1992-07

A Unified Model of Spatiotemporal Processing in the Retina

Gaudiano, Paolo

Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems

http://hdl.handle.net/2144/2102

Boston University
A UNIFIED MODEL
OF SPATIOTEMPORAL PROCESSING
IN THE RETINA

Paolo Gaudiano

July, 1992

Technical Report CAS/CNS-92-022

submitted to Nature
A unified model of spatiotemporal processing in the retina

Paolo Gaudiano

Boston University, Department of Cognitive & Neural Systems
111 Cummington Street, Boston, MA 02215 USA

A computational model of visual processing in the vertebrate retina provides a unified explanation of a range of data previously treated by disparate models. Three results are reported here: the model proposes a functional explanation for the primary feed-forward retinal circuit found in vertebrate retinae, it shows how this retinal circuit combines nonlinear adaptation with the desirable properties of linear processing, and it accounts for the origin of parallel transient (nonlinear) and sustained (linear) visual processing streams as simple variants of the same retinal circuit.

The retina, owing to its accessibility and to its fundamental role in the initial transduction of light into neural signals, is among the most extensively studied neural structures in the nervous system. Since the pioneering anatomical work by Ramón y Cajal at the turn of the last century, technological advances have abetted detailed descriptions of the physiological, pharmacological, and functional properties of many types of retinal cells. However, the relationship between structure and function in the retina is still poorly understood.

This article outlines a computational model developed to address fundamental constraints of biological visual systems. Neurons that process nonnegative input signals—such as retinal illuminance—are subject to an inescapable tradeoff between accurate processing in the spatial and temporal domains. Accurate processing in both domains can be achieved with a model that combines nonlinear mechanisms for temporal and spatial adaptation within three layers of feed-forward processing. The resulting architecture is structurally similar to the feed-forward retinal circuit connecting photoreceptors to retinal ganglion cells through bipolar cells. This similarity suggests that the three-layer structure observed in all vertebrate retinae is a required minimal anatomy for accurate spatiotemporal visual processing.

This hypothesis is supported through computer simulations showing that the model's output layer accounts for many properties of retinal ganglion cells. Moreover, the model shows how the retina can extend its dynamic range through nonlinear adaptation while exhibiting seemingly linear behavior in response to a variety of spatiotemporal input stimuli. This property is the basis for the prediction that the same retinal circuit can account for both sustained (X) and transient (Y) cat ganglion cells by simple morphological changes. The ability to generate distinct functional behaviors by simple changes in cell morphology suggests that different functional pathways originating in the retina may have evolved from a unified anatomy designed to cope with the constraints of low-level biological vision.

Parallel processing in the retina

The retina of all vertebrates is organized into three cellular (or nuclear) layers and two synaptic (or plexiform) layers. Information flows from the photoreceptors, through a layer of bipolar cells, and finally through the retinal ganglion cells, whose axons project via the optic nerve to subcortical and cortical areas. In addition to the feed-forward information processing carried out by these cell types, there exist two classes of cells, horizontal cells and amacrine cells, that carry signals laterally through the retina for additional processing.

Although there is general agreement that retinal cells can be subdivided into these broad classes, morphological and anatomical studies have shown the existence of a great number of cell types within each class, including dozens of ganglion and amacrine cell types. In addition to morphological and anatomical

---

classification schemes, electrophysiological studies have shown the existence of distinct functional classes, such as ON, OFF, and ON-OFF ganglion cells\textsuperscript{10, 11, 12}.

An additional classification scheme subdivides cat ganglion cells into two functional classes known as X and Y cells\textsuperscript{7}. The primary functional distinctions between X and Y cells are found through the null test\textsuperscript{7, 13, 14}: the response of a ganglion cell is recorded while a stationary sinusoidal grating (luminance distribution) is periodically flashed on the screen. The background luminance while the grating is off is equal to the grating's average luminance. Measurement of the cell's response to such spatiotemporal input modulation is repeated as the grating's location in space relative to the cell's receptive field center is systematically varied across trials.

