

The Visual Cortex as a Hierarchical Predictor

Rajesh P.N. Rao* and Dana H. Ballard

Department of Computer Science, University of Rochester,

Rochester, NY 14627-0226, USA

E-mail: {rao,dana}@cs.rochester.edu

Technical Report 96.4

National Resource Laboratory for the Study of Brain and Behavior

Department of Computer Science, University of Rochester

September 1996

(Revised December 1996)

Keywords: Computational Neurobiology, Visual Cortex, Feedback, Prediction, Learning.

*To whom correspondence should be addressed.

A fundamental feature of the visual cortex is the reciprocity of connections between its many distinct areas. Using the principle of Kalman filtering from classical control theory, we describe how this reciprocity may allow the cortex to function as a hierarchical predictor. Feedback connections in this model carry predictions of lower level neural activities, while feedforward connections convey to the higher level the differences between the predictions and the actual lower level activities. These differences allow the higher level to maintain optimal estimates of current visual recognition state and on a longer time scale, enable it to learn a dynamic internal model of the visual environment. A simulated Kalman filter network embodying these principles produced response properties that correlated closely with those of neurons in the primary visual cortex.

The remarkable uniformity and regularity in the structure of the neocortex has prompted an intense search over the past decade for the general computational mechanisms underlying cortical organization. Numerous models have been proposed to explain various aspects of cortical function [1, 2, 3, 4, 5], but a general computational principle has remained elusive. In this report, we suggest that an attractive candidate for this purpose is the statistical principle of *Kalman filtering* from classical control theory [6]. We describe how this principle allows the visual cortex to be modeled as a hierarchical predictor. Our model is motivated by two important properties of the cortex: (a) its distinctive laminar input-output structure, and (b) the reciprocity of connections between its constituent areas: if area A projects to area B, then area B almost invariably projects to area A [7, 8]. While cortico-cortical feedforward connections have been well-studied, the importance of the corresponding feedback connections has also been recently demonstrated in several neurophysiological experiments showing that neurons in lower areas such as striate cortex are heavily influenced by feedback from higher areas [9, 10]. An important functional role for feedback emerges if we posit that the feedback signals carry predictions of lower level neural activities from higher, more abstract areas [4]. The feedforward connections then need only convey the differences (or residuals [5, 28]) between the current activity and its prediction from the higher

area. The computationally attractive idea of encoding differences between an input signal and its prediction from a preexisting internal model is captured succinctly by the statistical construct of the Kalman filter [11].

The Kalman filter is a linear dynamical system that attempts to mimic the dynamic behavior of an observed natural process. It does so by calculating, at each time instant, an optimal estimate of the current state of the observed process, where optimality is defined in terms of maximizing the posterior probability of the state given the observed data [12]. The optimal state estimate is used to predict the next expected input using an internal model of the observed process. Given the next input, the filter computes the difference (or *residual error*) between its prediction and the actual input, and uses this residual to correct its estimate of the state. This involves (a) multiplying the residual error by a gain term that reflects the uncertainties in the internal model and the input measurement process, and (b) adding this weighted residual to the previous state estimate (see [13]). The new corrected estimate is then used to predict the next state, thereby completing one full cycle of filter operation. A small residual error in the Kalman filter implies that the input stimulus has been correctly predicted and recognized. A large residual, on the other hand, implies that the stimulus is possibly novel and therefore worthy of further attention.

Figure 1A depicts a neural network that can implement the Kalman filter outlined above. The canonical operation required by the Kalman filter is of the type: Av , where A is a matrix and v is a vector. In the standard neural implementation of this type of an operation [14], the matrix A represents the synaptic strength of neurons (each row represents the synapses of one neuron) and the components of the vector v denote the pre-synaptic inputs to these neurons. Each neuron computes a weighted sum of its pre-synaptic inputs according to the weights encoded by its synapses. Figure 1B illustrates this operation for the matrix U and the vector \bar{r} . The axonal output vector $U\bar{r}$ in this case forms the Kalman filter’s prediction of the next expected input \mathbf{I} . The subsequent residual error $(\mathbf{I} - U\bar{r})$ that is required by the filter is generated by predictive inhibition of the input \mathbf{I} as shown in Figure 1A. This residual error signal is then successively processed by neurons representing the “bottom-up” gain matrix G_{bu} , the feedforward matrix W , and the

normalization matrix N before being integrated by the prediction neurons V , which generate the next predicted state $\bar{\mathbf{r}}$ as their axonal output (see Figure 1A). The predicted state is transmitted to the lower level by the feedback neurons, whose synapses U transform the higher level state $\bar{\mathbf{r}}$ into the lower level signal $U\bar{\mathbf{r}}$. This signal, which is at the same abstraction level as the lower level inputs, is then used to inhibit the next set of incoming signals \mathbf{I} .

