Title

Binding of Object Representations by Synchronous Cortical Dynamic s Explains Temporal Order and Spatial Pooling Data

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Authors

Grunewald, Alexander
Grossberg, Stephen

Publication Date

1994

Binding of Object Representations by Synchronous Cortical Dynamics Explains Temporal Order and Spatial Pooling Data

Alexander Grunewald
Department of Cognitive and Neural Systems
Boston University
Boston, MA 02215
alex@cns.bu.edu

Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
Boston, MA 02215

Abstract

A key problem in cognitive science concerns how the brain binds together parts of an object into a coherent visual object representation. One difficulty that this binding process needs to overcome is that different parts of an object may be processed by the brain at different rates and may thus become desynchronized. Perceptual framing is a mechanism that resynchronizes cortical activities corresponding to the same retinal object. A neural network model based on cooperation between oscillators via feedback from a subsequent processing stage is presented that is able to rapidly resynchronize desynchronized featural activities. Model properties help to explain perceptual framing data, including psychophysical data about temporal order judgments. These cooperative model interactions also simulate data concerning the reduction of threshold contrast as a function of stimulus length. The model hereby provides a unified explanation of temporal order and threshold contrast data as manifestations of a cortical binding process that can rapidly resynchronize image parts which belong together in visual object representations.

Introduction

The primate visual system performs the complex task of analyzing the visual environment in several stages. At the first stage, the retina, the incoming image is transduced into neural signals. These signals are then transmitted to the lateral geniculate nucleus (LGN) and from there to the striate cortex (V1). Cells in all these stages have comparatively small receptive fields, with the biggest being in V1. Striate receptive fields have at most a diameter of about one degree in the fovea (Hubel & Wiesel, 1968). Unlike the receptive fields of cells in the retina and the LGN, receptive fields of striate neurons tend to have a preferred orientation. These cells fire optimally when a bar of their preferred orientation is in their receptive fields. Since the receptive fields are rather small, it can be said that striate neurons respond to local features, and hence they decompose the retinal image into its main local orientations.

It is known that the latency of the response onset of retinal and geniculate neurons is variable, even to identical stimuli (Shapley & Victor, 1978; Sestokas & Lehmkuhle, 1986). Moreover, the latency depends on stimulus parameters. More luminant stimuli are processed faster than less luminant stimuli, and higher spatial frequencies are processed faster than lower spatial frequencies (Bolz, Rosner, & Wässle, 1982; Sestokas & Lehmkuhle, 1986).

Since most images from a real environment contain a variety of luminances and spatial frequencies, processing of different parts of an image may happen at different rates, so that the cortical representation of the image may be desynchronized. As long as the retinal image is constant, this does not cause serious problems. However, when there is motion in the retinal image, the visual system needs to ensure that all the parts corresponding to the same retinal image are processed together, to avoid false conjunctions that could impair recognition of objects in a scene. Under extreme conditions, such as the rapid presentation of visual stimuli, it can happen that false conjunctions do occur (Intraub, 1985).

Perceptual framing is the process whereby the parts of an image are resynchronized (Varela, Toro, John, & Schwartz, 1981). In the present study, a neural network model is presented that exhibits perceptual framing. That is, temporally offset inputs to the network are resynchronized. The present study also shows that perceptual framing can be implemented with the same type of horizontal connections that have been postulated in a model of form perception and perceptual grouping (Grossberg & Mingolla, 1985a, 1985b).

Another issue that has to be solved by the visual system is that information in the visual cortex is spatially distributed. How does this information get bound together into coherent object representations? This is necessary since the interpretation of an image, which includes the recognition and the localization of objects in the image, requires global information. The anatomy of visual cortex suggests that horizontal connections within each area, and feedback connections between different areas, occur at multiple processing stages (Felleman & Van Essen, 1991; Gilbert, 1993). The present study focuses on interactions between neighboring neurons via feedback from subsequent processing stages. It is shown that horizontal integration within the visual cortex can enhance performance of single cortical cells, and thus can form a starting point for the global understanding of visual images.

