Principal component analysis learning algorithms: a neurobiological analysis

K. J. FRISTON†, C. D. FRITH AND R. S. J. FRACKOWIAK

MRC Cyclotron Unit, Hammersmith Hospital, London W12 0HS, U.K.

SUMMARY

The biological relevance of principal component analysis (PCA) learning algorithms is addressed by: (i) describing a plausible biological mechanism which accounts for the changes in synaptic efficacy implicit in Oja's 'Subspace' algorithm (*Int. J. neural Syst.* **1**, 61 (1989)); and (ii) establishing a potential role for PCA-like mechanisms in the development of functional segregation. PCA learning algorithms comprise an associative Hebbian term and a decay term which interact to find the principal patterns of correlations in the inputs shared by a group of units. We propose that the presynaptic component of this decay could be regulated by retrograde signals that are translocated from the terminal arbors of presynaptic neurons to their cell bodies. This proposal is based on reported studies of structural plasticity in the nervous system. By using simulations we demonstrate that PCA-like mechanisms can eliminate afferent connections whose signals are unrelated to the prevalent pattern of afferent activity. This elimination may be instrumental in refining extrinsic cortico-cortical connections that underlie functional segregation.

1. INTRODUCTION

An important class of learning algorithms, in unsupervised learning, are the principal component analysis (PCA) algorithms (Hornik & Kuan 1992; Hertz *et al.* 1991). These algorithms were originally developed from a theoretical perspective, with little reference to neurobiological implementation or implications. This article: (i) describes a plausible biological mechanism which could account for the synaptic changes implicit in PCA-learning algorithms; and (ii) illustrates a putative contribution of PCA-like behaviour to functional segregation in the sensory cortex.

(*a*) *PCA learning algorithms*

PCA learning algorithms are rules which govern changes in synaptic efficacy by using an associative Hebbian term in conjunction with a decay or forgetting term. The decay term minimizes the correlations among the outputs, or the connection strengths, and extracts the principal components (or some linear combination) of the input sequence. This behaviour emulates PCAs used in statistics to extract the principal dimensions or components embedded in some data. Extracting these principal components from sensory input is important from two points of view: feature detection (Oja 1989; Foldiak 1989, 1990; Rubner & Schulten 1990) and information theory (Linsker 1988; Friston *et al.* 1992).

PCA learning algorithms provide a partial solution to the problem of feature detection and perceptual categorization. The task of a recognition system has been described as dividing 'a set of high dimensional pattern vectors, such as images or sounds, into a finite number of classes', where the selection of features relies only on regularities in the input sequence (Foldiak 1989). Good features reduce dimensionality with a minimal loss of information (or maximal information transfer). Given certain assumptions, the eigenvector solution or principal components of the input covariances have these optimal properties.

(*b*) *Functional segregation*

One fundamental principle of cortical organization is functional segregation. This empirical phenomenon places constraints on any putative learning algorithms that may be implemented by the brain. PCA-like algorithms satisfy many of these constraints. The connections between cortical regions are not continuous but occur in patches or clusters. This patchiness, defined by extrinisic connections, has, in some instances, a clear relation to functional segregation. For example, V2 has a distinctive cytochrome oxidase architecture, consisting of thick stripes, thin stripes and interstripes. When anatomically identified recordings are made in V2, and then correlated with the cytochrome oxidase pattern, directionally selective (but not wavelength- or colour-selective) cells are found exclusively in thick stripes. Retrograde labelling of cells in V5 is limited to these thick stripes. All the available physiological evidence suggests that V5 is a functionally homogeneous area specialized for motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical correlate of functional segregation and specialization (see Zeki (1990) for a full discussion).

Functional segregation assembles functionally dis-

† To whom correspondence should be addressed at: The Neurosciences Institute, Suite 310, 3377 North Torrey Pines Court, La Jolla, California 92037, U.S.A.

tinct sets of signals into specialized areas, subareas and patches. This requires correlated activity in the convergent afferents which mediate this assembly. Conversely uncorrelated or orthogonal activity in divergent efferents is required to segregate and redistribute signals from a functionally heterogeneous area to a series of more functionally specialized areas. Functional segregation therefore suggests: (i) activity in convergent afferents is correlated; and (ii) activity in divergent efferents is uncorrelated or orthogonal. An anti-symmetrical arrangement, of this sort, satisfies both the requirements of functional segregation (Zeki 1990) and those predicted by the principle of maximum information transfer (Linsker 1988). See Friston *et al.* (1992) for a complete discussion. This commonsense analysis of functional segregation predicts that inputs to a small cortical region will show correlated activity. In other words, any inputs that arise (e.g. developmentally) and are not correlated with the prevalent pattern of incoming activity will be selectively eliminated.

We describe a PCA learning algorithm and: (i) develop a possible biological mechanism which depends on retrograde signalling from the terminal arbors to the cell body of the presynaptic neuron; and (ii) show that the algorithm can account for the elimination of uncorrelated inputs implied by functional segregation.

2. THEORY

(*a*) *The algorithm*

The PCA-like algorithm considered is one of the simplest and is known as Oja's 'Subspace' algorithm (Hornik & Kuan 1992):

$$\Delta Q = \zeta(Cx \cdot Q - Q \cdot Cy), \tag{1}$$

(our notation) where Q is the connection strength matrix and ΔQ its change. Cx and Cy are the input and output covariance matrices, respectively, and ζ is a constant. The algorithm has an associative Hebbian term $(Cx \cdot Q)$ and a decay term $(Q \cdot Cy)$. Computer simulations of this equation show that, with time, Q tends to have orthonormal columns which span the same subspace as the eigenvectors of Cx with the largest eigenvalues (Oja 1989). This observation simply confirms that the network is extracting the largest principal components of the inputs.

Equation (1) is related to a number of rules and architectures which show a similar behaviour. Mixed networks, combining associative Hebbian connections and decorrelating anti-Hebbian feedback (lateral) projections (Foldiak 1989, 1990; Rubner & Schulten 1990), also find the output space described by the eigenvectors and render the outputs uncorrelated. A rigorous convergence analysis of this class of learning algorithms is found in Hornik & Kuan (1992).

We propose that equation (1) may be implemented in the brain by: (i) a peripheral associative mechanism which reflects the conjunction of local pre- and postsynaptic depolarization; and (ii) a decay term that represents the local interaction between central signals, integrated over all the peripheral extensions of the pre- and postsynaptic neuron. For the postsynaptic neuron, this signal is the total anterograde influence of all presynaptic inputs to the dendritic tree. Similarly, for the presynaptic neuron, this signal would reflect the total retrograde influence of all the postsynaptic targets on the terminal arbors. This integration, in the postsynaptic neuron, could be the same as that subtending action potentials at the initial segment (e.g. electrotonic communication). In other words, the postsynaptic associative and decay terms share the same dependence on overall presynaptic input, but their intracellular mechanisms and timecourses may be very different. For the presynaptic neuron, integration over presynaptic terminals depends explicitly on retrograde axonal signals which converge on the cell body. The symmetry of these proposed associative and decay terms is apparent in figure 1, and they are expressed as:

$$\Delta Q_{ij} \propto \langle x_i\, y_j \rangle - \epsilon \langle (c_{\text{pre}} \otimes \Sigma\, Q_{ik}\, y_k) \cdot (c_{\text{post}} \otimes y_j) \rangle, \tag{2}$$

where y_j is the postsynaptic activity in unit j and is $\Sigma\, Q_{ki} x_k$; x_i is the input activity of afferent i, and Q_{ij} is the connection strength; $\langle \cdot \rangle$ is an averaging operator, $\otimes$ denotes convolution, and ϵ is a constant; c_{pre} and c_{post} are functions of time which model the delay and dispersion of weakening intracellular signals. These delay and dispersion affects may differ markedly in the pre- and postsynaptic cells. Equation (1) can be shown to be a special, but important, case of equation (2) (see Appendix 1).