We have developed a hierarchical form of the Kalman filter that allows both bottom-up signals from a lower level as well as “top-down” signals from a higher level to influence the state estimates being calculated at the current level [15]. As in the standard Kalman filter described above, the top-down signals \mathbf{r}_{td} from a higher level carry predictions of the lower level state $\bar{\mathbf{r}}$. These predictive signals generate the top-down residual error $(\bar{\mathbf{r}} - \mathbf{r}_{td})$ which is successively multiplied by the “top-down” gain matrix G_{td} and the normalization matrix N (see Figure 1A) before being integrated with the bottom-up weighted residual to generate the new optimal state estimate [16]. The optimal state estimate can be made robust to occlusions and other forms of noise in the input channels by allowing the gains G_{td} and G_{bu} to be *non-linear* functions of their respective residuals (for example, see [17]). In such a setting, the correlations between the various components of the residual vectors $(\bar{\mathbf{r}} - \mathbf{r}_{td})$ and $(\mathbf{I} - U\bar{\mathbf{r}})$ specify how the elements of the corresponding gain matrices G_{td} and G_{bu} should be set, thereby determining the degree of lateral interactions between neighboring neurons in the Kalman filter network.

The visual cortex is well-suited to implement a hierarchical form of the Kalman filter, given its roughly hierarchical organization and its distinctive laminar input-output structure [7, 8]. The filter in this case would compute optimal state estimates of visual events occurring in the distal environment, thereby enabling the cortex to recognize these events and predict future events based on a hierarchical internal model of the visual environment. The internal model, which is learned during exposure to the environment (see [18]), is jointly encoded by the Kalman filter parameters U , W , and V , which correspond to the modifiable synapses of neurons in the Kalman filter network. Figure 1D shows how a neural implementation of the different components of a hierarchical Kalman filter can fit comfortably within the laminar structure of a given cortical area.

In the case of the primary visual cortex (V1), the model assumes that the feedback connections from V1 to V1’s thalamic input site, the dorsal lateral geniculate nucleus (dLGN), convey predictions of the expected activities of dLGN relay neurons. These predictive signals (denoted by $U\bar{\mathbf{r}}$ in Figure 1) are assumed to inhibit, via inhibitory interneurons, the activities of their corresponding dLGN neurons to the extent that the predictions match the afferent retinal signals. The resulting *residual* activities of the dLGN neurons are conveyed to the cortex via the feedforward thalamo-cortical pathway. On their way to the cortex, the residuals are modulated by a “bottom-up” gain (analogous to the gain G_{bu} in the Kalman filter) which, for example, could be implemented by the inhibitory neurons of the thalamic reticular complex (including the perigeniculate nucleus) [19]. The modulated residual signal is then linearly filtered through the synapses of layer 4 cells in V1, which would correspond to the feedforward weights W in the Kalman filter (see Figure 1A). The signal subsequently undergoes a normalization (corresponding to the matrix N in the Kalman filter). A plausible site for this normalization is layer 2+3 (composed of layers 2 and 3), given that this layer receives a substantial projection from layer 4 [20, 21]. Normalization of responses has also been proposed by other authors to explain saturation of neural responses in V1 at high stimulus contrasts [22]. The preceding sequence of neural processing suffices to compute the weighted Kalman residual $NWG_{bu}(\mathbf{I} - U\bar{\mathbf{r}})$, which is used to correct the previous state estimate $\bar{\mathbf{r}}(t)$ at time instant t (see [13]). The corrected estimate then generates the next state prediction $\bar{\mathbf{r}}(t + 1)$ via the synapses V . Neurons in layer 5 appear to be ideally located for such a computation, given that layer 2+3 cells project extensively into layer 5 [20, 21]. Furthermore, layer 5 projects to layer 6, which is known to be a major source of cortico-thalamic feedback [20]. Layer 6 neurons could therefore synaptically encode the feedback weights U of the Kalman filter and convey the feedback signal $U\bar{\mathbf{r}}(t + 1)$ to the dLGN for predictive inhibition of the next input $\mathbf{I}(t + 1)$. Support for such an operation comes from the work of Murphy and Sillito [9] indicating that corticofugal feedback engages inhibitory mechanisms in the dLGN.

A final issue is the role of cortico-cortical feedback from a higher visual area, for example, V2. As in the case of cortico-thalamic feedback, the model assumes that the higher area (V2) conveys

predictions of the state at the lower area (V1). Just as in the bottom-up case, the predictive signals from V2 are assumed to generate residuals ($\bar{\mathbf{r}} - \mathbf{r}_{td}$) via interactions with inhibitory interneurons in the superficial layers. These residual activities are filtered by cells in layer 2+3, whose synapses would encode the top-down gain matrix G_{td} . The filtered top-down residuals $G_{td}(\bar{\mathbf{r}} - \mathbf{r}_{td})$ are conveyed to two sites: (a) via axon collaterals to the layer 2+3 “normalization” neurons N , which integrate these top-down residuals with the bottom-up residuals $WG_{bu}(\mathbf{I} - U\bar{\mathbf{r}})$ from layer 4, and (b) to the higher visual area (V2), where this signal serves as the bottom-up weighted residual for the higher level Kalman filter (see Figure 1C and 1D).