As shown in Fig. 1, two characteristics distinguish X and Y cells. First, the response of X cells to introduction and withdrawal of a grating at zero or 180 degrees spatial phase consists primarily of sustained components, whereas the response of Y cells consists primarily of transient components. Another difference is that, for an X cell, it is generally possible to find two distinct spatial phase angles, usually ±90°, at which no change in response is elicited by grating onset or offset (null response). On the other hand, at phase angles of ±90° the Y cell exhibits an on-off or frequency doubling response, that is, it exhibits a positive-going response to both onset and offset of the grating\textsuperscript{7}. These findings have been confirmed and extended in several studies\textsuperscript{13, 14, 15}, and a relationship has been established between the morphological classes of alpha and beta ganglion cells, respectively, and the functional classes of Y and X cells\textsuperscript{15, 16, 17}. A similar—though not identical\textsuperscript{18}—classification exists in primates, where midget and parasol ganglion cells give rise to the well-known parvocellular and magnocellular visual processing streams\textsuperscript{19, 20}.

The existence of X cell null responses at phase angles of ±90° indicates that the sudden decrease in input to one side of the cell's receptive field exactly cancels the concomitant increase in input to the other side of the receptive field, as illustrated in Fig. 2a. This property of X cells is largely independent of average luminance, contrast, or temporal frequency of modulation, suggesting that X cells perform linear spatial summation of inputs falling over their receptive field\textsuperscript{7}. Linear spatial summation requires that all retinal

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Response of an off-center X-cell (left column) and an off-center Y cell (right column) to introduction and withdrawal of a sinusoidal grating. Response measured in average pulses per second. Length of zero line represents duration of 2 sec. See text for details. Reprinted with permission from Enroth-Cugell and Robson\textsuperscript{7}.}
\end{figure}
Figure 2: Response of a hypothetical ganglion cell receiving inputs through different preprocessing mechanisms, when a spatial sinusoidal grating is introduced (thick solid line) and withdrawn (thick dashed line). Horizontal axis denotes phase angle in degrees between peak of sinusoidal grating and receptive field center of the cell being measured. Hypothetical response in arbitrary units.

(a): Linear preprocessing leads to exact cancellation of input increments (hatched area) and decrements (stippled area) when cell is located at ±90° spatial phase relative to the sinusoidal grating. (b): Static nonlinear preprocessing may distort the sinusoidal distribution depending on average input and contrast, so that input increments and decrements at ±90° may not cancel out, as indicated by different size of hatched and stippled areas. (c) A transient nonlinearity shows initial distortion (thick solid line), which rapidly equilibrates to approximate linearity (thin solid line), but the initial nonlinear distortion should prevent cancellation at ±90° immediately following grating onset.

Processing prior to the ganglion cell layer be perfectly linear: in the absence of compensatory mechanisms, any preprocessing nonlinearity should be detected in the X cell responses to a sinusoidal grating being turned on and off. The presence of on-off responses suggests instead that Y cells receive nonlinear inputs and perform a complex, nonlinear form of spatial summation.7, 14.

The issue of linear processing in the retina is of great importance to vision researchers: a number of vision models, both retinal and cortical, are based on the assumption that retinal elements behave approximately linearly. However, there are reasons to question the validity of this assumption. For example, it is known that photoreceptor responses are not linearly related to luminance; photoreceptors exhibit temporal adaptation that leads to a static nonlinear compression of the input range.2, 21. Because neural elements have a limited dynamic range, it is important that the photoreceptors be able to compress inputs into an appropriately narrow range, since any information that is lost at the photoreceptors cannot be recovered at later stages. Such a static nonlinearity should compromise linear spatial summation in the X cell, as diagrammed in Fig. 2b.

In addition to this static, compressive nonlinearity, it is known that photoreceptors and bipolar cells exhibit transient overshoots and undershoots to sudden input changes. These transients are the result of each photoreceptor trying to adjust its dynamic range in response to local changes in input intensity. However,
Figure 3: (a): Schematic of a one-dimensional feed-forward, distance-dependent shunting network. Membrane potential of each cell (gray and black disks) is depolarized by inputs falling within a small central area, and is hyperpolarized by inputs falling within a broader surrounding area. The curves describe the decrease in sensitivity of each neuron to excitatory (solid curve) and inhibitory (dashed curve) inputs with increasing distance from the neuron’s receptive field center. (b): Response of a shunting cell (lower trace), whose membrane potential obeys Eq. 1, to sinusoidal temporal modulation of an input pattern (upper trace) when passive decay is large relative to other cell parameters. Response in the presence of a dim background (solid lines) demonstrates poor temporal processing. Moreover, the large passive decay compromises the cell’s ability to discount uniform backgrounds, as shown by the saturating response to the same input in the presence of a more intense background (dashed lines). In this and subsequent figures, response units are arbitrary, and are assumed proportional to membrane potential as given by Eq. 1.

