Dynamic changes in receptive fields induced by cortical reorganization

Germán Mato and Néstor Parga
Departamento de Física Teórica C-XI
Universidad Autónoma de Madrid
28049 Madrid, Spain

Abstract

In this work we study the dynamical changes undergone by the receptive field of cortex cells after adaptation to stimuli with an artificial scotoma. In the model, the population of neurons in a given layer of a given column of the visual cortex is represented by a single effective unit. A two-layer architecture is used and the values of the horizontal connections in the output layer are updated in such a way that the redundancy between the activities of different effective units is minimized. This hypothesis is implemented using an algorithm with a simultaneous dynamics of the activities of the effective units and for their effective horizontal connections. The model predicts several experimental facts such as the expansion of the receptive fields and the bias in feature localization experiments. It also allows to make predictions on the *filling-in* phenomenon and on its relation to the expansion of receptive fields. High order terms in the updating rule of the effective horizontal connections, which we interpret as due to the effect of inhibitory interneurons, are essential to account quantitatively for the experimental results.

1.1 Introduction

It has been experimentally observed that the receptive fields (RF) of cortical cells have a dynamical nature. For instance, it was found that some time (of the order of minutes) after the occurrence of a retinal lesion the area of the RF increased by a factor of order 5 (Gilbert and Wiesel 1992), and that cortical cells with their classical RF inside the damaged region recovered their activity. A similar effect can be obtained without the existence of real lesions. Stimuli can emulate the lesion if they are localized; that is, if there is some small part of input space that receives stimulation strongly different from their surround. Lack of stimulation in a small region of the visual space produces an effect similar to a scotoma. Experiments with localized stimuli have been done in both the visual (Pettet and Gilbert 1992) and the somatosensory systems (Jenkins et al 1990).

These changes in the RFs of cortical neurons can be quantitatively studied with psychophysical experiments. For instance, changes in RF sizes are reflected in a systematic bias in feature localization tasks. It has been found (Kapadia, Gilbert and Westheimer 1994) that the ability to determine the relative position of a short line segment in the middle of another two, presented close to the border of the artificial scotoma, was strongly biased in a way that is consistent with the expansion of RFs of neurons in the cortical scotoma.

It has been speculated (Pettet and Gilbert 1992), (Gilbert 1992) that the expansion of RF sizes is responsible for the perceptual *filling-in* effect (Ramachandran and Gregory 1991) and other visual illusions. In perceptual *filling-in* a stimulus with an artificial scotoma is presented and it is observed that after a few seconds the central region is filled by the surrounding pattern. A related phenomenon, more similar to a permanent lesion, is the fact that under normal conditions the blind spot on the retina is not perceived but it also appears filled by the surrounding part of the visual scene. Experimental evidence suggests that the anatomical substrate of these changes is the cortex (Gilbert and Wiesel 1992), (Darian-Smith and Gilbert 1995).

Most of the true long range connections link pyramidal neurons in different columns. However some of these are disynaptic connections with inhibitory interneurons. Even if these neurons are only about a 20% of the total (McGuire, Gilbert, Rivlin and Wiesel 1991) its effect is not negligible. In particular a given stimulus can produce a facilitatory or suppresive effect depending on the way that it affects the inhibititory interneurons. Contextual dependence of the RF of neurons in V1 has been observed in other physiological (Hirsch, and Gilbert 1991), (Knierim and van Essen 1992), (Grinvald, Lieke, Frostig and Hildesheim 1994) and psychophysical works (Polat and Sagi 1993), (Polat and Sagi 1994), (Kapadia, Ito, Gilbert and Westheimer 1995) and in some cases the compatibility between the results of both approaches has been verified as is discussed in (Gilbert, Das, Ito, Kapadia and Westheimer 1996). The conclusion is that complex, strong stimuli tend to suppress the response of a neuron to its preferred stimulus while a weak stimulus tends to facilitate it. See also (Stemmler, Usher and Niebur 1995).

It is then justified to consider the behavior of *effective* long-range connections that contain the effect of inhibitory interneurons. Their dynamics under changes in the stimulus should be modeled such that they become effectively excitatory when the context facilitates the response and effectively inhibitory when it tends to suppress it. Here we describe a model for the dynamics of *effective* horizontal connections than can reproduce the observed dynamics of the RFs (Mato and Parga 1996). Neuronal activity is also defined in terms of effective units describing the state of a population of neurons in the same column. The model provides a coarse grained description of the visual cortex where the observed changes between facilitation and suppression according to the context can appear naturally.

The model explains correctly the expansion of the RFs under localized stimuli, the activity changes of neurons in the cortical scotoma, the recovery of the RFs to normal size once the artificial scotoma is taken away, and the perceptual shift in the psychophysical experiment done by (Kapadia, Gilbert and Westheimer 1994). It also predicts a relation between the expansion of the RFs and the appearance of a physiological *filling-in* effect in the same cortical area in a way which is consistent with (Pettet and Gilbert 1992).

The model can be interpreted in terms of information theory. The reason is that the RFs of neurons in the cortical scotoma must change because inactive neurons do not participate in the construction of an efficient code of the current statistical properties of the environment. ¿From this perspective the simultaneous dynamics of the neuronal activity and of the effective horizontal connections must implement an adaptive algorithm that suppresses the redundancy in the cortical code. This is precisely a possible interpretation of the dynamics of the effective horizontal connections as is used in the model.

The layout of the paper is as follows. In the next section we introduce the model, presenting the network architecture, the activity and effective synapses dynamics, and the statistics of the inputs. In section 1.3 we give a qualitative discussion of the model. In section 1.4 we show the results concerning the changes of RF sizes, their reversibility, the psychophysical shifts and the physiological *filling-in* effect. The last section contains the conclusions.

1.2 The Model

1.2.1 Architecture

We use a two-layered neural network. The input layer represents layer 4 of V1 while the output can be identified with the upper cortical layers 2+3 in the same column. Both layers consist of a square array of N^2 units. We are assuming that each unit represents a population of excitatory neurons in the same column and cortical layer. Our model has to be interpreted as a coarse grained description of V1 where the activity of each unit represents the average activity of the population of neurons that it describes. From now on we will refer to these populations as units or cells.