Perceptual Framing

Perceptual framing is the process of binding together parts of neural representations corresponding to the same image that may have come temporally out of register due to early processing. Bottom–up convergence of signals in visual cortex does not suffice as a mechanism for resynchronization, if only because cortical cells have a fast rate of integration (Mason, Nicoll, & Stratford, 1991), yet the responses of cortical cells within the first 5ms after response onset is a 95% accurate predictor of the entire response strength (Celebrini, Thorpe, Trotter, & Imbert, 1993; Oram & Perrett, 1992). It has also been shown that synchronization cannot be mediated by a clocking mechanism such as the cortical alpha–rhythm (Gho

& Varela, 1989).

Here we model how synchronization of distributed cortical activities can temporally realign out-of-phase image parts. The results model data showing that cortical activities synchronize in the cat and in the monkey when a stimulus is present in the visual field (Eckhorn, Bauer, Jordan, Brosch, Kruse, Munk, & Reitboeck, 1988; Gray & Singer, 1989), even when the receptive fields of the units recorded do not overlap.

This synchronization property of the neural network model means that if two stimuli are presented at two different locations, separated by a stimulus onset asynchrony (SOA), then the neural activities at the locations corresponding to those stimuli are separated by a smaller amount of time. This is how perceptual framing manifests itself in this model. Synchronization can only occur for SOAs that are not too big, so that perceptual framing only occurs for small SOAs.

A way to test this notion of perceptual framing is to link it to temporal order judgments (TOJs) of two separate visual stimuli. When perceptual framing breaks down, two stimuli will be perceived as successive, and thus it will be possible for observers to identify their temporal order. Hirsch and Sherrick (1961) have found a psychometric function of the accuracy of TOJs as a function of the stimulus onset asynchrony. Their results are shown in Figure 1. At an SOA of 20ms, observers can tell the temporal order with 75% accuracy.
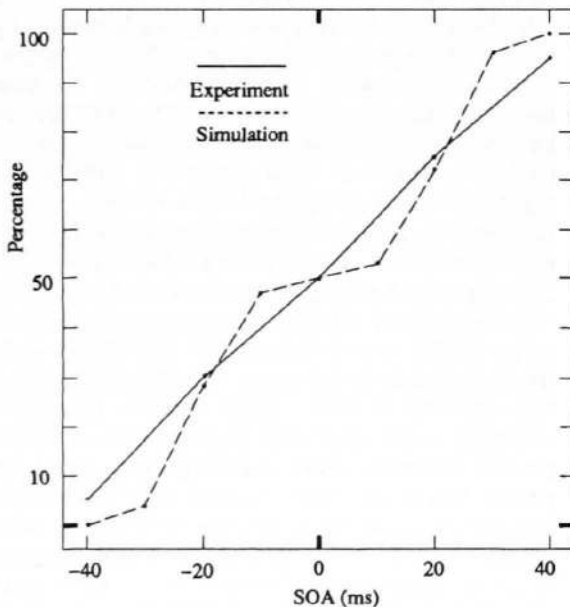

Figure 1: Accuracy of temporal order judgment as a function of SOA. Comparison between experimental results and the model proposed. SOA indicates the time by which stimulus one (e.g. the "right stimulus") leads the other stimulus in a two stimulus presentation task. The ordinate gives the percent responses that stimulus one appeared first. Solid line: results from experimental study. Dashed line: results from simulation of the model.

Spatial Pooling

Spatial summation is the effect that when stimuli are very small (typically smaller than the size of a striate receptive field), then an increase of stimulus size leads to a reduction of threshold contrast (Thomas, 1978). These experiments were conducted under the assumption that spatial summation only occurs within the range of a receptive field, and hence little data are available that indicate cooperative interactions over sizes that go significantly beyond a single receptive field. Gilbert (1993) has, however, reviewed evidence challenging the classical notion of a receptive field in the light of psychophysical, anatomical and physiological evidence suggesting the existence of extensive lateral interactions.
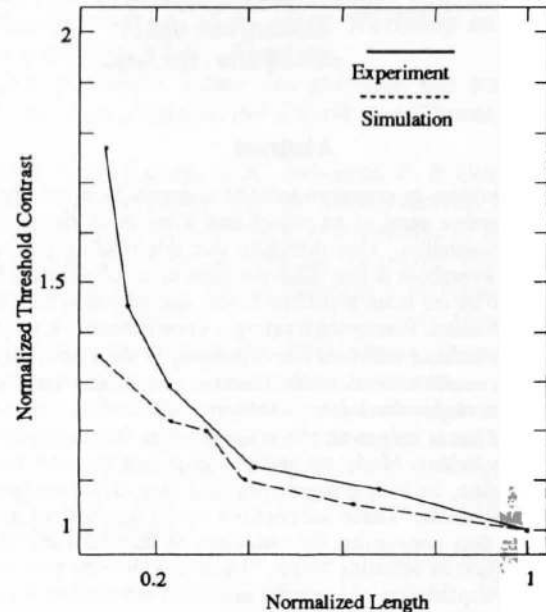

Figure 2: Comparison between psychophysical data and computer simulations. Normalized threshold contrast as a function of normalized stimulus size. Solid line: results from experimental study. Dashed line: results from simulations.