(*b*) *Biological evidence for centrally mediated synaptic weakening*

Equation (2) predicts something a little counter-intuitive, namely, synaptic weakening, following postsynaptic depolarization, in the context of no presynaptic activity. This phenomenon has been observed in electrophysiological studies of the ipsi- and contralateral synapses formed by bilateral entorhinal projections to the dentate gyrus (Lopez *et al.* 1990). How is the presynaptic component of this enduring synaptic change mediated?

In synaptic systems that have been studied in detail, an individual afferent makes contact with postsynaptic cells through complex pre-terminal arborizations, ending in multiple *en passant* and terminal boutons (Burke 1987). These multiple synaptic connections provide a rich repertoire for functional modulation of each terminal (Gustafsson & Wigstrom 1988), transitions between different structural configurations (Desmond & Levy 1990) or structural remodelling of the terminal arbor (Mattson 1988). An equivalence between functional and structural plasticity is implicit. If valid, it follows that observations on axonal remodelling, development and regeneration also apply to functional plasticity. The lines of evidence supporting an equivalence between structural and functional plasticity are: (i) many intracellular systems act as both mediators of neural outgrowth and morphological responses to altered input; (ii) bidirectional transition between presynaptic terminal and growth

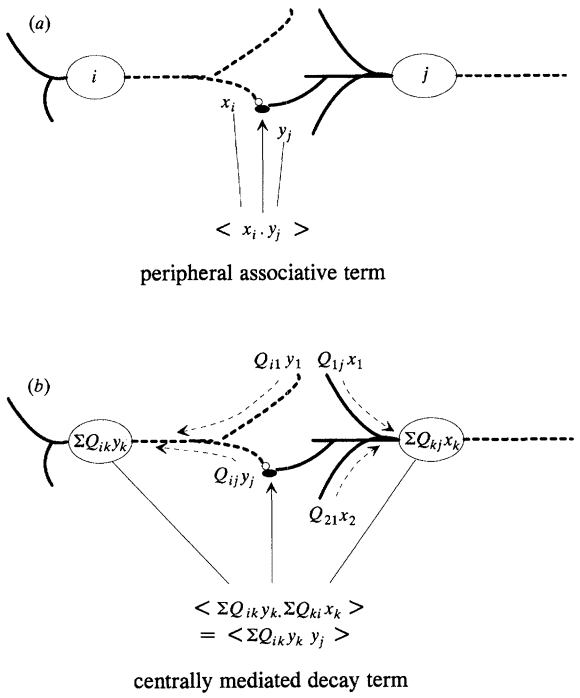


Figure 1. Schematic illustration of signalling between neurons responsible for associative strengthening of synaptic efficacy and its decay. The variables are defined in the text. Solid line, dendrites; broken lines, axons.

cone; (iii) ultrastructural neuronal changes associated with learning and memory; (iv) cytoarchitectural response to changes in sensory input (impoverished or enriched); (v) *in vivo* monitoring of neuronal cytoarchitecture; and (vi) structural responses to lesions. See Mattson (1988) for a review of this evidence in the context of neurotransmitter regulation of neuronal remodelling. Rotshenker (1988) suggests that neurons are in dynamic equilibrium with regards to axonal growth. The processes leading to neurite extension depend on growth promoting steps that involve either a stimulus for growth or a removal of inhibitory influences. It is proposed that growth-promoting mechanisms are peripheral, in the sense that they act directly on peripheral extensions (axons and terminals). Motor axons can be induced to sprout and form functional synapses while separated from the cell body (before dying). Conversely, inhibitory central mechanisms implicate the somata as sites of growth regulation. Target-derived signals may be translocated to the cell body where they exert inhibitory influences on growth. A central inhibitory mechanism is based on the following evidence: (i) motorneurons that are partly or totally deprived (with the myotoxin carbocaine) of their target muscle fibres respond with sprouting and synapse formation; (ii) axotomized frog motorneurons regenerate their axons into areas that lack any obvious source of growth factors; and (iii) developing sensory neurons innervate the skin before nerve growth factor (NGF) synthesis or synthesis of NGF receptors on the nerve (Rotshenker 1988). This, and other, evidence suggests that the cell body of the presynaptic neuron is capable of mediating a weakening of its peripheral extensions in developmental, regenerative and, by implication, functional plasticity.

The proposed interaction between associative strengthening and centrally mediated weakening results in a pleasing dialectic. Dendrites will sample afferents with highly correlated activity, whereas axonal arbors will resist driving correlated dendrites. The latter is a consequence of the weakening term. An axon strongly and multiply connected to a series of dendrites with correlated postsynaptic depolarization will be subject to a substantial weakening of its synapses. This resistance to driving correlated postsynaptic neurons prevents convergence among outputs that are connected to the same axon, and segregation can thus ensue.

3. SIMULATIONS

By using simulations we tested the hypothesis that the 'Subspace' algorithm (and implicitly all PCA algorithms) could account for the elimination of afferent connections whose signals are uncorrelated with the dominant input patterns. We present: (*a*) an idealized simulation to demonstrate this behaviour is a solid and clear way; and (*b*) a more realistic (but still simple) simulation addressing functional segregation in terms of ocular dominance.

(*a*) *An idealized simulation*

A network with 64 inputs and 8 outputs was simulated on a slow timescale by using a stationary input pattern described by Cx and using equation (1) to update the connection strengths (with $\zeta = 0.05$). Figure 1 shows the assumed covariance matrix of the inputs (Cx). The first 16 inputs were mutually orthogonal (the corresponding 16×16 subpartition was the identity matrix). The remaining inputs were substantially intercorrelated (see the figure legend for details of how Cx was generated). The first 16 inputs represent afferentation which bore no relation to the prevalent pattern of input activity. They could be thought of as afferents from wavelength-selective units in V2 which have found themselves impinging, inappropriately, on V5.

Equation (1) has the effect of driving the sum of squares of Q_{ij} over a dendritic tree to unity (cf. Theorem 3 in Hornik & Kuan (1992)) and consequently $|Qij| < 1$ for all i,j. However, this conservation does not apply to the Q_{ij} which correspond to a terminal arbor. It is possible for an input to develop Q_{ij}s which are all zero. In this case the afferent does not contribute to any output, and has been eliminated. We predicted the first 16 afferents would suffer this fate.