The activity of the input unit j is denoted by v_j and the state of the unit i in the second layer by u_i. The feedforward connections from the input to the output layer are denoted by J_{ij}. These connections are fixed and perform a topographic mapping between the two layers. The divergent connections from a unit in the input layer take their maximum value when they reach a unit in the equivalent position in the second layer. Their strengths decay according to a Gaussian function of the distance $d(i,j)$ between the input and the output units:

$$J_{ij} = \frac{1}{2\pi\sigma_f^2} e^{-d(i,j)^2/2\sigma_f^2}. \tag{1.1}$$

The dispersion σ_f describes the spread of the feedforward connections. $d(i,j)$ is measured in terms of the number of lattice spacings between cells. The first factor was chosen in such a way that the total strength of the divergent connections is one (for each input unit) in the limit of a dense input layer. The horizontal connections between neurons i and k in the second layer will be constructed as the product $D_{ik}T_{ik}$ of a truly dynamical part T_{ik} and a static factor D_{ik}. The first connects all possible pairs of cortical units while the static piece effectively limits the range of the horizontal connections. The D_{ik}'s are chosen once and for all as a gaussian function (with a variance σ_h) of the distance between the units i and k. Again distances between units in this layer are measured in terms of the number of lattice spacings between cells. Both quantitities, σ_f and σ_h, are relevant for the size of the spike and optical point spreads. According to (Das and Gilbert 1995) the area of the spike PS is about 5% of the optical PS. This represents a value of σ_h larger than σ_f by roughly a factor 2.

Both feedforward and horizontal connections contribute to define the classical RF. The difference between the classical and the dynamic RFs is that the latter results from adaptation to persistent features in the stimuli. The problem here is how to choose the ensemble of visual stimuli that defines "normal" visual conditions and that determines the classical RF. This point will be made more precise in section (1.2.3) where the statistics of the inputs is discussed.

1.2.2 Dynamics

The dynamics for the activity of units in the second layer is taken as:

$$\tau \frac{du_i}{dt} = -u_i + H\left(\sum_{j=1}^{N^2} J_{ij}v_j - \sum_{k=1}^{N^2} D_{ik}T_{ik}u_k\right), \tag{1.2}$$

where v_j denotes the strength of the stimulus on the cell j in the input layer and u_k is the activity of the unit k in the output layer. Let us observe that as D_{ik} is always positive, a positive value of T_{ik} means an inhibitory interaction. The parameter τ is a time constant of the order of the membrane time constant and H is an activation function that has the form

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } \theta > x \geq 0 \\ \theta & \text{if } x \geq \theta \end{cases} \tag{1.3}$$

Saturation is introduced to prevent the effective horizontal connections from producing a divergence of the network activity. As we shall see, this can happen only for certain values of the parameters of the updating rule for the J_{ik}'s. This point will be discussed in subsection 1.4.1.

The time constant τ is expected to be of order 10 msec. This is faster than the other time constants present in the problem (e.g. those related to the modification of the effective synapses and to the presentation of the visual stimuli). Therefore we can assume that the system is always at the fixed point of eq.(1.2):

$$u_i = H\left(\sum_{j=1}^{N^2} J_{ij} v_j - \sum_{k=1}^{N^2} D_{ik} T_{ik} u_k\right). \tag{1.4}$$

Now we need to define the mean value of a generic function f of the firing rates in a time window of size T

$$< f >_T (t) = \frac{1}{T} \int_{-\infty}^{t} f(t') e^{(t'-t)/T} dt'. \tag{1.5}$$

The instantaneous separation from this mean will be denoted by:

$$\delta_T f(t) = f(t) - < f >_T (t). \tag{1.6}$$

The effective horizontal connections will be updated according to the equation:

$$\tau_s \frac{dT_{ik}}{dt} = \alpha \delta_T u_i \delta_T u_k + \beta \delta_T u_i^2 \delta_T u_k + \gamma \delta_T u_i \delta_T u_k^2, \tag{1.7}$$

where $i \neq k$. In order to fix the scale of the outputs we must chose the value of N^2 weights. We chose to keep constant the N^2 self-interaction terms fixing $D_{ii} T_{ii} = J_{ii} - 1$. This corresponds to a scaling factor of 1 between outputs and inputs for a transfer function $H(x) = x$.

Let us remark that this updating rule has two time constants. One is given by the averaging time window size T and the other by the typical value of the constants α, β and γ (that has been absorbed in τ_s). Apart from the usual quadratic term this updating rule contains higher order contributions. These can be introduced phenomenologically by arguing that they represent the effect of interneurons in the effective synapses. We want to emphasize that since we are dealing with a coarse grained model of the cortex these terms should not be interpreted (exclusively) as changes of the real synapses.

Even if the stimuli v_j were uncorrelated, the activity of the units in the second layer will be correlated because of the spread of the feed-forward connections. Eq.(1.7) is designed to eliminate correlations of any order. It implements an algorithm proposed by Jutten et al. to perform independent component analysis (Hérault and Jutten 1994); this means that (hopefully) it has as a fixed point a solution for the T_{ik}'s such that the cortical units retrieve the statistically independent components of a signal (i.e. of the stimulus v_j). It has been proved (Nadal and Parga 1996) that if this rule does find a fixed point such that each of the terms on the r.h.s. of eq.(1.7) is zero then it has found the independent components of the input.

This fixed point gives an efficient code of the typical stimuli because all redundancy has disappeared. This solution would not be reasonable at the neuron level where one could argue, for instance, that since the system is noisy a good cortical code would need some degree of redundancy to compensate for the noise. What makes this optimal solution acceptable is the fact that the units used in the model represent populations of the true neurons located in the same column. Let us also remark that eq.(1.7) could also converge to a situation where its r.h.s. is zero globally. Because of this we will explicitly check the kind of fixed point where it arrives.