In order to validate the model, we trained a three-level hierarchical network (Figure 1C) on a sample of natural images as shown in Figure 2A. For each input, the network was allowed to converge according to the Kalman filter dynamics described above, and the feedforward, feedback and prediction synapses W , U and V were adapted to further minimize the prediction error using Kalman filter-based update equations, operating on a slower time scale at the synaptic level (see [18]). Upon convergence of the synapses, we examined the response properties of neurons situated in the different laminae of the model and compared these properties to those observed in the mammalian visual cortex. A surprisingly large class of model neurons (87% in layer 2+3) exhibited the classical phenomenon of *endstopping*: neurons that respond vigorously to a bar have diminished responses as the bar extends beyond the classical receptive field [23].

Endstopping has previously been characterized as a feedforward effect, caused mainly by lateral inhibition from neighboring cells [3, 23]. When viewed as a purely feedforward effect, endstopping appears complex and highly non-linear, and has been interpreted as arising due to the brain’s need to represent curvature [3] (see Figure 4). Results from our simulations however suggest a more general interpretation of “extra-classical” effects such as endstopping in terms of feedback and predictive encoding. Our interpretation is supported by recent neurophysiological experiments in V1 showing that extra-classical effects in layer 2+3 neurons often manifest themselves only 80-100 milliseconds after stimulus onset, strongly suggesting a role for higher level feedback in mediating these effects [24]. We tested this hypothesis by examining the response properties of layer 2+3

neurons in level 1 (corresponding to V1) of the Kalman filter network that was trained on natural images (Figure 2). Note that a class of model neurons in layer 2+3 compute the filtered residual $G_{td}(\bar{\mathbf{r}} - \mathbf{r}_{td})$. As shown in Figure 3A, the typical response of such a model neuron (solid line) drops off sharply as a test bar extends beyond its classical receptive field (RF). The attenuation in response can be attributed to diminishing residuals $(\bar{\mathbf{r}} - \mathbf{r}_{td})$ caused by better predictions \mathbf{r}_{td} from level 2 of the state $\bar{\mathbf{r}}$ at level 1, as stimulus length is incrementally increased up to the size of the larger RF of the level 2 model neurons (see Figure 2D and 2E for an example). Such an explanation is supported by the RF profiles of level 2 neurons (Figure 2C), some of which appear to be tuned towards long line segments. Further confirmation of the role of feedback in endstopping was obtained by disabling the feedback connections from level 2 to level 1, which eliminated endstopping in most of the level 1 model neurons (Figure 3B and 3C). Thus, a general prediction of the model is that inactivation of a higher cortical area such as V2 or MST should cause substantial disinhibition and elimination of extra-classical effects such as endstopping in layer 2+3 neurons in lower areas such as V1 or MT respectively. Preliminary results from experiments involving the inactivation of V2 appear to lend some support to these predictions [10].

In summary, our results show that many of the apparently complex, extra-classical responses of cortical neurons can be parsimoniously explained by considering the interactions between feedforward and feedback signals, as captured by the functional responses of model neurons in the different laminae of the hierarchical Kalman filter network. Thus, rather than characterizing visual cortical neurons unconditionally as feature detectors, the model suggests that it may be useful to view some of them, especially those conveying feedforward signals to a higher area (such as the layer 2+3 neurons), as *feature-difference* or *residual detectors*. These residuals play a crucial role in enabling the hierarchical Kalman filter network to continually adapt the activities of its constituent neurons to accurately predict incoming stimuli. Simultaneously, but at a slower rate, the residuals also drive the adaptation of neuronal synapses in the network, thereby further minimizing the errors in prediction. In doing so, the network learns a hierarchical and distributed internal model of its environment that can be used to efficiently subserve the larger cognitive goals of the organism.