$|Q_{ij}|$ was interpreted as the probability of a functionally active connection between units i and j. As the Q_{ij} develop over time, different permutations of the afferents will constitute the input sampled. Each input is, at any time, functionally connected to one or more outputs, with a probability p_i, which is simply one minus the probability that input i is not connected to any outputs, $p_i = 1 - \Pi\,(1 - |Q_{ij}|)$. We examined p_i over 100 iterations, starting with a random connectivity matrix Q (elements were selected from a Gaussian distribution and Euclidean normalized

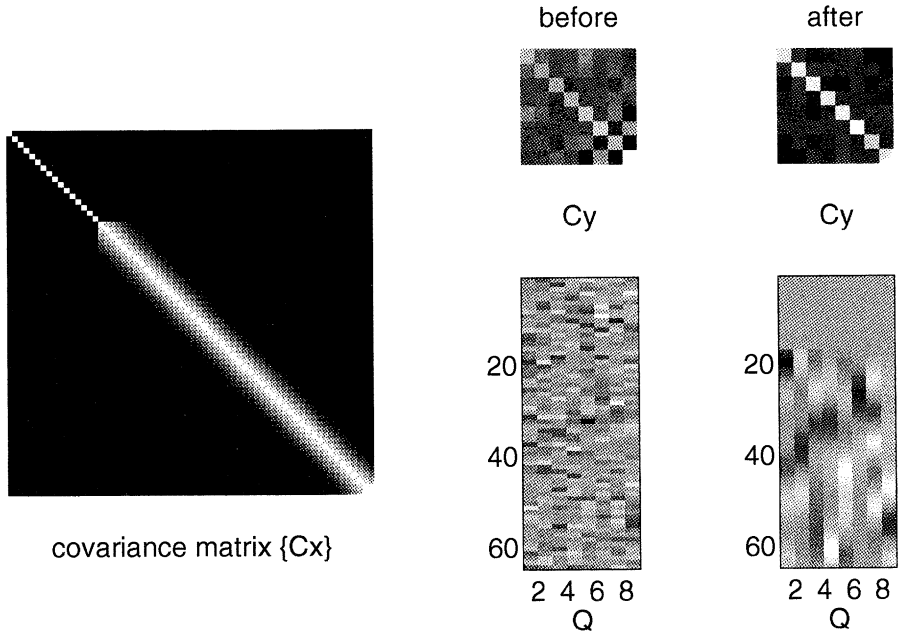


Figure 2. *Cx*: the (64 × 64) input covariance matrix *Cx* used in the first simulation. The last 48 units are substantially intercorrelated, whereas the first 16 were mutually independent (orthogonal). The 48 × 48 subpartition of *Cx* was a Toeplitz (autocorrelation) matrix of a Gaussian function of parameter 2. The 16 × 16 subpartition was the identity matrix. *Q*: connection strengths mapping the (64) inputs to the (8) outputs before and after 100 iterations. Note that the connection strengths from the first 16 inputs (top portion), to the outputs, have been eliminated. The orderly and structured connections to the remaining 48 units have segregated in such a way as to render the outputs largely uncorrelated. *Cy*: covariance matrix of the (8) outputs. At the end of the simulation the outputs were more orthogonal. The grey scale is scaled to the maximum of each matrix.

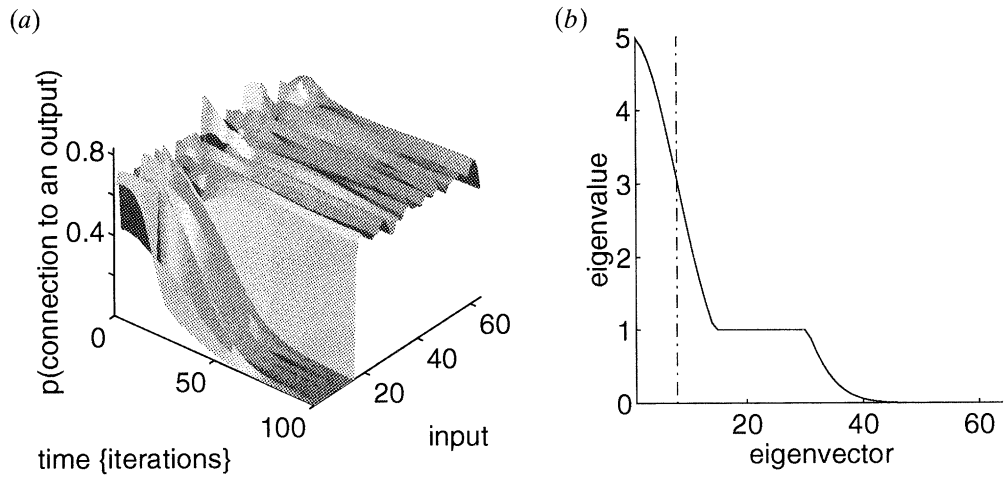


Figure 3. (*a*) The time-dependent probability (p_i) that an input is connected to one or more outputs over 100 iterations (time). In accord with figure 1, (Q—after) the last 48 inputs have been 'selected' by the outputs and the first 16 eliminated. (*b*) The eigenvalues associated with the eigenvectors of *Cx*. The broken line separates the first 8 eigenvalues from the rest. The flat portion of the curve corresponds to the eigenvector patterns due to the first 16 inputs.

($\Sigma Q_{ij}^2 = 1$, sum over i)). Simulations were done using MATLAB (MathWorks Inc., Sherborn, Massachusetts, U.S.A.).

Figure 2 shows the idealized *Cx*, the connection strengths (Q), and output covariance matrices *Cy*, before and after 100 iterations. The two effects, which bear directly on functional segregation, are evident: (i) orthogonalization of the outputs, reflected in the prominent leading diagonal of *Cy* (however, note that output decorrelation is not necessarily a convergent property of the 'Subspace' algorithm); and (ii) the elimination of afferent connections with orthogonal signals (top portion of Q). This elimination preserves specialization of the simulated region for the prevalent stimulus attribute(s). Figure 3*a* shows the probabilities (p_i) that each afferent is connected to one or more outputs. As predicted, the first 16 afferents were selectively eliminated. Figure 3*b* shows the distribution of the eigenvalues associated with *Cx*. Note that the eigenvalues corresponding to the first 16 inputs (flat portion of the curve) are smaller than the eigenvalues of the eight most prominent eigenvectors represented in the outputs (broken line). This is why they were eliminated.