Membrane dynamics

The solution to this puzzle arises from a general model of spatiotemporal processing, based on a network of neurons that obey membrane, or shunting dynamics. Feed-forward shunting interactions are described by a first order differential equation equivalent to the classical membrane equation used by Hodgkin:

\[ C \frac{dv}{dt} = (v^p - v)g^p + (v^+ - v)g^+ + (v^- - v)g^- \]  

(1)

The constant terms \( v^p, v^+, \) and \( v^- \), respectively (with \( v^+ > v^p > v^- \)), represent passive, depolarized, and hyperpolarized equilibrium potentials, while the constant \( C \) represents membrane capacitance. Inputs modulate membrane conductance (\( g^+ \) and \( g^- \)) of two ionic species that have opposite effects on membrane potential \( v \). In the absence of inputs (\( g^+ = g^- = 0 \)), membrane potential decays to the resting potential \( v^p \) at a constant rate \( g^p \). Increases in conductance drive the membrane potential toward the corresponding equilibrium potential. An on-center, off-surround receptive field structure is incorporated into this equation.
by assuming that (excitatory) inputs falling within a small central area increase the depolarizing conductance, whereas (inhibitory) inputs falling within a broader area increase the hyperpolarizing conductance (Fig. 3a).

Equation 1 is defined as a linear time-varying differential equation: it is linear because it contains only linear factors of the dependent variable $v$; it is time-varying because $v$ is multiplied by conductance terms that are time-dependent functions of the input. In spite of the lack of nonlinear factors of the dependent variable, linear time-varying systems behave nonlinearly because the effect of a given input depends on the state of the system as a result of the multiplicative relationship between the input and the dependent variable. For example, inputs that cause an equal change in the depolarizing ($g^+$) and hyperpolarizing ($g^-$) conductance terms of Eq. 1 will not cancel out because they are multiplied, respectively, by the terms $(v^+ - v)$ and $(v^- - v)$, which are generally unequal.

Equation 1 embodies a number of important properties: membrane potential is bounded between the equilibrium potentials $v^+$ and $v^-$ regardless of input size\(^2\), \(^5\), \(^26\). The nonlinear interaction between center and surround results in a form of gain control that can retune the cell’s steady-state sensitivity\(^2\), \(^3\), \(^25\). It is possible to show that such a network, at equilibrium, exhibits spatial adaptation and responds to luminance ratios (contrast) across space while ignoring uniform changes in background luminance\(^27\).

**Space-time tradeoff**

In order to achieve desirable spatial information processing at equilibrium, the shunting network sacrifices its ability to respond to certain types of temporal modulation\(^3\). This irreducible tradeoff results from the time-varying nature of Eq. 1 coupled with the nonnegative nature of visible light (light is nonnegative in the sense that photoreceptor responses depend on the number of absorbed photons, which cannot be a negative quantity). Briefly, when inputs are increased from zero, the membrane potential in Eq. 1 changes at a rate that depends on the size of the inputs, so that larger inputs lead to faster changes in membrane potential. On the other hand, reducing all inputs to zero leaves the membrane potential to decay passively at a rate that only depends on the decay constant $g^P$ (Fig. 3b). No evidence for such a limitation is found experimentally in the undamaged early visual system, a fact that seems to cast doubt on the validity of the shunting network as a model of biological visual processing mechanisms.

The space-time tradeoff can be circumvented, however, by requiring that incoming signals be duplicated into two equal and opposite, or push-pull pathways before reaching a shunting neuron: one pathway, whose activation increases with increasing inputs, modulates the depolarizing conductance of the target neuron; the other, whose activation increases with decreasing inputs, modulates the hyperpolarizing conductance of the target neuron\(^3\). Fig. 4 shows that addition of a push-pull mechanism enables the shunting network to follow input increments and decrements equally well.

The existence of pathways carrying equal and opposite signals requires that the input first be compressed to a finite range\(^3\), \(^5\). A preprocessing mechanism for temporal adaptation, i.e., one that adjusts to input changes over time, can provide adequately compressed signals in response to arbitrary inputs.

