semir Zeki, Richard Frackowiak, Erik Lumer, Dave Chawla, Christian Büchel, Chris Frith, Cathy Price and the reviewers for their scientific input.

REFERENCES

Atick, J. J. & Redlich, A. N. 1990 Towards a theory of early visual processing. *Neural Comput.* **2**, 308–320.

Barlow, H. B. 1961 Possible principles underlying the transformation of sensory messages. In *Sensory communication* (ed. W. A. Rosenblith). Cambridge, MA: MIT Press.

Bell, A. J. & Sejnowski, T. J. 1995 An information maximization approach to blind separation and blind de-convolution. *Neural Comput.* **7**, 1129–1159.

Bell, A. J. & Sejnowski, T. J. 1997 The independent components of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338.

Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. 1995 The Helmholtz machine. *Neural Comput.* **7**, 889–904.

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.

De Yoe, E. A. & Van Essen, D. C. 1985 Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature* **317**, 58–61.

Friston, K. J., Frith, C., Passingham, R. E., Dolan, R., Liddle, P. & Frackowiak, R. S. J. 1992 Entropy and cortical activity: information theory and PET findings. *Cerebr. Cortex* **3**, 259–267.

Gawne, T. J. & Richmond, B. J. 1993 How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* **13**, 2758–2771.

Hubel, D. H. & Wiesel, T. N. 1977 The Ferrier Lecture: functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond.* B **198**, 1–59.

Hubel, D. H. & Livingstone, M. S. 1987 Segregation of form color and stereopsis in primate area 18. *J. Neurosci.* **7**, 3378–3415.

Jones, D. S. 1979 *Elementary information theory.* Oxford, UK: Clarendon Press.

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

Moutoussis, K. & Zeki, S. 1997 Functional segregation and temporal hierarchy of the visual perceptive system. *Proc. R. Soc. Lond.* B **264**, 1407–1414.

Oja, E. 1989 Neural networks, principal components, and subspaces *Int. J. Neural Syst.* **1**, 61–68.

Olshausen, B. A. & Field, D. J. 1996 Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609.

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

Pratt, W. K. 1978 *Digital image processing.* New York: Wiley.

Rolls, E. T. & Treves, A. 1997 *Neural networks and brain function.* Oxford University Press.

Shipp, S. & Zeki, S. 1985 Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature* **315**, 322–325.

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.

Van Hateren, J. H. & Van der Schaaf 1998 Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond.* B **265**, 359–366.

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.

Zeki, S. 1990 The motion pathways of the visual cortex. In *Vision: coding and efficiency* (ed. C. Blakemore), pp. 321–345. Cambridge University Press.

Zeki, S. 1993 *A vision of the brain.* Oxford, UK: Blackwell Scientific.