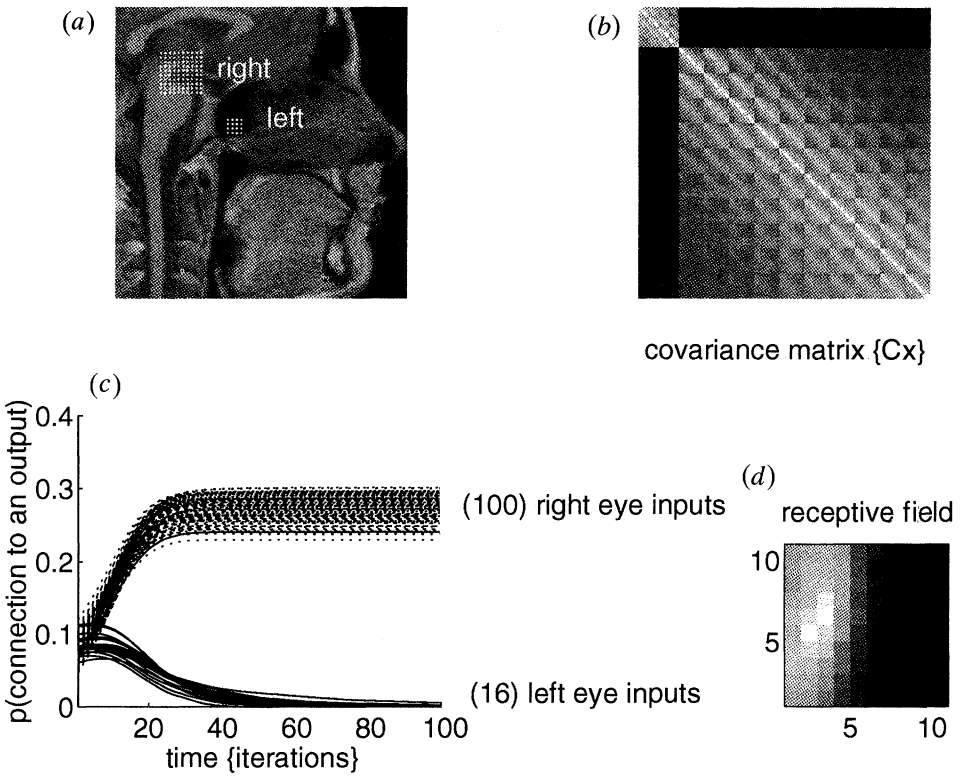


Figure 4. (*a*) The MRI images used to generate an input sequence. This 256 × 256 pixel mid-sagittal section of the human brain was sampled independently (1000 times) by two square arrays of simulated receptors spaced 1 pixel apart. The larger (10 × 10) array modelled right-eye input (right), and the smaller (4 × 4) array, left-eye input (left). (*b*) *Cx* is the covariance matrix of the input sequence thus obtained. The top left subpartition is that representing the smaller left-eye input. (*c*) The time-dependent probability (p_i) that an input is connected to one or more outputs over 100 iterations (time). Solid lines, left-eye inputs; broken lines, right-eye input. (*d*) The receptive field for one output unit after 500 iterations.

(*b*) *A more realistic simulation*

The same procedure was then applied to a covariance matrix derived from a real time series. This series was meant to represent input from two (uncoordinated) eyes, and was obtained by repeated local sampling of a picture (a magnetic resonance image (MRI) of the human brain). Most inputs modelled afferent signals from the right eye using a 10 × 10 array of simulated receptors, spaced 1 pixel apart. The left input came from a similar (4 × 4) array. Both receptor arrays were moved randomly over the picture 1000 times to create a sequence of 116 (= 10^2+4^2) inputs. The covariance matrix of this input sequence was calculated and the connection strengths to ten output units were updated for 500 iterations according to equation (1) with $\zeta = 0.002$. To make these simulations more realistic, connection strengths were not allowed to change sign. Q was a matrix of random elements selected with uniform probability in the range [0, 1] which were scaled by a factor of $1/n$, where n is the number of inputs.

The covariance matrix of the simulated visual input is seen in figure 4*b* and shows the segregation of covariance into two partitions corresponding to the two eyes. The development of ocular dominance is apparent in figure 4*c*, which shows the selective elimination of the 16 afferents from the left eye. This elimination reflects the disproportionate influence of right eye inputs, which only results from their greater number. The receptive field of a typical output unit is also shown.

4. DISCUSSION

This paper makes two separate but related points. First, by invoking a presynaptic retrograde signal in the regulation of synaptic efficacy, a biological basis for one PCA learning rule emerges. This signal relies on retrograde axonal transport from presynaptic terminals to the cell body of an afferent neuron and is distinct from (but depends on) retrograde signalling between pre- and postsynaptic sites (see, for example, Gally *et al.* 1990). Second, the resulting PCA-like behaviour can account for segregation of mixed inputs and the elimination of orthogonal or unrelated afferents predicted by functional segregation.

(*a*) *Axonal retrograde signalling*

We have derived a PCA learning algorithm (Oja's 'Subspace' Algorithm) from a synaptic modification rule that comprises a Hebbian associative term and a biologically motivated decay term. The decay term models a plastic mechanism acting at the level of the cell body. The result is two complementary, interacting and anti-symmetrical influences on synaptic connec-

tivity: (i) synaptic consolidation, which depends on the conjunction of local pre- and postsynaptic depolarization events; and (ii) synaptic weakening, which again depends on a local interaction between a pre- and postsynaptic signal but where these signals come from the cell body. These signals reflect postsynaptic or target activity 'seen' by all the presynaptic terminals of cell i and, equivalently, the presynaptic influences integrated over all the dendritic synapses on cell j. The cell body is naturally implicated as the site of this convergent integration, particularly on the presynaptic side where there is no facility for electrotonic signalling over the pre-terminal arbors.

As suggested by Rotshenker (1988), synaptic induction (consolidation) and regression (weakening) interact to render the endstate of neuronal extensions in equilibrium. It is possible that one mechanism mediating the weakening effect relies on inhibition of the expression of morpho-regulatory proteins, for example, growth associated protein, GAP-43. Support for the notion that axonal structural changes have a key role in dynamic (as opposed to developmental or regenerative) plasticity derives from the involvement of GAP-43 in long-term potentiation and synaptic regulation (Benowitz & Routtenberg 1987). GAP-43 is a rapidly transported axonal protein most prominently expressed in regenerating and developing nerves. However, the low-level persistence of GAP-43 in the adult central nervous system (CNS), where growth and regenerative capacity are minimal, may suggest a role for this molecule in neuronal remodelling (De la Monte *et al.* 1989). GAP-43 expression is strikingly high in the adult rat hippocampus, and has also been shown to be localized with monoaminergic neurons in the brain stem suggesting that 'this phosphoprotein might be involved in the functional plasticity and synaptic transmission of monoaminergic neurons' (Bendotti *et al.* 1991).

The notion that axonal growth is under retrograde, centrally (cell body) directed inhibitory control suggests that collateral axonal growth should be potentiated when that signal is attenuated by: (i) removal of postsynaptic depolarization (see above, and Rotshenker 1988); and (ii) removal of collaterals distant to the axonal extensions potentiated. There is some evidence to suggest the latter. Stanfield (1989) has reported that, if axons ascending from the locus coeruleus are cut in rats then a more widespread distribution of descending coeruleospinal neurons is retained beyond the perinatal period. These results not only suggest 'that the absence of the normally retained collateral of the locus coeruleus neuron is sufficient to prevent the elimination of a collateral which would otherwise be lost, but also may imply that the presence of the maintained collateral is somehow causally involved in the elimination of the transient collateral' (Stanfield 1989). This is clear evidence of a central (cell body) mechanism.