References

- [1] H.G. Barrow, *Proc. IEEE Int. Conf. on Neural Networks*, 115 (1987); R. Linsker, *Computer* **21(3)**, 105 (1988); J.S. Albus, *IEEE Trans. Sys., Man, and Cyb.* **21(3)**, 473 (1991); P.J.B. Hancock, R.J. Baddeley, and L.S. Smith, *Network* **3**, 61 (1992); A.P. Pentland, in *Neural Networks for Vision and Image Proc.*, G.A. Carpenter and S. Grossberg, Eds. (MIT Press, Cambridge, MA, 1992), pp. 133-159; C.C. Law and L.N. Cooper, *Proc. Natl. Acad. Sci. USA* **91**, 7797 (1994); T.A. Poggio and A. Hurlbert, in *Large-Scale Neuronal Theories of the Brain*, C. Koch and J.L. Davis, editors (MIT Press, Cambridge, MA, 1994), pp. 153-182; S. Ullman, *ibid.*, pp. 257-270; D.C. Van Essen, C.H. Anderson, and B.A. Olshausen, *ibid.*, pp. 271-299; D.J. Heeger, E.P. Simoncelli, and J.A. Movshon, *Proc. National Acad. Sciences* **93**, 623 (1996).
- [2] B.A. Olshausen and D.J. Field, *Nature* **381**, 607 (1996).
- [3] A. Dobbins, S.W. Zucker, and M.S. Cynader, *Nature* **329**, 438 (1987).
- [4] U. Grenander, *Lectures in Pattern Theory I, II and III* (Springer-Verlag, Berlin, 1976-81); G.M. Edelman, in *The Mindful Brain*, V.B. Mountcastle and G.M. Edelman, Eds. (MIT Press, Cambridge, MA, 1978) pp. 51–100; G.A. Carpenter and S. Grossberg, *Computer Vision, Graphics, and Image Processing* **37**, 54 (1987); E. Harth, K.P. Unnikrishnan, and A.S. Pandya, *Science* **237**, 184 (1987); M. Kawato, H. Hayakawa, and T. Inui, *Network* **4** 415 (1993); G.E. Hinton, P. Dayan, B.J. Frey, and R.M. Neal, *Science* **268**, 1158 (1995); W.R. Softky, in *Advances in Neural Info. Proc. Systems* 8, D. Touretzky, M. Mozer, and M. Hasselmo, Eds. (MIT Press, Cambridge, MA, 1996), pp. 809–815.
- [5] D. Mumford, in *Large-Scale Neuronal Theories of the Brain*, C. Koch and J.L. Davis, editors (MIT Press, Cambridge, MA, 1994), pp. 125-152; H. Barlow, *ibid.*, pp. 1–22; D.M. MacKay, in *Automata Studies*, C.E. Shannon and J. McCarthy, Eds. (Princeton University Press, Princeton, NJ, 1956).

- [6] A.E. Bryson and Y.-C. Ho, *Applied Optimal Control* (John Wiley and Sons, NY, 1975); P.S. Maybeck, *Stochastic Models, Estimation, and Control* (Academic Press, NY, 1979).
- [7] K.S. Rockland and D.N. Pandya, *Brain Res.* **179**, 3 (1979); D.J. Felleman and D.C. Van Essen, *Cerebral Cortex* **1**, 1 (1991).
- [8] D.C. Van Essen, in *Cerebral Cortex* **3**, A. Peters and E.G. Jones, Eds. (Plenum, NY, 1985), pp. 259–329; J.H.R. Maunsell and W.T. Newsome, *Annual Review of Neuroscience* **10**, 363 (1987); R. Desimone and L.G. Ungerleider, in *Handbook of Neuropsychology* **2**, F. Boller and J. Grafman, Eds. (Elsevier, NY, 1989), pp. 267–299.
- [9] J.H. Sandell and P.H. Schiller, *Journal of Neurophysiology* **48(1)**, 38 (1982); P.C. Murphy and A.M. Sillito, *Nature* **329**, 727 (1987); M. Mignard and J.G. Malpeli, *Science* **251**, 1249 (1991).
- [10] A.C. James, J.M. Hupe, S.L. Lomber, B. Payne, P. Girard, and J. Bullier, *Soc. Neurosci. Abstract* **21**, 904 (1995).
- [11] R.E. Kalman, *Trans. ASME J. Basic Eng.* **82**, 35 (1960); R.E. Kalman and R.S. Bucy, *Trans. ASME J. Basic Eng.* **83**, 95 (1961).
- [12] The Kalman filter assumes that the observed dynamical process can be modeled as:

$$\mathbf{I}(t) = U\mathbf{r}(t) + \mathbf{n}_{bu}(t) \quad (1)$$

where $\mathbf{r}(t)$ is the actual hidden state of the process at time instant t , $\mathbf{I}(t)$ is the observable output (for example, an image), U is a generative (or feedback) matrix that relates the hidden state to the observable output, and $\mathbf{n}_{bu}(t)$ is a “bottom-up” Gaussian noise process with mean zero and covariance $\Sigma_{bu} = E[\mathbf{n}_{bu}\mathbf{n}_{bu}^T]$ (E denotes the expectation operator and T denotes transpose). In addition, the state \mathbf{r} follows the dynamics:

$$\mathbf{r}(t) = V\mathbf{r}(t-1) + \mathbf{n}(t-1) \quad (2)$$

where V is the *state transition (or prediction) matrix* and \mathbf{n} is a Gaussian noise process with mean $\bar{\mathbf{n}}(t)$ and covariance $\Sigma(t) = E[(\mathbf{n}(t) - \bar{\mathbf{n}}(t))(\mathbf{n}(s) - \bar{\mathbf{n}}(s))^T]$. The filter minimizes an optimization function J given by:

$$J(\mathbf{r}) = (\mathbf{I} - U\mathbf{r})^T \Sigma_{bu}^{-1} (\mathbf{I} - U\mathbf{r}) + (\mathbf{r} - \bar{\mathbf{r}})^T M^{-1} (\mathbf{r} - \bar{\mathbf{r}}) \quad (3)$$

where $\bar{\mathbf{r}}$ is a prior estimate of \mathbf{r} and M is the associated covariance matrix (see below). Minimizing J is equivalent to maximizing the posterior probability of \mathbf{r} given input data \mathbf{I} [6].