1.2.3 Statistics of the inputs

An important point to determine the RFs is the description of the visual stimuli. The equations above give the evolution of cortical activity and synaptic weights in terms of the inputs v_j and they will undergo strong changes each time their statistics changes. The system will be presented with two different image ensembles. The first will represent normal visual stimuli, selected with the second order statistics obtained experimentally for ensembles of natural images, i.e., a power spectrum $S(f)$ given by:

$$S(f) \propto |f|^{\eta-2} \tag{1.8}$$

where η is a small number. This power spectrum was found in studies of the statistics of natural images (Field 1987). The precise value of η depends on the particular set of images, in (Ruderman 1994) it was found to be $\eta \approx 0.2$. This describes the conditions before the lesion, and it will give rise to classical RFs. The second ensemble will represent images with an artificial scotoma. These images consist of uncorrelated inputs (i.e. random dots) with a given mean luminosity and a uniform central region of intensity equal to that mean.

1.3 Discussion of the model

We discuss here some qualitative predictions from the dynamics of the effective long range connections. A qualitative analysis can help us to identify the influence of the different parameters on the dynamics of the RFs.

After the system has adapted to the statistics of normal visual conditions the *effective* horizontal connections are inhibitory, with a range σ_h. This is because the rule in eq.(1.7) forces the output units to detect different features of the inputs. When stimuli with an artificial damage are shown the activity of a unit inside the cortical scotoma will fall to zero because the input is absent, the horizontal connections give an inhibitory contribution and the activation function gives zero output for any negative argument. A unit with its RF only partially inside the scotoma will decrease its activity in proportion to the fraction of its RF inside the scotoma. Let us call i (k) the index of a unit totally inside (partially outside) the scotoma. Initially the modification of the weight between units k and i will be, approximately:

$$\tau_s \frac{dT_{ik}}{dt} \approx -\alpha < u_i >_T \delta_T u_k + \beta < u_i >_T^2 \delta_T u_k - \gamma < u_i >_T \delta_T u_k^2 \qquad (1.9)$$

Since the introduction of the scotoma decreases the activity of the unit k, $\delta_T u_k$ will be initially smaller than zero. This means that the first term will have the sign of α, while the second and the third will have signs opposite to β and γ respectively. Therefore a positive α will (at least at the beginning) make the connections even more *inhibitory.* This instead of expanding the RF will tend to contract it. The opposite effect is found for β and γ. This analysis suggests that one should search in a region with small positive or even negative α and positive β and γ.

We can immediately see that the averaging time T is essential for the expansion of the RF. If it were zero or very small the average activity of unit

i, $< u_i >_T$, would go very fast to zero and there would be no modification of the connections impinging on this unit. The necessity of having some kind of memory effect was also noticed in other problems by several authors. Clothiaux et al. (Clothiaux, Bear and Cooper 1991) use a memory effect in the context of synaptic changes of feedforward connections based on the BCM model (Bienenstock, Cooper and Munro 1982) and it is also a relevant ingredient in some models for invariant object recognition (Foldiak 1991), (Wallis and Rolls 1996).

We can expect that the expansion of the RFs will increase as T increases, although a quantitative analysis requires numerical simulations. On the other hand a high value of T will have the effect of making the RFs (and the psychophysical shifts) to grow more slowly, because it will take a longer time to "forget" the state of the system previous to the presentation of the scotoma. This effect can be partially compensated by decreasing τ_s.

1.3.1 Implementation of the algorithm

The inputs were generated according to the statistics defined in the previous section. A new input was presented every 10 time steps, both for normal visual stimuli and for the artificial scotoma. This fixes the time scale unit as the number of presented patterns and allows the comparison with psychophysical experiments. For instance the dynamic background of random dots used with the artificial scotoma in (Kapadia, Gilbert and Westheimer 1994) was refreshed at a rate of 30 Hz, this means that one time unit expressed in physical units is about 33 msec. The time scale T (used in the evaluation of activity averages) will be also expressed in the same units. For instance, $T = 50$ means that averages will be evaluated in a time window corresponding to the presentation of 50 patterns or, again in the case of (Kapadia, Gilbert and Westheimer 1994), about 1.6 sec. The numerical results presented in the next section were obtained by integrating eq.(1.7) using a first order Euler integration scheme. The time step was set at $\delta t = 0.1$. We checked that using smaller time steps (keeping the total integration time fixed) or a fourth order Runge-Kutta integration scheme does not change substantially the results. The mean value of the inputs v_j in the stimuli with scotoma was chosen to be 1.2 and the dispersion 0.3 (except in the scotoma where the dispersion is 0). The stimuli with the statistics of the natural images have the same mean value. The saturation value of the transfer function H was chosen as $\theta = 2$. With this choice, the internal field $\sum_{j=1}^{N^2} J_{ij} v_j - \sum_{k=1}^{N^2} D_{ik} T_{ik} u_k$ is typically in the linear region of the activation function. Let us remark that an arbitrary scaling factor in the values of u_i can always be introduced by rescaling the inputs and the constants α, β and γ in a suitable way. For this reason the interpretation of u_i in terms of firing rates is quite arbitrary.

Most of the simulations have been performed in systems with 16×16 units in each layer. The artificial scotoma was taken as a square of 7×7 units. The range of the connections are chosen as $\sigma_f = 2$ and $\sigma_h = 4$. The spread of the feedforward connections is taken to be smaller than that of the horizontal ones because thalamic connections have a shorter range than intracortical connections (Gilbert and Wiesel 1992) and also to represent properly the relative size of the spike and optical PS. The absolute value of the spreads is chosen in such a way that the classical RF are smaller than the size of the artificial scotoma by a factor approximately 2. These relative sizes are quite compatible with those used in some of the experiments (Pettet and Gilbert 1992).

1.3.2 Simulation of the experiments

The simulations are composed of several operations that depend on the experiment one wants to simulate:

- First the network has to be put in a state corresponding to normal visual conditions, as defined before. To do this the horizontal connections are first initialized to zero (except the self-interaction terms) and stimuli generated with the statistics corresponding to eq.(1.8), are presented at the input layer. Horizontal connections are then updated according to eq.(1.7). We check convergence and the nature of the fixed point by measuring activity correlations up to fourth order as a function of time.

- Once the network is in its normal state we evaluate the (classical) RF of some of the output units by presenting a point stimulus on the input layer at several positions and simultaneously measuring the response of the selected unit on the second layer.