(*b*) *Functional segregation and PCA-like algorithms*

The elimination of trivial and uncorrelated afferentation by a PCA-like mechanism is not surprising. The important aspect of this phenomenon is its relevance to functional segregation. Functional segregation depends on divergent and convergent extrinsic cortico-cortical connections which assemble functionally related signals in specialized areas and subareas and then redistribute segregated, less-correlated signals to other areas. Intrinsic connections within some cortical regions may behave according to a PCA-like mechanism and, in so doing, segregate mixed inputs. Furthermore, by eliminating uncorrelated afferentation, the same mechanism could also help to establish appropriate patterns of extrinsic connectivity.

Extrinsic connectivity becomes committed at an early stage of CNS development (McConnell 1989; Sur *et al.* 1990). A selective elimination of synaptic connections (which would otherwise confound functional segregation) may be instrumental in the refinement and sharpening of axonal arbors during development. During development, many bifurcating axons could derive from a group of neurons (Edelman 1978). These axons may impinge in a largely arbitrary fashion on distant groups, however, only those convergent axons with non-trivial intercorrelations will be selected (not eliminated) by PCA-like mechanisms. In this way, topographic segregation of function is preserved. Note that this scenario precludes two uncorrelated sets of axonal efferents from the same group being connected to the same target group. In this way convergence and divergence of extrinsic connectivity (Zeki 1990) is ensured.

The elimination of afferents by the mechanisms modelled in this paper depend on a complicated interaction between eigenvalues of the input sequence covariance matrix, the contribution of each input to these eigenvectors and the degree of dimension reduction. Elimination ensues whenever there is a significant dimension reduction and a small subset of inputs that are statistically independent of the dominant input patterns.

It should be noted that there are many more comprehensive mechanisms proposed for the formation of patchy, ordered connectivity which take explicit account of spatial interactions by using limited lateral connectivity or lateral diffusion (see von der Malsburg & Willshaw 1979; Kohonen 1982; Miller 1992; Montague *et al.* 1991).

(*c*) *Conclusion*

We hope to have extended the biological relevance of PCA-like plasticity by proposing a simple neuronal mechanism and demonstrating a consistent relation between the phenomenological aspects of PCA-like mechanisms and functional segregation.

K.J.F. was funded by the Wellcome Trust.

APPENDIX 1

The postsynaptic response of unit $j(y_j)$ to many presynaptic inputs (x_i) is $\Sigma Q_{ij} x_i$, where $\langle x_i \rangle = 0$ (x_i can be interpreted as the deviation from mean firing

rate). Let $\gamma^x_{ij}(u)$ denote the cross-covariance, at lag u, between x_i and x_j, $\gamma^x_{ij}(u) = \langle x_i(t+u) \cdot x_j(t) \rangle$. We assume that x_i is stationary and $\gamma^x_{ij}(u) = 0$ for $u \neq 0$. This assumption can be partly relaxed (see below). Therefore

$$\gamma^y_{ij}(u) = 0, \quad \text{for} \quad u \neq 0. \tag{A 1}$$

In what follows we use a discrete time formulation and rely on some standard results derived in the frequency domain. Cross-spectral density $g^y_{ij}(\omega)$ between processes y_i and y_j is (Cox & Miller 1980):

$$g^y_{ij}(\omega) = 1/2\pi \Sigma \gamma^y_{ij}(u)\, e^{-i\omega u}, \tag{A 2}$$

$$= \gamma^y_{ij}(0)/2\pi \quad \text{(from A 1)}. \tag{A 3}$$

Changes in connection strength Q_{ij} are effected by an associative term and a centrally mediated decay term (see text):

$$\Delta Q_{ij} \propto \langle x_i\, y_j \rangle - \epsilon \langle c_{\text{pre}} \otimes \Sigma Q_{ik}\, y_k) \cdot (c_{\text{post}} \otimes y_j) \rangle, \tag{A 4}$$

where $\otimes$ denotes convolution; c_{pre} and c_{post} are kernels which account for the delay and dispersion associated with convergence of intracellular signals, onto the cell body, and redistribution to peripheral extensions. Let c_{pre} have an associated transfer function $\lambda_{\text{pre}}(\omega)$:

$$\lambda_{\text{pre}}(\omega) = \Sigma c_{\text{pre}}(t) \cdot e^{-i\omega t}, \tag{A 5}$$

and similarly for c_{post}. If $y'_k = c_{\text{pre}} \otimes y_k$ and $y'_j = c_{\text{post}} \otimes y_j$ then equation (A 4) can be written more compactly:

$$\Delta Q_{ij} \propto \Sigma Q_{kj} \gamma^x_{ik}(0) - \epsilon \Sigma Q_{ik} \gamma^{y'}_{kj}(0), \tag{A 6}$$

where

$$\gamma^{y'}_{ij}(0) = \int g^{y'}_{ij}(\omega)\, \mathrm{d}\omega = \int \lambda_{\text{pre}}(\omega) \cdot \lambda_{\text{post}}(\omega)^* \cdot g^y_{ij}(\omega)\, \mathrm{d}\omega$$

$$= \gamma^y_{ij}(0)/(2\pi) \cdot \int \lambda_{\text{pre}}(\omega) \cdot \lambda_{\text{post}}(\omega)^*\, \mathrm{d}\omega. \tag{A 7}$$

If we let, for simplicity, $1/\epsilon$ be equal to the integral in equation (A 7) (they are both constants), then the expression for change in Q_{ij} is

$$\Delta Q_{ij} \propto \Sigma Q_{kj} \gamma^x_{ik}(0) - \Sigma Q_{ik} \gamma^y_{kj}(0), \tag{A 8}$$

or, in matrix notation:

$$\Delta Q \propto Cx \cdot Q - Q \cdot Cy. \tag{A 9}$$

This is the 'Subspace' logarithm. The main point made here is that delay and dispersion do not affect the relative covariances between two multidimensional processes if the cross-spectral density is distributed uniformly over all frequencies (equivalently, if the cross- and auto-covariance functions decay very quickly). Furthermore, the relative cross-covariances are preserved even if the two delay and dispersions are different for each process. The amount of cross-covariance that remains after convolution depends on the 'overlap' between the two convolution kernals (c_{pre} and c_{post}). This dependency is expressed in frequency space in equation (A 7).

All these observations hold for any two processes whose cross-spectral densities can be factorized into a term which depends only on $\{i,j\}$ and a second frequency-dependent term which does not depend on $\{i,j\}$. An analysis of the general case of arbitrary $\gamma^x_{ij}(u)$ will be presented in a further paper.

Finally, it should be noted that the arguments presented here could also apply to the associative term in equation (A 4). For simplicity we have assumed that the signalling which mediates Hebbian plasticity is sufficiently fast that the associated convolution functions can be modelled as δ functions. In this case, the associative terms reduces to its usual form ($\langle x_i\, y_j \rangle$).

REFERENCES

Bendotti, C., Servadio, A. & Samanin, R. 1991 Distribution of GAP-43 mRNA in the brain stem of adult rats as evidenced by in situ hybridization: localization within monoaminergic neurons. *J. Neurosci.* **11**(3), 600–607.

Benowitz, I. I. & Routtenberg, A. 1987 A membrane phosphoprotein associated with neural development, axonal regeneration, phospholipid metabolism, and synaptic plasticity. *Trends Neurosci.* **10**, 527–532.