[13] Using the dynamic model described in [12], the Kalman filter provides an optimal estimate $\hat{\mathbf{r}}$ of the true hidden state \mathbf{r} at time instant t :

$$\hat{\mathbf{r}}(t) = \bar{\mathbf{r}}(t) + NWG_{bu}(\mathbf{I}(t) - U\bar{\mathbf{r}}(t)) \quad (4)$$

where $\bar{\mathbf{r}}(t)$ is the predicted state for time t generated from the prior state estimate $\hat{\mathbf{r}}(t-1)$ using the internal model of the dynamics of the process (see [12] above):

$$\bar{\mathbf{r}}(t) = V\hat{\mathbf{r}}(t-1) + \bar{\mathbf{n}}(t-1) \quad (5)$$

The two equations above can be combined to yield the complete Kalman filter equation:

$$\bar{\mathbf{r}}(t) = V(\bar{\mathbf{r}}(t-1) + NWG_{bu}(\mathbf{I}(t-1) - U\bar{\mathbf{r}}(t-1))) + \bar{\mathbf{n}}(t-1) \quad (6)$$

which can be efficiently implemented in the neural circuit of Figure 1A. The term $K = NWG_{bu}$ is known as the *Kalman gain* and modulates the degree to which the prediction $\bar{\mathbf{r}}(t)$ is corrected by the incoming sensory residual $(\mathbf{I}(t) - U\bar{\mathbf{r}}(t))$. The “bottom-up” gain matrix G_{bu} is the inverse of the covariance matrix Σ_{bu} , the feedforward matrix W is essentially the transpose of U (see [18]), and N is a normalization matrix that maintains the covariance of the estimated state: $N(t) = (M^{-1}(t) + WG_{bu}(t)U)^{-1}$, where $M(t) = VN(t-1)V^T + \Sigma(t-1)$ (see [6] for details).

[14] B. Widrow and S.D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985); P.S. Churchland and T.J. Sejnowski, *The Computational Brain* (MIT Press, Cambridge, MA, 1992).

[15] R.P.N. Rao and D.H. Ballard, *Neural Computation*, (1996), in press; R.P.N. Rao and D.H. Ballard, *Computational Neuroscience '96*, J. Bower Ed. (Plenum, NY, 1997), in press.

[16] The hierarchical Kalman filter minimizes an optimization function based on the minimum description length (MDL) principle by penalizing deviations from a hierarchical stochastic model of the image generation process [15] (see also [25]). It allows a top-down feedback signal $\mathbf{r}_{td} = U\bar{\mathbf{r}}(t)$ from a higher level, and corrects the predicted state estimate $\bar{\mathbf{r}}$ using the bottom-up and top-down residuals $(\mathbf{I} - U\bar{\mathbf{r}}(t))$ and $(\mathbf{r}_{td} - \bar{\mathbf{r}}(t))$:

$$\hat{\mathbf{r}}(t) = \bar{\mathbf{r}}(t) + NWG_{bu}(\mathbf{I} - U\bar{\mathbf{r}}(t)) + NG_{td}(\mathbf{r}_{td} - \bar{\mathbf{r}}(t)) - Ng(\bar{\mathbf{r}}(t)) \quad (7)$$

where $\bar{\mathbf{r}}$ is computed as in [13] above. The residuals are weighted by their respective Kalman gain matrices $K_1 = NWG_{bu}$ and $K_2 = NG_{td}$, where G_{td} is the “top-down” gain matrix and is essentially the inverse of the top-down covariance matrix Σ_{td} [15]. As shown in Figure 1A, the prediction $\bar{\mathbf{r}}$ is conveyed as output to the lower level by multiplying it with the feedback matrix U . The outputs $\bar{\mathbf{r}}$ of spatially adjacent modules at each level are also conjoined and fed as input to the next higher level as shown in Figure 1C. As a result, higher level model neurons predict based on inputs from a larger spatial extent. Consequently, the receptive fields of model neurons become progressively larger as one ascends the hierarchical network, in a manner similar to that observed in the occipitotemporal pathway [8]. The decay term $-Ng(\bar{\mathbf{r}}(t))$ arises from the MDL-based optimization function [15] and penalizes overfitting of data, thereby encouraging the network to generalize to new situations. In addition, one can obtain a temporal hierarchy by using exponentially decreasing decay functions g for successively higher levels; this causes the states at the higher levels to decay at a slower rate, thereby allowing prediction based on longer historical memories at the higher levels. The decay term also causes the network to seek higher-order statistical correlations in the input data as suggested by Olshausen and Field [2] in their sparseness maximization approach, which is a maximum likelihood special case of the Kalman filter framework.

[17] P.J. Huber, *Robust Statistics* (John Wiley and Sons, NY, 1981).