Hence a network of neurons subject to membrane dynamics can remain sensitive to temporal as well as spatial input modulation if it is embedded within a three layer architecture: an initial, spatially localized mechanism of temporal adaptation compresses several orders of magnitude of visual inputs into a narrow range of neural signals. These signals are transmitted through a second layer consisting of cell pairs, one depolarized and the other hyperpolarized by increasing inputs. A third layer of cells collects push-pull inputs over an extended spatial area. This architecture, illustrated in Fig. 4a, is analogous to the three layer structure of the vertebrate retina, which includes photoreceptors, ON and OFF bipolar cells, and ganglion cells\(^2\).

The architecture depicted in Fig. 4a differs from the classical view, which holds that ON and OFF bipolar cells feed separately into ON and OFF ganglion cells\(^2\), \(^28\). The push-pull mechanism can give rise to either ON or OFF ganglion cell classes even though bipolar cells of both polarities converge onto a single ganglion
Figure 4: (a): Schematic of the push-pull mechanism. Light signals \([I(t)]\) impinging upon each photoreceptor cause equal and opposite responses in two bipolar cells of opposite polarity \([R(t)\) and \(M - R(t)\), where \(M\) is the maximum photoreceptor activation]. These cells in turn activate depolarizing \((g^+)\) and hyperpolarizing \((g^-)\) membrane conductances in a single ganglion cell, whose membrane potential \([v(t)]\) is bounded between depolarizing \((v^+)\) and hyperpolarizing \((v^-)\) equilibrium potentials. Similar push-pull pairs occur throughout the cell’s receptive field. (b): Response of a push-pull shunting neuron (lower trace) to sinusoidal temporal modulation of an input pattern (upper trace) equal to the input shown in Fig. 3b. The push-pull cell is able to follow input modulation regardless of the background intensity. Inputs and responses in arbitrary units, plotted on the same scale as Fig. 3b for comparison.

cell\(^5\). Evidence for simultaneous (push-pull) modulation of depolarizing and hyperpolarizing ganglion cell conductances has been found in some preparations\(^{29,30,31,32}\).

The preceding observations suggest that the ubiquitous three layer structure of the retina evolved in order for biological organisms to respond to spatial and temporal distributions of a nonnegative input signal. This hypothesis is strengthened in the remainder of the article by showing that the push-pull mechanism endows the network with other useful properties, and that the model is able to simulate physiological data on both X and Y retinal ganglion cells.

**Linearity and adaptation**

The need for a push-pull network was based on mathematical analyses showing a tradeoff between accurate processing in the temporal and spatial domains. A similar tradeoff exists between the relative merits of linear and nonlinear processing.

Linearity is important for the process of encoding light into neural signals without loss of useful information. From a mathematical standpoint, computational models of visual processing are greatly simplified when linearity is assumed\(^{33}\). However, the assumption of linearity is restricting, and frequently inaccurate in the context of biological systems.

Nonlinearity can be used to improve a system’s performance. For instance, the use of a threshold can remove noise from a signal, while a compressive nonlinearity can encode a broad range of inputs into a narrow output range. In many cases, however, nonlinear processing introduces unwanted distortion factors
and removes useful information, thus corrupting the signal that reaches subsequent processing stages. From a practical standpoint, nonlinearity usually increases the complexity of a computational model, often making it analytically intractable.

Biological visual processing systems—and the computational models that simulate them—must balance the power of nonlinear processing with the advantages of linearity. A good example of this tradeoff is found in the transduction of light into neural signals carried out by the retina: the range of light intensity that could be coded accurately if the retina were linear is limited by the dynamic range of retinal cells and by the presence of noise. Spatial and temporal adaptation instead allow retinal cells to encode several orders of magnitude of light intensity within their rather limited dynamic range.

Spatial and temporal adaptation instead allow retinal cells to encode several orders of magnitude of light intensity within their rather limited dynamic range. While certain types of adaptation can be effected with a linear system, there is abundant evidence for nonlinear adaptation mechanisms acting in the retina, beginning as early as the outer segment of photoreceptors. The existence of nonlinearities in the first stages of retinal processing is balanced by the apparent linear spatial summation observed in X cell responses to modulated sinusoidal gratings. Somehow the transformations occurring between photoreceptors and ganglion cells must restore some linearity in retinal processing.

Mathematical analysis has shown that the same push-pull mechanism that is necessary for accurate spatial and temporal processing is also responsible for the linear spatial summation of X cells. Briefly, the convergence of two equal and opposite signals onto the depolarizing and hyperpolarizing conductance terms in Eq. 1, reduces the membrane equation to a linear, time-invariant (LTI) system.