Burke, R. E. 1987 Synaptic efficacy and the control of neuronal input–output relations. *Trends Neurosci.* **10**, 42–45.

Cox, D. R. & Miller, H. D. 1980 *The theory of stochastic processes*. London: Chapman and Hall.

De la Monte, S. M., Federoff, H. J., Ng, S. C., Grabczyk, E. & Fishman, M. C. 1989 GAP-43 gene expression during development: persistence in a distinct set of neurons in the mature central nervous system. *Brain Res. Dev. Brain Res.* **46**, 161–168.

Desmond, N. L. & Levy, W. B. 1990 Morphological correlates of long term potentiation imply the modification of existing synapses, not synaptogenesis, in the hippocampal dentate gyrus. *Synapse* **5**, 139–143.

Edelman, G. M. 1978 Group selection and phasic reentrant signalling: a theory of higher brain function. In *The mindful brain* (ed. G. M. Edelman & V. B. Mountcastle), pp. 55–100. Cambridge, Massachusetts: MIT Press.

Foldiak, P. 1989 Adaptive network for optimal linear feature extraction. In IEEE/INNS Joint Conference on Neural Networks, pp. 401–405. New York: IEEE Press.

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

Friston, K. J., Frith, C. D., Passingham, R. E., Dolan, R. J., Liddle, P. F. & Frackowiak, R. S. J. 1992 Entropy and cortical activity: Information theory and PET findings. *Cerebr. Cortex* **2**, 259–267.

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

Gustafsson, B. & Wigstrom, H. 1988 Physiological mechanisms underlying long term potentiation. *Trends Neurosci.* **11**, 156–162.

Hertz, J., Krogh, A. & Palmer, R. G. 1991 *Introduction to the theory of neural computation*. Addison-Wesley.

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

Kohonen, T. 1982 Self organized formation of topologically correct feature maps. *Biol. Cyber.* **43**, 59–69.

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

Lopez, H. S., Burger, B., Dickstein, R., Desmond, N. L. &

Levy, W. B. 1990 Associative synaptic potentiation and depression: Quantification of dissociable modifications in the hippocampal dentate gyrus favors a particular class of synaptic modification equations. *Synapse* **5**, 33–47.

Mattson, M. P. 1988 Neurotransmitters in the regulation of neuronal cytoarchitecture. *Brain Res. Rev.* **13**, 179–212.

McConnell, S. K. 1989 The determination of neuronal fate in the cerebral cortex. *Trends Neurosci.* **12**, 342–349.

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

Montague, P. R., Gally, J. A. & Edelman, G. M. 1991 Spatial signalling in the development and function of neural connections. *Cerebr. Cortex* **1**, 199–220.

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

Rotshenker, S. 1988 Multiple nodes and sites for the induction of axonal growth. *Trends Neurosci.* **11**, 363–366.

Rubner, J. & Schulten, K. 1990 Development of feature detectors by self organization: A network model. *Biol. Cyber.* **62**, 193–199.

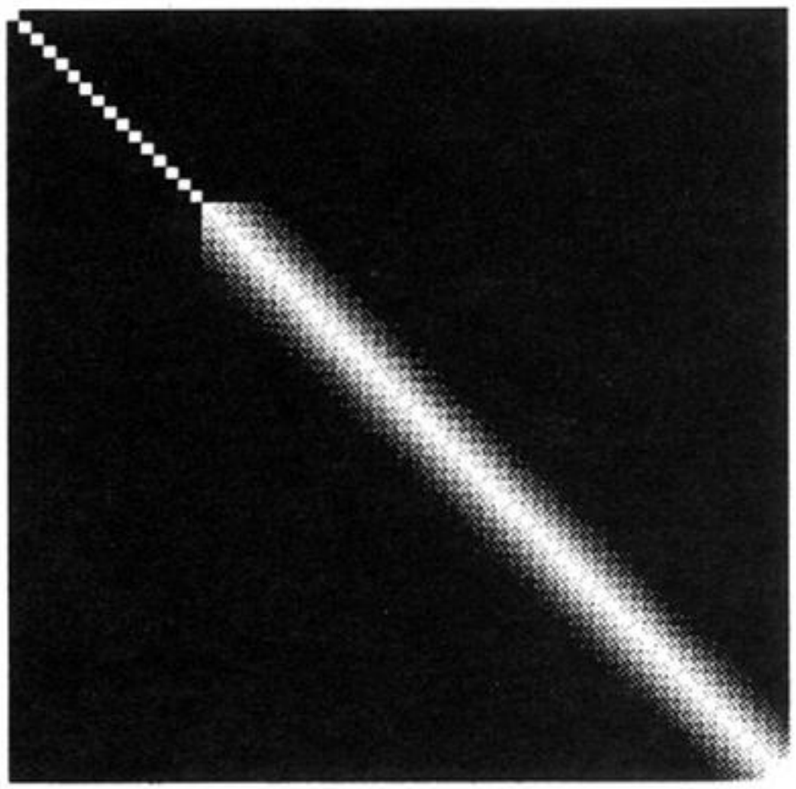
Stanfield, B. B. 1989 Evidence that dorsal locus coeruleus neurons can maintain their spinal cord projections following neonatal transection of the dorsal adrenergic bundle in rats. *Expl Brain Res.* **78**, 533–538.

Sur, M., Pallas, S. L. & Roe, A. W. 1990 Cross-modal plasticity in cortical development: differentiation and specification of sensory cortex. *Trends Neurosci.* **13**, 227–232.