[18] A Hebbian learning rule for adapting the synapses U was obtained by first collapsing the rows of the matrix U into a vector \mathbf{u} and then deriving a Kalman filter update equation similar to the one for \mathbf{r} [15]: $\hat{\mathbf{u}}(t) = \bar{\mathbf{u}}(t) + C_1(\mathbf{I} - \bar{U}\bar{\mathbf{r}}(t)) - C_2\bar{\mathbf{u}}(t)$. This learning rule involves the familiar addition of a correction residual $(\mathbf{I} - \bar{U}\bar{\mathbf{r}}(t))$ to the prior estimate $\bar{\mathbf{u}}(t)$, which is defined to be $\hat{\mathbf{u}}(t - 1)$ plus an additive noise term. The correction residual is weighted by a gain matrix C_1 computed from the state and bottom-up error covariances [15]. The linear decay term $-C_2\bar{\mathbf{u}}(t)$ serves two purposes. First, it acts as a regularizer by penalizing overfitting, thereby increasing the potential for generalization. Second, a similar learning rule for W causes W to converge to the transpose of the matrix U [15]. This allows the network of Figure 1A to implement the Kalman filter update equation described in [16] using W as the transpose of U . A synaptic learning rule similar to the one given above for U may also be formulated for the prediction weights V for time-varying input stimuli [15]. Since the endstopping simulations reported here involved static stimuli, we used $\bar{\mathbf{r}}(t + 1) = \hat{\mathbf{r}}(t)$ and identity matrices for the bottom-up and top-down gains. For a given input, the activities of neurons were updated iteratively according to the Kalman filter in [16] and allowed to stabilize at all levels before updating \mathbf{u} according to the equation above. This combination of estimation of state and synaptic learning of filter parameters can be regarded as an on-line form of the expectation-maximization (EM) algorithm from statistics [26] (see [15] for details).

[19] F. Crick, *Proc. National Acad. Sci. USA* **81**, 4586 (1984).

[20] J.S. Lund, in *The Organization of the Cerebral Cortex*, F.O. Schimdt *et al.*, Eds. (MIT Press, Cambridge, MA, 1981), pp. 105-124; C.D. Gilbert and T.N. Wiesel, *ibid.*, pp. 163-191.

[21] J. Bolz and C.D. Gilbert, *Nature* **320**, 362 (1986); J. Bolz, C.D. Gilbert, and T.N. Wiesel, *Trends in Neurosciences* **12(8)**, 292 (1989).

[22] J.G. Robson, G.C. DeAngelis, I. Ohzawa and R.D. Freeman, *Invest. Ophthalm. Vis. Sci. Supp.* **32**, 429 (1991); D.G. Albrecht and W.S. Geisler, *Vis. Neurosci.* **7**, 531 (1991); D.J. Heeger,

Vis. Neurosci. **9**, 181 (1992); M. Carandini and D.J. Heeger, *Science* **264**, 1333 (1994).

- [23] D.H. Hubel and T.N. Wiesel, *J. Neurophysiol.* **28**, 229 (1965).
- [24] K. Zipser, V.A.F. Lamme, and P.H. Schiller, *J. Neurosci.* **16(22)**, 7376 (1996).
- [25] K.C. Chou, A.S. Willsky, and A. Benveniste, *IEEE Trans. Automatic Control* **39(3)**, 464 (1994); M.R. Luetzgen and A.S. Willsky, *IEEE Trans. Image Proc.* **4(2)**, 194 (1995).
- [26] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, *Annals Math. Stat.* **41**, 164 (1970); A.P. Dempster, N.M. Laird, and D.B. Rubin, *J. Royal Stat. Soc. Series B* **39**, 1 (1977).
- [27] The images were first filtered using a center-surround difference-of-Gaussians operator to approximate processing at the levels of the retina and the LGN (see also [2]). During training, three 16×16 overlapping Gaussian-windowed image patches (offset by 5 pixels horizontally) were fed to the three level 1 modules (see Figure 1C). The responses from the level 1 modules were then conjoined as a single vector and fed to the sole level 2 module. The effective level 2 RF thus encompassed a 16×26 image region spanned by the three overlapping circles as shown in the enlarged level 2 RF in Figure 2A.
- [28] J.G. Daugman, *IEEE Trans. Acoustics, Speech, and Signal Proc.* **36(7)**, 1169 (1988); A.E.C. Pece, in *Artificial Neural Networks 2*, I. Aleksander and J. Taylor, Eds. (Elsevier Science, Amsterdam, 1992), pp. 865–868.
- [29] The neurons that continued to exhibit endstopping without feedback were found to be those whose receptive field orientations were not aligned with that of the bar used for testing. For these neurons, the bar of the shortest length excited part of the receptive field but the bar of the next length, though still within the overall first-level receptive field, partially or fully negated the first response, resulting in a form of endstopping, though not of the extra-classical kind.
- [30] This work was supported by research grants from the National Institute of Health (NIH) and the National Science Foundation (NSF).