- In another operation, performed after the network converged under normal conditions, the inputs display the statistics of images with an artificial scotoma. The change of ensemble induces a sudden increase in the correlations, that should decrease again to zero as the algorithm converges. In this case we are interested in the dynamics itself of the network.

- We will simulate the experiment of (Kapadia, Gilbert and Westheimer 1994) at several intermediate steps. This means that taking the current state of the network a stimulus made of three small horizontal bars is presented and the response of cortical units is measured. This operation is repeated placing the stimulus at different relative positions with respect to the scotoma.

- We can also evaluate the RF as a function of time by presenting again a point stimulus at different places on the input layer, but in particular we are interested in the RF when a stationary state is reached. This should give an expansion of the RF that can be compared with (Pettet and Gilbert 1992).

1.4 Results

1.4.1 Changes of receptive field size

We first show the results obtained when the parameters of the algorithm are $\alpha = 0.2$, $\beta = 0.8$, $\gamma = 1$ and $T = 200$. In Fig. 1.1a we can see the time evolution of the activity correlations of two and three cortical units given by the average of each of the terms that appear in eq.(1.7).

The artificial scotoma was introduced at $t = 200$, after adaptation to normal conditions had occurred; we can see that the change in the statistics makes the correlations to increase but after some time the network adapts to the new stimuli and they again converge to zero. In Fig. 1.1b we show that also a fourth order and another third order correlations cancel, suggesting that the algorithm has indeed succeeded in reducing the redundancy in the output units. We want to remark that even for the simple image ensembles that we are considering here the processing taking place between the two layers (such as the instantaneous average defined in eq.(1.5)) introduces non trivial correlations. Also these should be suppressed by eq.(1.7), and they are as shown in Figs. 1.1.

The RFs obtained after adaptation to the statistics of normal visual conditions and after adaptation to the inputs with the scotoma are shown in Fig. 1.2a (for a unit in the cortical scotoma) and in Fig. 1.2b (for a unit with the RF well outside the scotoma). We can see that a unit inside the

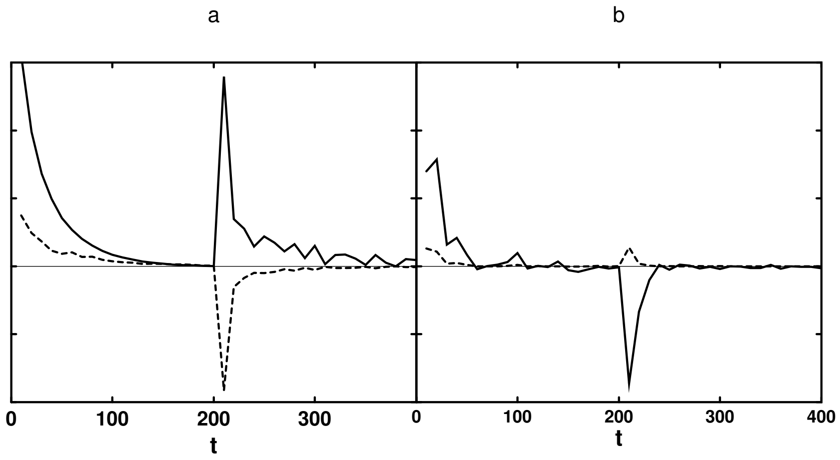

Figure 1.1: Second, third and forth order statistics as a function of time, $\tau_s = 1$, $\alpha = 0.2$, $\beta = 0.8$, $\gamma = 1$ and $T = 200$ (the activity is measured in arbitrary units). ¿From time 0 to 200, the inputs have the statistics of the natural images. ¿From 200 to 400 they have the scotoma. Stimuli are presented regularly and one time unit corresponds to the presentation of a single pattern. To compare with the experiments in (Kapadia, Gilbert and Westheimer 1994) this time unit has to be taken equal to 33 msec. **a)** Solid line: $\overline{<\delta_T u_i \delta_T u_j>}$; dashed line: $\overline{<\delta_T u_i \delta_T u_j^2>}$. The upper bar denotes the average over the units. **b)** Solid line: $\overline{<\delta_T u_i \delta_T u_j \delta_T u_k>}$; dashed line: $\overline{<\delta_T u_i \delta_T u_j \delta_T u_k \delta_T u_l>}$.

cortical scotoma has increased its maximum response by about 100 %, while the other is unchanged.

The size of the RF expansion depends on the parameters of the model. It is easy to understand the effect of T on the expansion. When the artificial damage is presented the activity of units in the cortical scotoma drops immediately to zero. If T is relatively small, their average activity $< u_i >_T$ would also vanish and then there would be no substantial modifications of the horizontal connections affecting those units and no expansion of their RFs. Fig. 1.3 shows the dependence of the ratio between the maximum responses after and before the adaptation to the scotoma on the time constant T. We see that as T decreases the ratio tends to 1, meaning that no expansion has occurred.

The effect of the other parameters is more complex. In Fig. 1.4a we

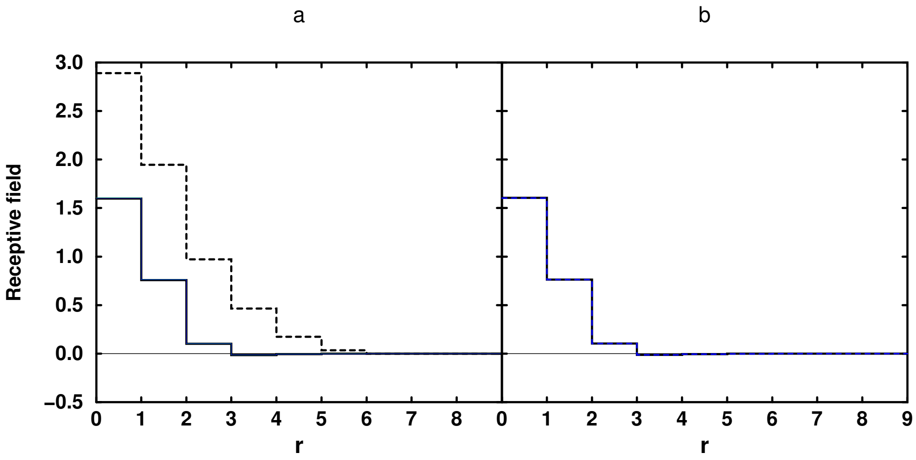

Figure 1.2: RFs are measured after averaging the internal field of the unit when the stimulus is presented at a distance r from the center of the system. Same parameters as in Fig. 1.1. Solid line: RF after adaptation to normal visual conditions. Dashed line: RF after adaptation to images with a scotoma. **a)** Change in RF for a neuron in the cortical scotoma. **b)** Change in RF for a unit outside the cortical scotoma.

show the effect of varying α fixing $\gamma = 1$ and $\beta = 0$; and also for $\beta = 1$ and $\gamma = 0$ and for $\beta = 0.5$ and $\gamma = 0.5$.