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

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

Received 29 April 1993; accepted 30 July 1993



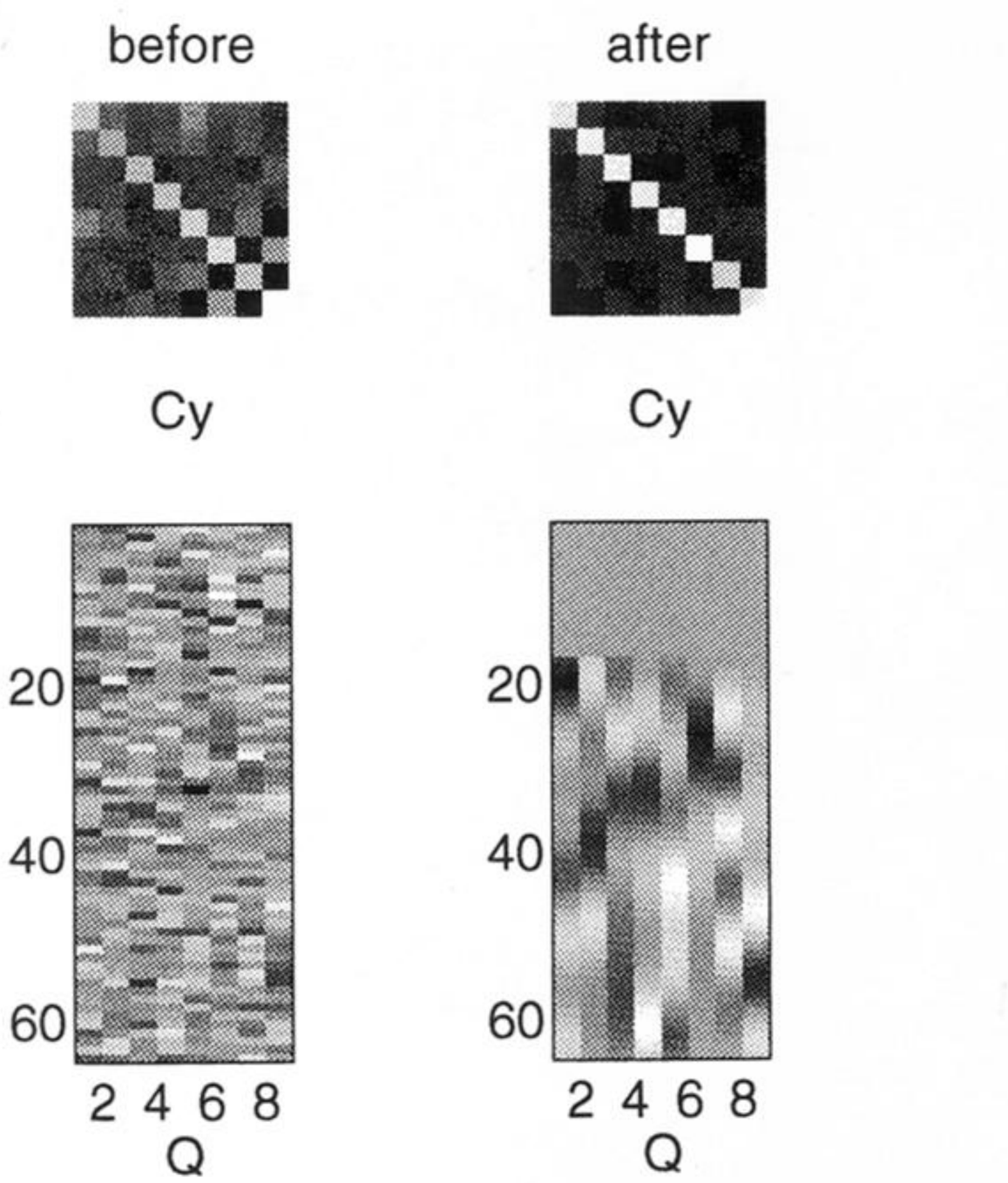


Figure 2. *Cx*: the (64 × 64) input covariance matrix *Cx* used in the first simulation. The last 48 units are substantially intercorrelated, whereas the first 16 were mutually independent (orthogonal). The 48 × 48 subpartition of *Cx* was a Toeplitz (autocorrelation) matrix of a Gaussian function of parameter 2. The 16 × 16 subpartition was the identity matrix. *Q*: connection strengths mapping the (64) inputs to the (8) outputs before and after 100 iterations. Note that the connection strengths from the first 16 inputs (top portion), to the outputs, have been eliminated. The orderly and structured connections to the remaining 48 units have segregated in such a way as to render the outputs largely uncorrelated. *Cy*: covariance matrix of the (8) outputs. At the end of the simulation the outputs were more orthogonal. The grey scale is scaled to the maximum of each matrix.

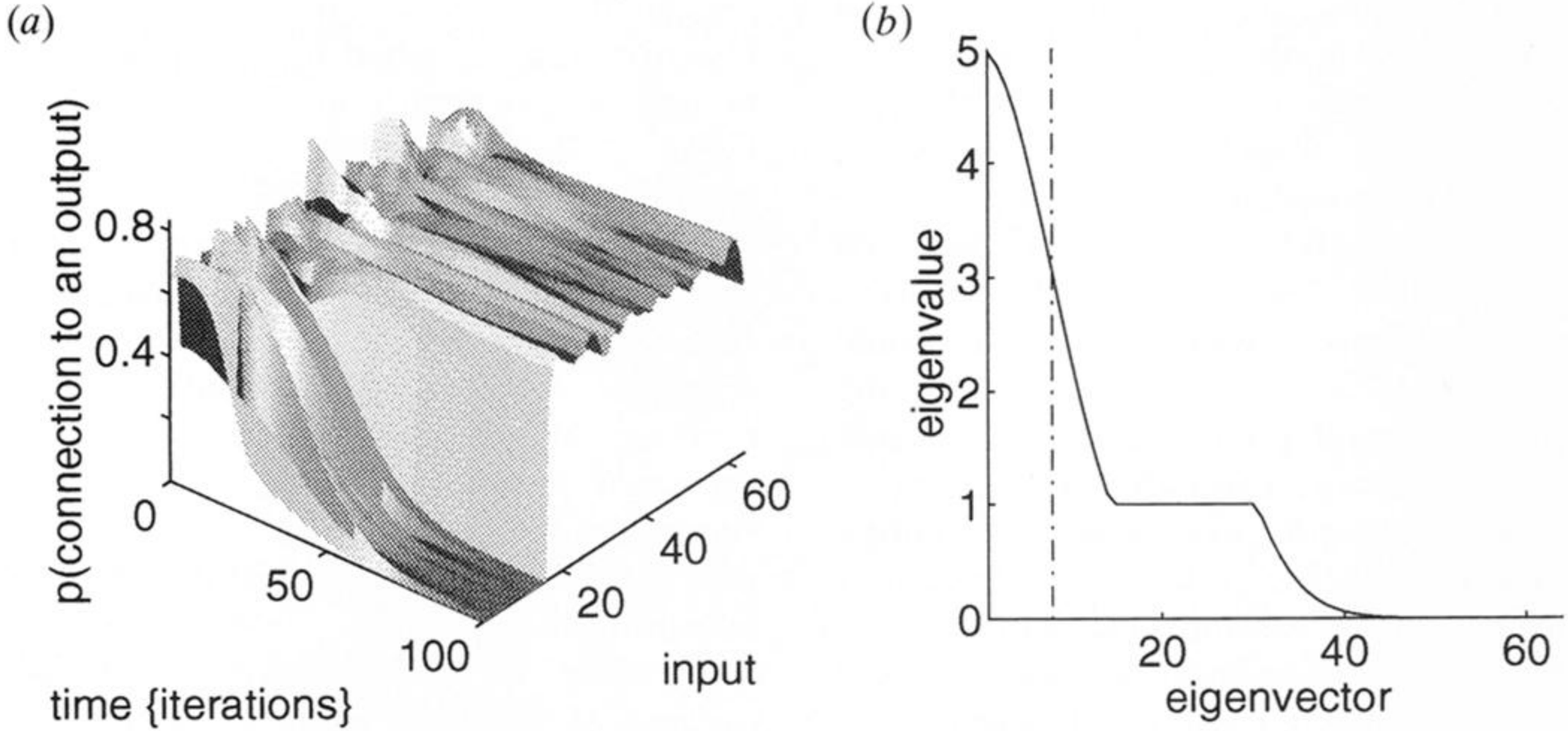
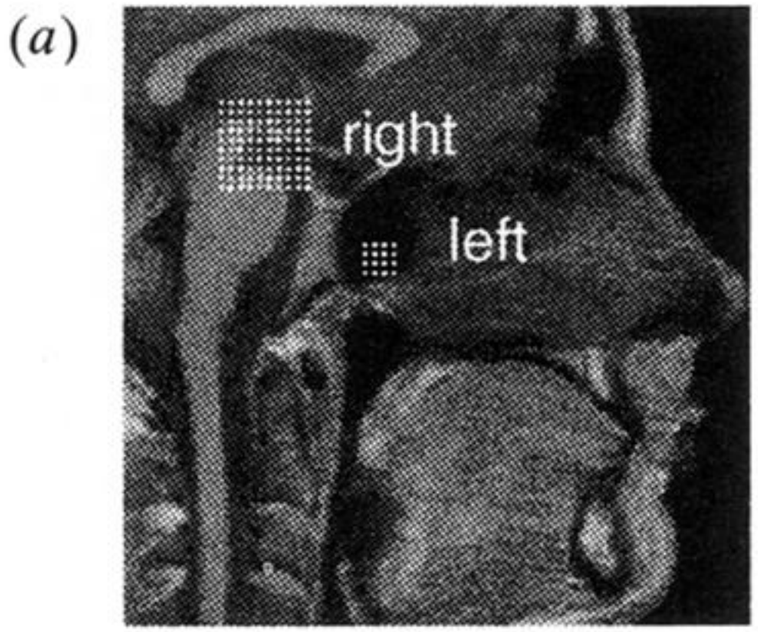
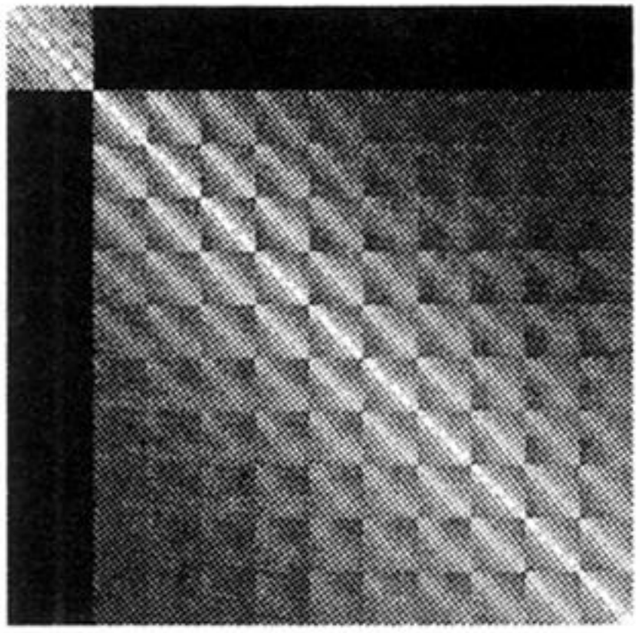