The present authors are aware of only one study that compares contrast thresholds when stimulus lengths go beyond that of the diameter of a single striate receptive field. Essock (1990) showed a reduction of threshold contrast up to grating lengths of 5.5 degrees. The results of Essock (1990) are replotted in Figure 2. It can be seen how the threshold decreases with increasing stimulus length. The length over which pooling occurs is too long to allow an explanation within a striate receptive field, thus suggesting some kind of horizontal cortical cooperation. In the present study we call this effect spatial pooling to distinguish it from spatial summation proper.

Description of the Model

Grossberg and Mingolla (1985a, 1985b) developed a model called the Boundary Contour System (BCS) for the generation of emergent boundary segmentations by the visual cortex. This model was later adapted to show that cortical synchronization of neural activities does not require the presence of a central clocking mechanism (Grossberg & Somers, 1991). In the present study, we further develop and modify this model. There are two layers, one consisting of fast-slow neural oscillators (Ellias & Grossberg, 1975), and the other of bipole cells, that receive input to two separate receptive field lobes,

in addition to receiving direct bottom–up input. In the present simulations, bipole cells fire if at least two of its three receptive zones are activated. The architecture of the model is shown in Figure 3.

probability P. In Figure 1 the experimental results of Hirsch and Sherrick (1961) about temporal order judgments and the simulation results are compared. The simulations match the data closely.

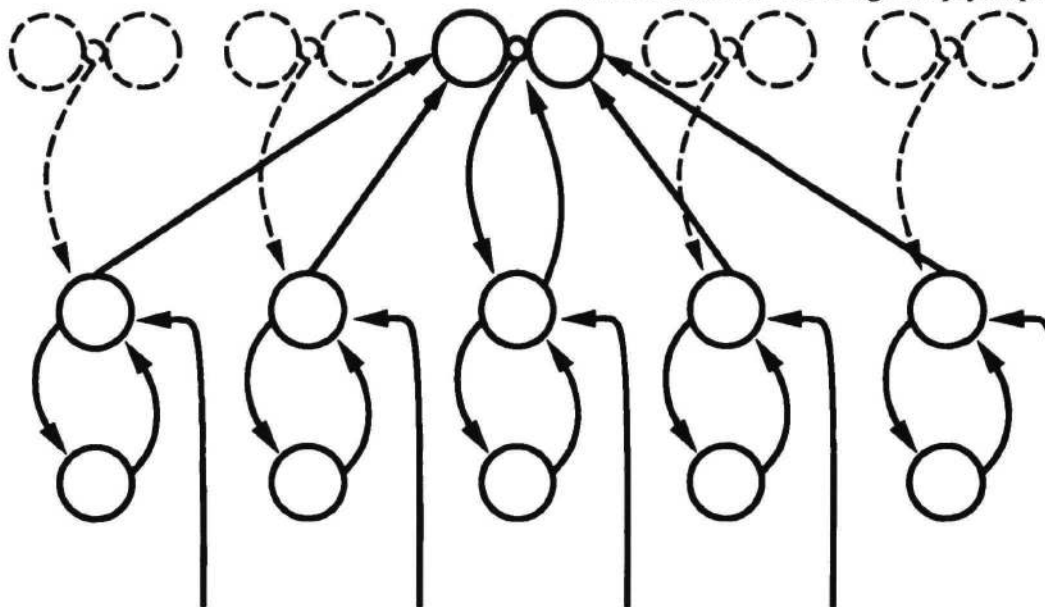We also tested the model against psychophysical data on

Figure 3: The architecture of the model proposed. A layer of fast–slow oscillators is coupled via a layer of bipole cells. In contrast to previous versions of the model, there is a direct signal from each oscillator to its corresponding bipole cell which facilitates boundary completion.