For all these cases we observe that the expansion is a monotonously decreasing function of α, for positive values of this parameter. Let us also remark that these results are consistent with the first-order analysis performed in the previous section, where it was predicted that a positive value of α would lead to inhibitory connections and contraction of the RF. Negative values of α generate a contribution that tends to *correlate* the activity of the output units. This means that excitatory connections will be generated. If $-\alpha$ is large enough these excitatory connections become so strong as to make the network unstable, in the sense that even a very small input will lead the whole system to its saturation value. In this situation the output does not contain any information about the stimuli. This gives a lower bound on the admissible values of α. Let us remark that saturation is relevant only in this case. If α is positive then the system never leaves the linear region of the transfer function. Therefore the results for zero or positive *alpha* would be the same for a semilinear transfer function. For a positive value of α absence of saturation can lead to diverging values for the activities of the

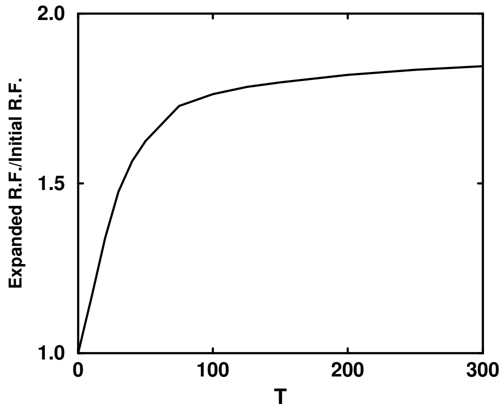

Figure 1.3: Ratio between the final and the initial RFs at its maximum value as a function of T for $\tau_s = 1$, $\alpha = 0.2$, $\beta = 0.8$ and $\gamma = 1$.

cortical units and values of the intracortical connections. In this case the excitatory connections become so strong that there is no enough inhibition in the system to offset its effect.

The higher order terms in the updating rule, i.e. the ones proportional to β and γ, *are indeed necessary* in order to obtain an expansion of the RFs as large as the one observed in the experiments (Pettet and Gilbert 1992). Fig. 1.4b shows the results for the relative expansion as a function of α when $\beta = 0$ and $\gamma = 0$ for different values of T (for values of α smaller than those shown in the figure the algorithm fails to decorrelate). We can see that even a large T does not give a sufficiently large expansion. By running simulations with networks of size 32×32, we have checked that this result is not affected by finite size effects. We have also checked that the expansion of the RFs can be reversed by re-stimulating the region of the artificial damage after the network has adapted to it.

1.4.2 Perceptual Shifts

We now analyze some psychophysical consequences of the RF expansion. In particular we will be concerned with the alterations that it induces in spatial localization task. In the experiment described by (Kapadia, Gilbert and Westheimer 1994) the subject first adapts to stimuli with an artificial scotoma, after some time these are replaced by another stimulus composed

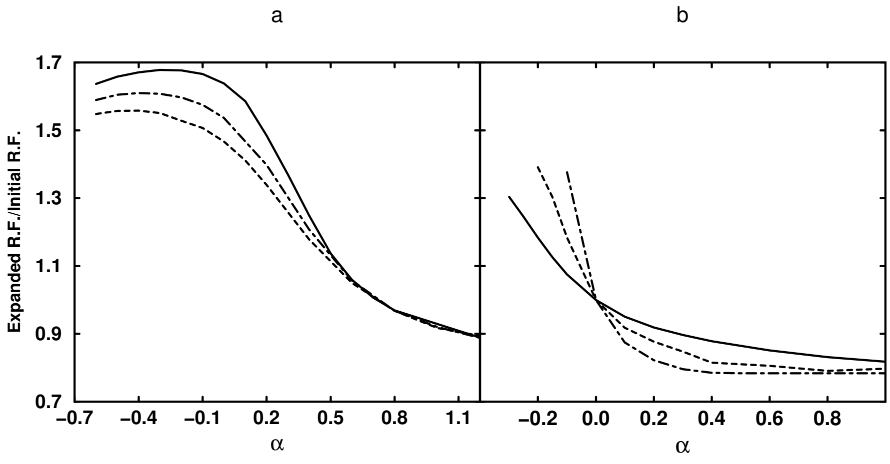

Figure 1.4: The ratio between the final and the initial RFs at their maximal values. For each one of the curves, if smaller values of α are used the algorithm does not converge to the solution with zero correlations. **a)** The ratio as a function of α ($T = 200$): Solid line: $\beta = 0$ and $\gamma = 1$; dashed line: $\beta = 1$ and $\gamma = 0$; dot-dashed line: $\beta = 0.5$ and $\gamma = 0.5$. **b)** The ratio as a function of α for $\beta = \gamma = 0$ and different values of T. Solid line: $T = 50$; dashed-line: $T = 100$; dot-dashed line: $T = 200$.

of three horizontal bars disposed along the vertical axis. The central bar is positioned in one of seven positions, either centered between the top and bottom bars or shifted by a fixed distance one, two or three units above or below the center. The task consists in saying whether the middle bar appears closer to the top or to the bottom bars. The experiment shows that if the test bars are presented in a position close to the place where the border of the artificial scotoma was located there is a systematic bias in the perceptual responses. The central bar appears to be shifted upwards, if it is presented near to the lower border of the scotoma.