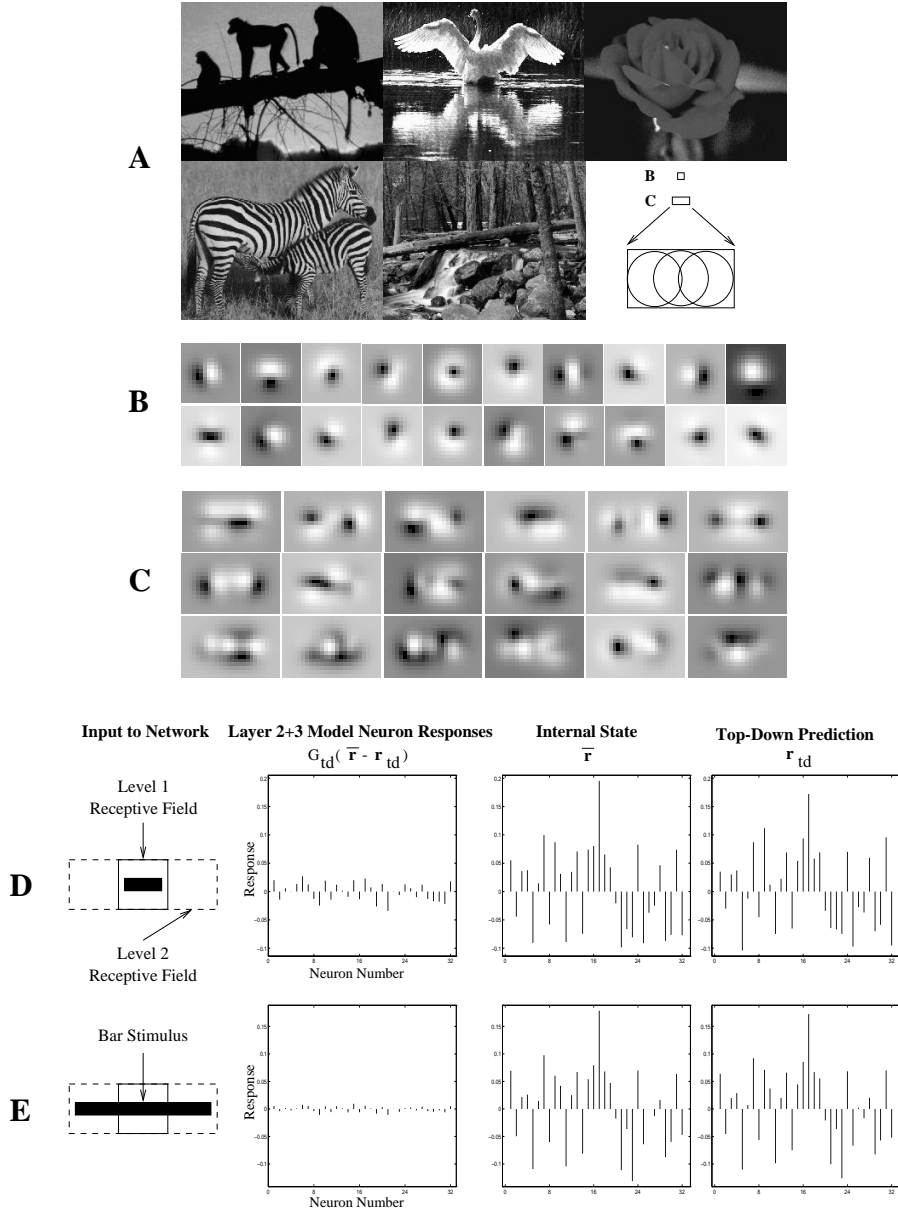


Figure 2: Receptive Fields and Model Neuron Responses Observed after Training. (A) Five natural images used for training the three-level hierarchical network of Figure 1C (see [27]). The two upper boxes in the bottom right corner show relative sizes (16×16 and 16×26 pixels) of level 1 and level 2 receptive fields (RFs). (B) Learned synaptic weights (RF weighting profiles) of 20 of the 32 layer 4 model neurons in the level 1 KF module analyzing the central image region. Flanking image regions were analyzed by two other level 1 modules (Figure 1C), each with 32 model neurons. Values for these synapses, which form rows of W , can be positive or negative (positive values = bright regions; negative values = dark regions). These RF profiles resemble classical oriented edge/bar detectors characteristic of simple cells [23], previously modeled using difference of offset Gaussians and Gabor functions [3, 28]. (C) RF profiles of 18 of the 128 layer 4 model neurons in the level 2 KF module. (D) Responses of the 32 layer 2+3 model neurons in the central level 1 module to a dark bar (positive values = upward bars; negative values = downward bars). Positive and negative values may be coded by separate neurons in the cortex). (E) Increase in top-down prediction accuracy as the bar extends beyond the classical RF (solid box), up to the size of the level 2 RF (dashed box) causes a reduction in the level 1 top-down residuals, which manifests itself as endstopping in the layer 2+3 model neurons.