In the simulations two stimuli of the same length are presented to different locations of the neural network, separated by an SOA. The internal time difference Δt is the minimal time between peaks of neural activity of the fast neurons at the two sites corresponding to the stimuli. For different SOAs we found the internal time difference Δt for the corresponding neural signals in our model. The time of the response peak corresponding to each of the two stimuli is a random variable, and the mean of the difference between the two random variables corresponding to the two stimuli is the internal time difference Δt. The probability that each of those neural signals occurs at any given time follows the normal distribution, where the mean of the first signal can be set to zero, and the mean of the second can be set to Δt. The standard deviation of the time of the peak response σ is the same for both, and has been reported to be 6ms (Maunsell & Gibson, 1992; Zack, 1973). The probability that the signal corresponding to the first stimulus is perceived first can be found by taking the difference of the two random variables, which is also a normal distribution, with mean Δt and standard deviation $\sqrt{2}\sigma$. Thus the probability that the first stimulus in a two stimuli paradigm is perceived first, and hence that the temporal order of the stimuli is perceived correctly is given by

$$P = \Phi\left(\frac{\Delta t}{\sqrt{2}\sigma}\right),$$

where Φ is the cumulative normal distribution function. Each SOA leads to a different value for Δt, and hence a different

spatial pooling (Essock, 1990) by finding the minimal inputs necessary to yield oscillations (and therefore activities above baseline level) for different input sizes. The results from the simulations and from the experiment were normalized. Normalization was performed by dividing the input (or the contrast) by the value at which it asymptotes for large stimulus sizes. Since smaller stimuli require more contrast to be detected, the normalized contrast for these stimuli is greater than one. Size was also normalized with respect to the asymptoting size. The normalized experimental data and computer simulations are shown in Figure 2. The difference between the two curves at small stimulus lengths can be attributed to the fact that in the present study it was assumed that the stimulus spans the full size of the cortical receptive fields. Had we taken into account that very small stimuli only excite parts of the receptive field of a single neuron, and thus the bottom–up input is weakened and needs to be compensated by higher contrast, the match would be very good. However, in this study the focus was on lateral interactions via feedback.

Discussion

In this study, we have shown data about that perceptual framing and spatial pooling can be quantitatively explained using a model neural process of cortical cooperation across space and time. Such an architecture helps to bind perception of a whole with its parts, and thus provides a key step in the global integration of a percept. It is also known that synchronization of convergent neural inputs drive long-term potentiation (LTP) (Bliss & Collingridge, 1993). Thus perceptual framing may be utilized not only for recognition, but also for learning about the visual environment. Indeed it is known that perceptual learning can occur quickly, with effects lasting for a long time (Karni & Sagi, 1993). Together these results sug-

gest that perceptual framing can set up a resonant state that drives learning processes, as in Adaptive Resonance Theory (Grossberg, 1976, 1980).

Appendix: Details of the model

In the simulations of the model there were 64 oscillators arranged along a ring. Each oscillator consisted of two nodes each, one fast and one slow. The activity of the fast node is denoted by x_i, of the corresponding slow node by y_i. The index i denotes the position of the oscillator, and ranges from 1 to 64. Oscillators with indices differing by one are neighbors. Since the oscillators are arranged as a ring, units indexed by 1 and 64 respectively are also neighbors. This structure was chosen to avoid edge effects. Care was taken to ensure that input was sufficiently far removed from the wrap around position to avoid undesirable side effects. The input to the network is denoted by I_i and it is position specific. Associated with every oscillator there is a bipole cell, whose activity is denoted by z_i. The equations governing the oscillators are

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)\left(Cf_o(x_i) + f_o(z_i) + I_i\right)$$
$$\qquad\qquad - Dx_i f_o(y_i)$$
$$\frac{dy_i}{dt} = E(x_i - y_i)$$

where the signal function f_o is given by

$$f_o(x) = \frac{x^{n_o}}{Q_o^{n_o} + x^{n_o}}$$

and A, B, C, D and E are parameters of the network. The parameters n_o and Q_o determine the signal function of the oscillator. The equation governing the bipole cells is