To relate the results of the simulations with these psychophysical results it is necessary to propose how the system estimates the position of the bars from the activity of cortical cells. One estimator of the position along one of the axis (say, y) is the *center-of-gravity* of the neuronal activation profile (Baldi and Heilimberg 1988). It can be obtained in the following way: let us de denote by w_y the average of the activities of all the units whose RF center is at coordinate y. An estimator E of the position of any stimulus will be

$$E = \frac{\sum_y y w_y}{\sum_y w_y}. \tag{1.10}$$

If we present as a stimulus the three bars with the one in the middle placed at the center, the shift in the estimation will be

$$S = E_2 - \frac{E_1 + E_3}{2}, \tag{1.11}$$

where E_1, E_2 and E_3 denote the estimation of the position of the three bars. In Fig. 1.5a we show the time evolution of this quantity. It grows fast as soon as the stimuli with a lesion are presented, reaching saturation after a time that depends on T and on the values of the parameters in the updating rule (α, β and γ). For the same parameters as in Fig. 1.1, we find that the shift saturates about 30-40 time units after the stimuli with scotoma are presented. If the patterns are presented with a rate of 30 Hz, as in (Kapadia, Gilbert and Westheimer 1994), this means a convergence time of about 1-1.2 secs, which is roughly the same as in the experiments. Increasing α, β and γ or decreasing T makes the convergence faster. On the other hand a smaller T implies that changes take place faster and then the asymptotic value of the shift will be smaller. The relation between this time scale and the one at which the *filling-in* effect occurs will be discussed in the next subsection.

We also evaluated the dependence of the shift with the position of the stimulus. We find that the shift is maximum when the stimulus is centered just at the border of the scotoma (see Fig. 1.5b), agreeing with the results of (Kapadia, Gilbert and Westheimer 1994). The numerical data also show that if the test stimulus is presented at a suitable distance from the scotoma we find that the sign of the shift is reversed; this effect could be due to the expansion of cells with their RF only partially inside the scotoma.

1.4.3 Filling-in effect

The physiological correlate of the perceptual fill-in is the recovery of the activity of the neurons in the cortical scotoma (in V1 or in higher areas) to the level it had in the absence of the artificial damage. Equivalently, one can compare the activity of those neurons with the average activity of the neurons outside the scotoma. Let us remark that the recovery of activity in V1 is a controversial fact. Experiments performed by de Weerd et al. (DeWeerd et al 1995) have failed to find recovery of activity in V1, but it was found in V2 and in V3. On the other hand the expansion of the RF of

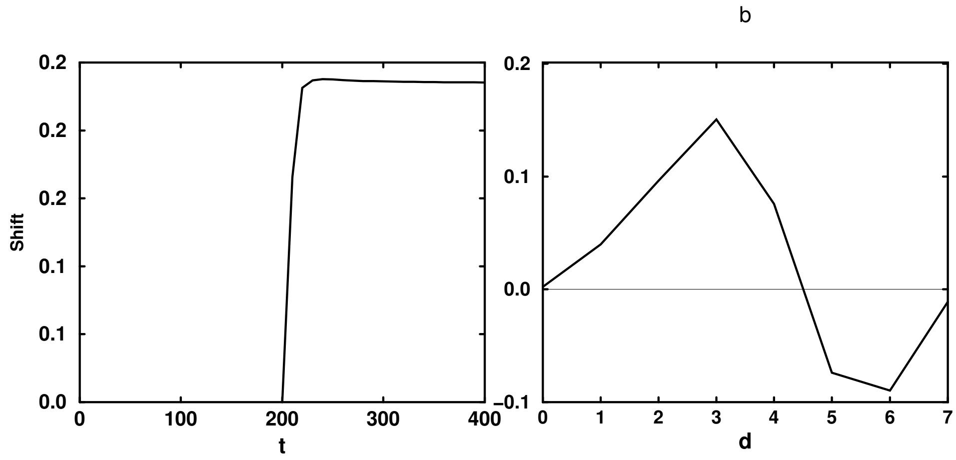

Figure 1.5: Perceptual shift. The test stimulus is centered at the border of the scotoma. Same parameters as in Fig. 1.1. **a)** The shift as a function of time. In physical units it reaches its asymptotic value at about 1 sec. **b)** The shift for different positions of the test stimulus relative to the scotoma. The distance between the center bar of the test stimulus and the center of the scotoma is d. A positive value of the perceptual shift means that the middle bar is perceived above its true position.

a cell to a region outside the cortical scotoma found in (Gilbert and Wiesel 1992) indicates that the recovery does indeed take place and that the same phenomenon will induce the appearance of perceptual fill-in, as noticed in (Pettet and Gilbert 1992).

We now examine the question of the recovery of the activity of a unit in the cortical scotoma and compare its time course to the one of the perceptual phenomenon. We interpret that the *filling-in* effect occurs when the adaptation of the horizontal connections to the artificial damage leads the activity of a unit in the cortical scotoma to a value comparable to the one of the units outside it. In Fig. 1.6 we show the level of activity of one unit in the cortical scotoma compared with the average activity of the whole system for $\alpha = 0.2$, $\beta = 0.8$, $\gamma = 1$ and $T = 200$. We can see that they become similar about 70 time units after the stimuli with the scotoma are presented. In Fig. 1.5 we showed the time evolution of the shift for the same parameters. We can see that it reaches its saturation value about 30-40 time units after the presentation of the scotoma. The comparison of the time scales shows that it takes longer to achieve the *filling-in* than to saturate the shift, although the difference is smaller than the factor 5-10 found in (Kapadia, Gilbert and Westheimer 1994). The value of the saturation activity of the unit in the scotoma depends on the parameters α, β, γ and T. For fixed values of α, β and γ it increases monotonically with T. If T is too small the activity will saturate to a value smaller than the average. If it is too large it will saturate to a value larger than the average. For that reason the *filling-in* phenomenon imposes a constraint on the model parameters.