The reduction of Eq. 1 to an LTI system is advantageous for several reasons. First, as discussed above, such a system is able to follow input increments and decrements equally well even when the inputs are strictly nonnegative. Second, the response of an LTI system to a combination of static and superimposed modulated input components consists of a sustained response component, whose amplitude only depends on the amplitude of the static input component, and a modulated response component, whose amplitude only depends on the amplitude of the modulated input component. In other words, an LTI system segregates the transient and sustained components of an input signal.

A similar result holds when the input is preprocessed by a nonlinear mechanism, in which case, however, some nonlinearity is present in the output of the push-pull mechanism. The mathematical form of the push-pull system makes it possible to control independently the network's sensitivity to transient and sustained components of the (possibly nonlinear) input. In agreement with the anatomical data linking beta cells to X cells and alpha cells to Y cells, an increase in the radius of a cell's receptive field center relative to the surround can shift the cell's response from X-like (sustained, linear) to Y-like (transient, nonlinear). Specifically, simulated X cells exhibit primarily sustained responses, and null responses to modulated gratings at ±90° spatial phase independent of contrast and average illuminance. Simulated Y cells exhibit large transient responses, on-off responses at ±90° for gratings of low spatial frequency, and on-off responses at all values of relative spatial phase for gratings of high spatial frequency. These results are shown in Fig. 5.

The prediction that the same neural architecture can account for the behavior of both X and Y cells represents a striking departure from the classical view, which holds that X and Y cells are formed by different receptive field mechanisms. By refuting the a priori assumption of linear preprocessing and taking advantage of two stages of nonlinearity, the present model is able to capture the fundamental aspects of linear and nonlinear processing characteristics of both X and Y ganglion cells. The ability to modify the network's behavior by simple parametric changes suggests that the model could simulate other classes of ganglion cells, including those that give rise to the parvocellular and magnocellular processing streams in the primate.
Figure 5: Simulation results from numerical integration of push-pull network including nonlinear photoreceptors, push-pull bipolar cells, and X and Y ganglion cells that obey membrane dynamics. The primary difference between simulated X and Y cells is width of receptive field center, which is three times larger for Y than X cells. Time and response scales, and all plotting conventions are matched to those of Fig. 1. Parameters have been reported elsewhere. (a): Simulation of null test for X and Y cells (see Fig. 1a). (b) Population responses for X cells (middle graph) and Y cells (bottom graph) stimulated by square-wave temporal modulation of sinusoidal grating (top graph). Each one-dimensional population is spatially aligned along x axis. Time evolves along t axis. Amplitude of input [I(t), top] and ganglion cell responses [v(t), middle and bottom] are represented along the vertical axis (arbitrary units). The thick solid and dashed lines trace input and responses for spatial phases of 0° and 270°, respectively. These lines correspond to individual traces in part (a), to help visualize the phase-dependent null responses in X cells and on-off responses in Y cells. (c) Same as part (b), but spatial frequency is doubled. X cells respond as before, with null responses at 0° and 270°. Y cells exhibit on-off responses independent of spatial phase. This result is due to the photoreceptor nonlinearity carried through push-pull bipolar cells. The observation that Y cell responses in the cat are independent of spatial phase was the basis for the proposal of small rectifying subunits located throughout the Y cell receptive field.
Toward a unified theory

The push-pull model provides a unified theoretical explanation of several previously unexplained aspects of retinal processing. This article has focused on anatomical and physiological data, but the model can also be tested and expanded on the basis of pharmacological and psychophysical data\textsuperscript{5, 40}.

The constraints that guided the development of this model are not restricted to retinal processing, and apply to other cases where neural populations obeying membrane dynamics must process nonnegative spatiotemporal inputs. This condition occurs for instance whenever neurons collect inputs that are encoded presynaptically by axonal spiking rate. For example, ON and OFF retinal ganglion cells, which use action potentials to carry information extraretinally, should converge in a push-pull fashion onto simple and complex cortical cells, as suggested by recent experimental and analytical results\textsuperscript{28, 41}.

The model’s ability to exhibit diverse functional behaviors within a unified anatomical structure suggests that different classes of cells in the retina and in later stages of the vertebrate visual system may have evolved as variants of a single general-purpose neural mechanism for spatiotemporal processing.

Acknowledgments

The author wishes to express his gratitude to Ennio Mingolla and Michael Cohen for useful comments on an earlier version of this manuscript.
References

1. Ramón y Cajal, S. La cellule 9, 119–257 (1892).