$$z_i = \left[f_b(L_i) + f_b(R_i) + F f_b(C_i) - \Gamma_{cpl}\right]^+$$

where

$$[x]^+ = \max(x, 0),$$

and the bipole signal function is

$$f_b(x) = \frac{x^{n_b}}{Q_b^{n_b} + x^{n_b}}.$$

and F and Γ_{cpl} are parameters. The parameters n_b and Q_b determine the signal function of the bipole cell. The bottom up input to bipole cells is given by

$$L_i = \frac{1}{w} \sum_{j=1}^{w} f_o(x_{i-j})$$
$$R_i = \frac{1}{w} \sum_{j=1}^{w} f_o(x_{i+j})$$
$$C_i = f_o(x_i)$$

where w is the halfwidth of the kernel. The initial conditions of the network where chosen to be $x_i = 0.15$, $y_i = 0.15$, and $z_i = 0$ for all i. The initial value of the slow variable is maintained by tonic input, which is quenched when an input comes on. Scaling of time was done by taking into account recent findings that the period of oscillations in primates is about 15

ms (Eckhorn, Frien, Bauer, Woelbern, & Kehr, 1993). It was found that putting a timestep of 1 unit in the model equal to 1 ms yields good results. The integration stepsize used was $H = 0.1$ ms. The parameters used throughout this report are $A = 1, B = 1, C = 20, D = 33.3, E = 0.1, F = 0.5, n_o = 4, Q_o = 0.9, n_b = 5, Q_b = 0.001, \Gamma_{cpl} = 1, w = 6$.

In the temporal order judgment simulations, each node received a constant level of background activity ($I_i = 0.15$). Two nodes received an input ($I_i = 0.65$) which lasted for 250 ms. The first input ($i = 33$) comes at simulation onset, the second input ($i = 34$) comes on later by an amount specified with SOA.

In the spatial pooling simulations the background activity was set to zero, to avoid unwanted lateral interactions. For each stimulus size all units that received input received the same value. Threshold input was the lowest value (up to 0.01) that led to oscillations.

Reference

Bliss, T. V. P., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature, 361,* 31–39.

Bolz, J., Rosner, G., & Wässle, H. (1982). Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. *Journal of Physiology (London), 328,* 171–190.

Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience, 10,* 811–825.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex?. *Biological Cybernetics, 60,* 121–130.

Eckhorn, R., Frien, A., Bauer, R., Woelbern, T., & Kehr, H. (1993). High frequency (60-90 Hz) oscillations in primary visual cortex of awake monkey. *NeuroReport, 4,* 243–246.

Ellias, S. A., & Grossberg, S. (1975). Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics, 20,* 69–98.

Essock, E. A. (1990). The influence of stimulus length on the oblique effect of contrast sensitivity. *Vision Research, 30*(8), 1243–1246.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1–47.

Gho, M., & Varela, F. J. (1989). A quantitative assessment of the dependency of the visual temporal frame upon the

cortical rhythm. *Journal of Physiology (Paris), 83*(2), 95–101.

Gilbert, C. D. (1993). Circuitry, architecture, and functional dynamics of visual cortex. *Cerebral Cortex, 3*(5), 373–386.

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA, 86*, 1698–1702.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: feedback, expectation, olfaction, illusions. *Biological Cybernetics, 23*, 187–202.

Grossberg, S. (1980). How does a brain build a cognitive code?. *Psychological Review, 87*, 1–51.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92*, 173–211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38*(2), 141–171.

Grossberg, S., & Somers, D. (1991). Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. *Neural Networks, 4*, 453–466.

Hirsch, I. J., & Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62*(5), 423–432.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London), 195*, 215–243.

Intraub, H. (1985). Visual dissociation: an illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 431–442.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*, 250–252.

Mason, A., Nicoll, A., & Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *Journal of Neuroscience, 11*(1), 72–84.

Maunsell, J., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68*(4), 1332–1343.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*(1), 70–84.

Sestokas, A. K., & Lehmkuhle, S. (1986). Visual response latency of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. *Vision Research, 26*(7), 1041–1054.

Shapley, R. M., & Victor, J. D. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *Journal of Physiology (London), 285*, 275–298.

Thomas, J. P. (1978). Spatial summation in the fovea: asymmetrical effects of longer and shorter dimensions. *Vision Research, 18*, 1023–1029.

Varela, F. J., Toro, A., John, E. R., & Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia, 19*(5), 675–686.

Zack, J. L. (1973). Estimation of the variability of the latency of responses to brief flashes. *Vision Research, 13*, 829–835.