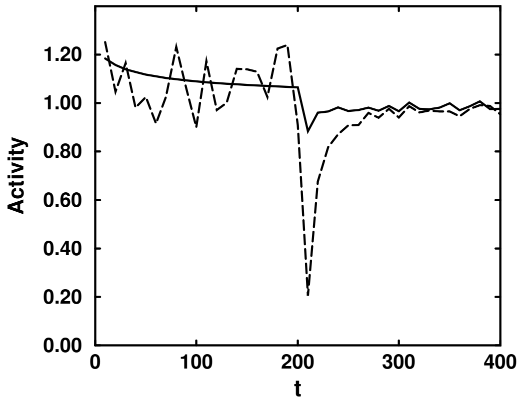

Figure 1.6: Physiological filling-in effect. Same parameters as in Fig. 1.1. Activity is measured in arbitrary units. Solid line: average activity of the system; dashed line: activity of a neuron in the cortical scotoma.

Let us remark that for the set of parameters used above all the experimental constraints (expansion of RFs, psychophysical shifts, *filling-in*) are satisfied. However a precise tuning to these values is not necessary to account for the experiments. There is a whole region in parameter space that does it. For instance, all those experiments are still explained by the model if α is taken in the range $[-0.25 : 0.5]$. On the other hand, changes of T affect more strongly the results, specially the time when the *filling-in* effect occurs. Depending on the precision required for the similarity between the firing rate of the unit in the cortical scotoma and the its average value on the layer, a change in T will have to be compensated by changes in the rest of the parameters.

1.4.4 Discussion of the results

As we have seen in section 1.4.1, the lowest order term in the equation for the synaptic plasticity, eq.(1.7), cannot account by itself for the phenomena we want to explain. For the values of the time constants required for the appearance of the perceptual shifts or the *filling-in* effect one cannot find a value of the parameter α giving the correct expansion of the RF if $\beta = \gamma = 0$. However the maximum expansion that can be achieved for $\beta = \gamma = 0$ is not very far from the wished value. This result cannot be qualitatively improved by changing the slope or the form of the activation function. We conclude that the usual covariant rule cannot explain all the phenomena. To obtain the correct expansion of the RF within a time scale of the order of a few seconds, as observed in psychophysical experiments, requires the higher order terms in the learning rule. In the next section we discuss some possible modifications of the learning rule that could be done in order to check the robustness of this conclusion.

1.5 Conclusions

In this work we have analyzed the dynamics of RFs on the basis of some experimental results on damaged (artificial or real) sensory inputs (Gilbert and Wiesel 1992), (Pettet and Gilbert 1992), (Kapadia, Gilbert and Westheimer 1994), (Jenkins et al 1990), (Merzenich et al 1984) and (Recanzone et al 1992). The dynamics we used can be put into a wider theoretical framework based on information theory criteria such as minimal redundancy (Barlow 1961). In our model redundancy is minimized by updating the effective horizontal connections with a rule, eq. (1.7), that has a fixed point where second- and a higher-order correlations are zero. This dynamics induces changes in the RFs of cortical cells which have allowed us to explain several basic experiments. Besides the expansion of the RF size (Pettet and Gilbert 1992) we have also analyzed the generation and time evolution of shifts in feature localization tasks (Kapadia, Gilbert and Westheimer 1994) and the *filling-in* of stimuli with an artificial scotoma (Ramachandran and Gregory 1991), (Kapadia, Gilbert and Westheimer 1994). We have found that for a suitable region in the space of parameters it is possible to reproduce most of the experimental results. The relevant time scales for the expansion are T and τ_s. The averaging time window T plays an important role in the dynamics. If it is too small the average activities go to zero too fast after the stimulus with the scotoma is presented and there is no time to generate the additional horizontal connection that will expand the RF. On the other hand if it is too large the network will take too long to adapt to that stimulus. We can express the value of T required to fit the data in physical units. In (Kapadia, Gilbert and Westheimer 1994) the images with the artificial scotoma are presented at a rate of 30 Hz. On the other hand the expansion of the RF (shown in Fig. 1.2a), the shifts in the psychophysical task (shown in Fig.1.5b) and the *filling-in* effect (shown in Fig.1.6), that are compatible with the experimental results, are obtained for $T = 200$, which means $T \sim 6$ secs.

RF dynamics based on information theory ideas have been also proposed by Sirosh et al. (Sirosh, Miikkulainen and Bednar 1996). Their rule however only decorrelates the output activities instead of removing the redundancy. This is because the higher order terms are not included in their learning rule. An expansion of the RF is found in this work but the problem of the time scale where the perceptual shifts occur is not addressed in this work. According to a different strategy suggested in (Phillips, Kay and Smyth 1995) the output code is constructed by maximizing the information shared by the RF and the context. The RF is generated by the feed-forward connections and the context by the lateral connections. In this case no expansion of the RF should be observed because in the absence of RF input (as is the case after the damage) the state of a neuron is silent independently of the value of the context. The reason for this behavior is that in Phillips et al. (1995) the objective is to maximize the information that the units transmit about their RF inputs irrespective of the context, while in our work we maximize the information about the input without discriminating between RF and context.

The dynamical changes have been studied in the context of the modifications of horizontal connections in V1. However similar changes can be observed when different kinds of inputs (that include more complex stimuli) are presented to the system, suggesting that connections in different regions of the cortex can also be modified. In (Ramachandran and Gregory 1991) it is found that if the pattern surrounding the scotoma contains type-written text, the scotoma is filled-in by text (although subjects could not actually read the letters in the filled-in region). Moreover, if the surrounding pattern suddenly changes (for instance from twinkling dots to flickering lines) the process of filling-in has to begin again. This is consistent with a change of the locus of the synaptic modification.

It would be worthwhile to explore the relation between this model for effective synapses and models defined at the cell level, e.g. the one discussed in (Stemmler, Usher and Niebur 1995). This would allow us to express the high order terms in eq. (1.7) in terms of parameters describing the inhibitory interneurons and the true synaptic plasticity of horizontal connections. Another possibility would be to generalize the model described in (Amit and Brunel 1995) including a more structured inhibition. Another direction for future reasearch would be to compare different ways of implementing the memory effect; this could be done for instance as in (Foldiak 1991), (Wallis and Rolls 1996).

Acknowledgements

We warmly thank Jean-Pierre Nadal, Edmund Rolls and Antonio Turiel for discussions. G.M. thanks the spanish Ministery of Education and Science for financial support. An E.U. grant CHRX-CT92-0063 and a spanish grant AEN96-1664 are also gratefully acknowledged.