Figure 3. (*a*) The time-dependent probability (p_i) that an input is connected to one or more outputs over 100 iterations (time). In accord with figure 1, ($Q-$after) the last 48 inputs have been 'selected' by the outputs and the first 16 eliminated. (*b*) The eigenvalues associated with the eigenvectors of Cx. The broken line separates the first 8 eigenvalues from the rest. The flat portion of the curve corresponds to the eigenvector patterns due to the first 16 inputs.





covariance matrix {Cx}

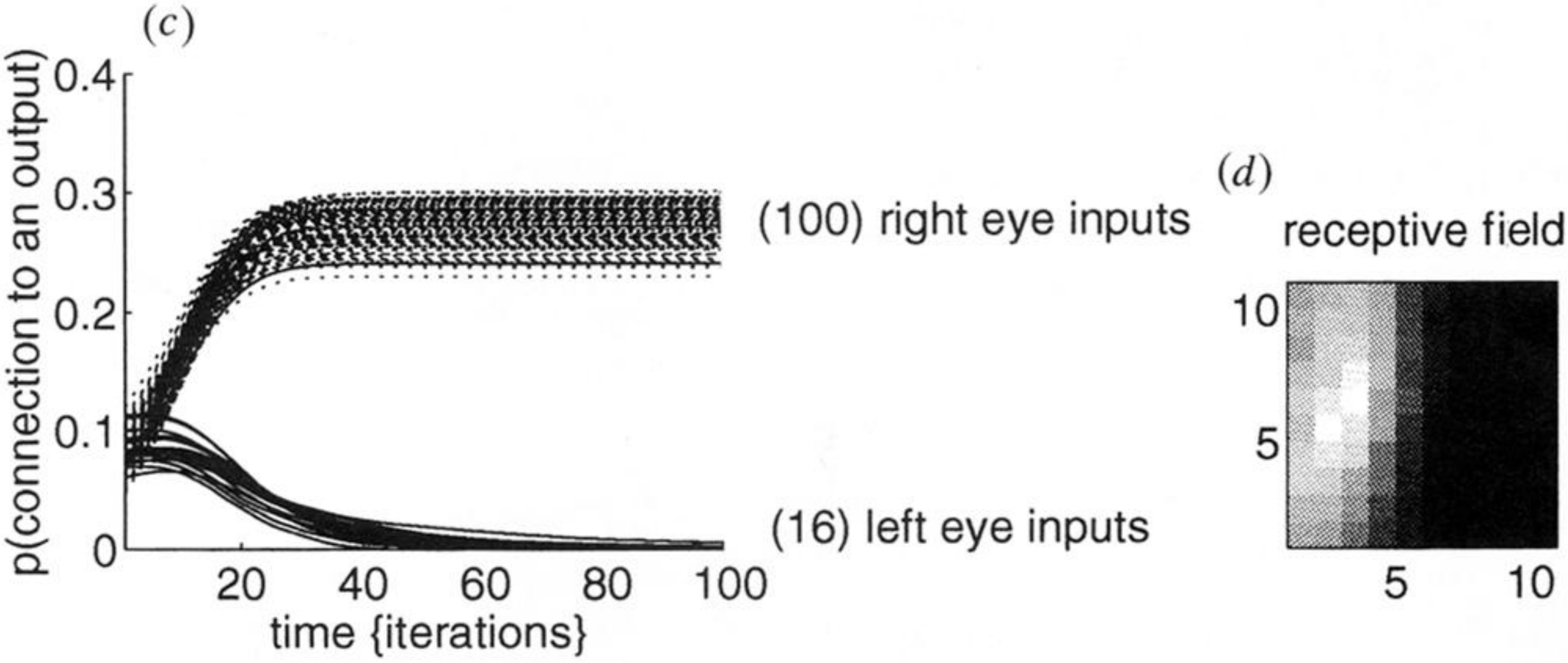


Figure 4. (*a*) The MRI images used to generate an input sequence. This 256×256 pixel mid-sagittal section of the human brain was sampled independently (1000 times) by two square arrays of simulated receptors spaced 1 pixel apart. The larger (10×10) array modelled right-eye input (right), and the smaller (4×4) array, left-eye input (left). (*b*) *Cx* is the covariance matrix of the input sequence thus obtained. The top left subpartition is that representing the smaller left-eye input. (*c*) The time-dependent probability (p_i) that an input is connected to one or more outputs over 100 iterations (time). Solid lines, left-eye inputs; broken lines, right-eye input. (*d*) The receptive field for one output unit after 500 iterations.