Bibliography

[1] Amit D. and Brunel N. 1995, Learning internal representations in an attractor neural network, *Network* 6:359-388

[2] Baldi P. and Heilimberg W. 1988, How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers, *Biol. Cybern.* 59:313-318

[3] Barlow H. B. 1961 Possible principles underlying the transformation of sensory messages. In Rosenblith W., editor, *Sensory Communication*, page 217. Cambridge:M.I.T. Press

[4] Bienenstock E. L., Cooper L. N. and Munro P. W. 1982, Theory for development of neuron selectivity: orientation specificity and binocular interaction in visual cortex *J. Neurosci.* 2:32-48

[5] Clothiaux E. E., Bear M. F. and Cooper L. N. 1991, Synaptic plasticity in visual cortex: comparison of theory with experiment, *J. Neurophysiol.* 66:1785-1804

[6] Darian-Smith C. and Gilbert C. D. 1995, Topographic reorganization in the striate cortex of the adult cat and monkey is cortically mediated, *Journal of Neuroscience* 15:1631-1647

[7] Das A. and Gilbert C. D. 1995, Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex, *Nature* 375:780-784

[8] Field D. J. 1987, Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am.* 4:2379-2394

[9] Foldiak P. 1991, Learning invariance from transformation sequences, *Neural Computation* 3:194-200

[10] Gilbert C. D. 1992, Horizontal integration and cortical dynamics *Neuron* 9:1-13

[11] Gilbert C. D., Das A., Ito M., Kapadia M. and Westheimer G 1996, Spatial integration and cortical dynamics, *Proc. Natl. Acad. Sci. USA* **93**, 615-622

[12] Gilbert C. D. and Wiesel T. 1992, Receptive field dynamics in adult primary visual cortex *Nature* 356:150-152

[13] Grinvald A., Lieke E. E., Frostig R. D., and Hildesheim R. 1994, Cortical Point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex, *Journal of Neuroscience* **14**, 2545-2568

[14] Hérault J. and Jutten C. 1994, *Réseaux neuronaux et traitment du signal*, Paris:Hermes

[15] Hirsch J. A. and Gilbert C. D. 1991, Synaptic Physiology of Hori zontal Connections in the Cat's Visual Cortex, *Journal of Neuroscience* **11**, 1800-1809,

[16] Jenkins W. M., Merzenich M. M., Ochs M. T., Allard T., and Guic-Robles E. 1990, Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation, *J. Neurophys.* 63:82-104

[17] Kapadia M. K., Gilbert C. D. and Westheimer G. 1994, A quantitative measure for short-term cortical plasticity in human vision, *J. Neurosci.* 14:51-457

[18] Kapadia M.K., Ito M., Gilbert C.D. and Westheimer G. 1995, Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys, *Neuron* **15**, 843-856

[19] Knierim J. J. and van Essen D. C. 1992. Neuronal Responses to Static Texture Patterns in Area V1 of the Alert Macaque Monkey, *Journal of Neurophysiology* **67**, 961-980

[20] Mato G. and Parga N. 1996, Dynamics of receptive fields in the visual cortex, *Network* (submitted).

[21] McGuire B.A., Gilbert C.D., Rivlin P.K. and Wiesel T.N. 1991. Targets of horizontal connections in macaque primary visual cortex, *Journal of Comparative Neurology* **305**, 370-392

[22] Merzenich M. M., Nelson R. J., Stryker P., Cynader S., Schoppmann A. and Zook J. M. 1984, Somatosensory cortical map changes following digit amputation in adult monkeys, *J. Comp. Neurol.* 224:591-605

[23] Nadal J.-P. and Parga N. 1996, Redundancy reduction and independent component analysis: algebraic and adaptive approaches, to appear in *Neural Computation*

[24] Pettet M. W. and Gilbert C. D. 1992, Dynamic changes in receptive-field size in cat primary visual cortex, *Proc. Natl. Acad. Sci. USA* 89:8366-8370

[25] Phillips W. A., Kay J. and Smyth D. 1995, The discovery of structure by multi-stream networks of local processors with contextual guidance, *Network* **6**, 225-246

[26] Polat U. and Sagi D. 1993, Lateral Interactions Between Spatial Channels: Suppression and Facilitation Revealed by Lateral Masking Experiments, *Vision Research* **33**, 993-999

[27] Polat U. and Sagi D. 1994, The Architecture of Perceptual Spatial Interactions, *Vision Research* **34**, 73-78

[28] Ramachandran V. S. and Gregory R. L. 1991, Perceptual filling in of artificially induced scotomas in human vision, *Nature* 350:699-702

[29] Recanzone G. H., Merzenich M. M., Jenkins, W. M., Kamil A. G. and Dinse H. R. 1992, Topographic reorganization of the hand representation in cortical area 3b of owl monkeys trained in a frequency discrimination task, *J. Neurophysiol.* 67:1031-1056

[30] Ruderman D. L. 1994, The statistics of natural images, *Network* 5:517-548

[31] Sirosh J., Miikkulainen R. and Bednar J. 1996, Self-Organization of Orientation Maps, Lateral Connections, and Dynamic Receptive Fields in the Primary Visual Cortex. In *Lateral Interactions in the Cortex: Structure and Function.* Eds: Sirosh, J., Miikkulainen, R., and Choe, Y..

The UTCS Neural Networks Research Group, Austin, TX. Electronic book, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96.

[32] Stemmler M., Usher M. and Niebur E. 1995, Lateral interactions in primary visual cortex: a model bridging physiology and psychophysics, *Science* 269:1877-1880

[33] Wallis G. and Rolls E. 1996, A model of invariant recognition in the visual system, *Progress in Neurobiology* **51** 167-194

[34] DeWeerd P., Gattass R., Desimone R. and Ungerleider L. 1995, Responses of cell in monkey visual cortex during perceptual filling-in of an artificial scotoma, *Nature* 377:731-734

[35] Xing J. and Gerstein G. 1994, Simulation of dynamic receptive fields in primary visual cortex 1994, *Vision Res.*, 34:1901-1911