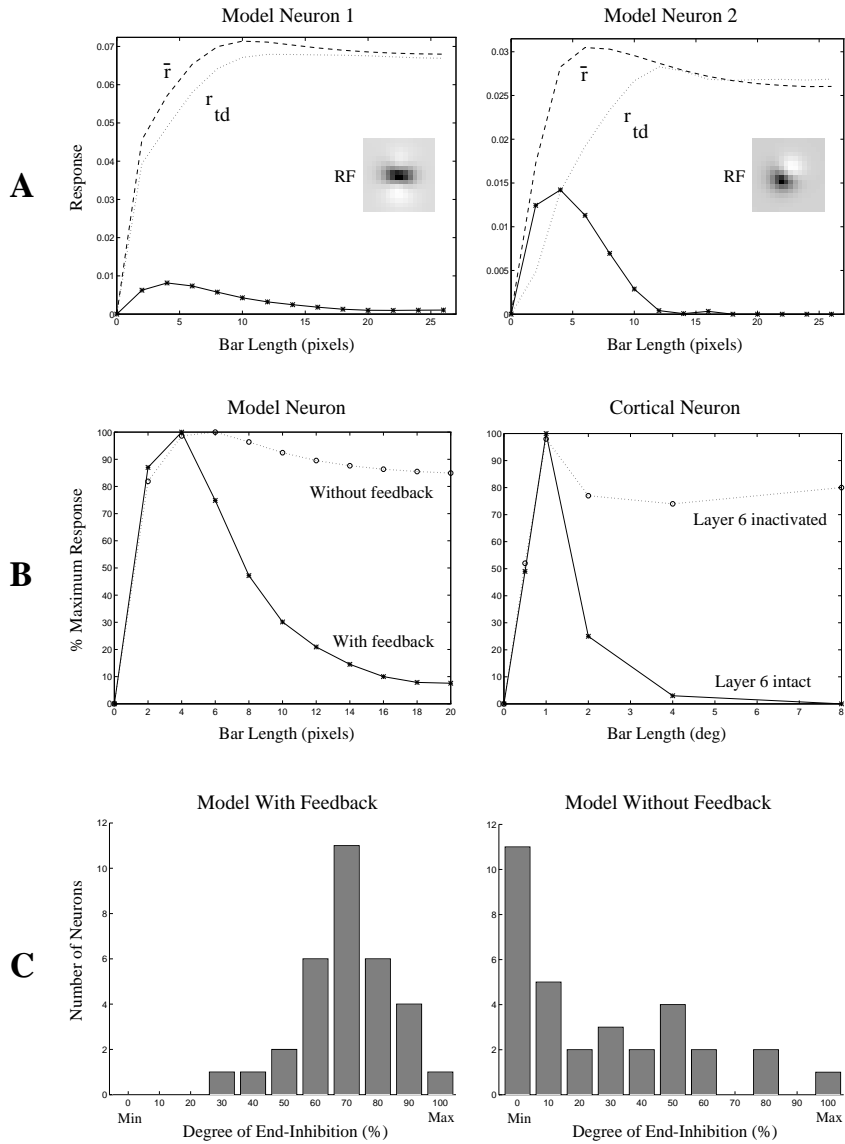


Figure 3: Endstopping Results. (A) Length tuning curves for two layer 2+3 model neurons at level 1 with even-symmetric (left) and odd-symmetric (right) RF profiles. Both model neurons exhibit the decrease in response characteristic of endstopping as the bar extends beyond the classical RF. The dashed line represents the corresponding activity \bar{r} of a layer 5 model neuron, and the dotted line represents the top-down (inhibitory) prediction r_{td} from the corresponding layer 6 neuron in the level 2 module. (B) Effect of inactivating feedback from level 2. Plotted on the left are the length tuning curves for a layer 2+3 model neuron at level 1 with and without feedback from level 2 (solid and dotted line respectively). Tuning curves for a layer 2+3 complex cell in cat striate cortex (V1) reported by Bolz and Gilbert [21] are plotted on the right for comparison. Disabling top-down feedback eliminated endstopping in the model neuron in a manner qualitatively similar to that observed in the cortical neuron after layer 6 inactivation (dotted line). Elimination of feedback from V2 is known to dramatically affect neurons in layer 6 of V1 [9]. (C) Block histograms summarizing distribution of length tuning in all 32 layer 2+3 model neurons in the central level 1 KF module with feedback (left) and without feedback (right) from level 2. End-inhibition was quantified as the percentage difference between peak response and average plateau response for lengths greater than 18 pixels: $(\text{peak} - \text{plateau})/\text{peak} \times 100$. Model neurons were classified into 10 categories according to their degree of end-inhibition, with 100% inhibition denoting a plateau response of zero to long bars. As shown, disabling the feedback connections eliminated endstopping (defined as greater than 50% inhibition) in 84% of the layer 2+3 model neurons [29].

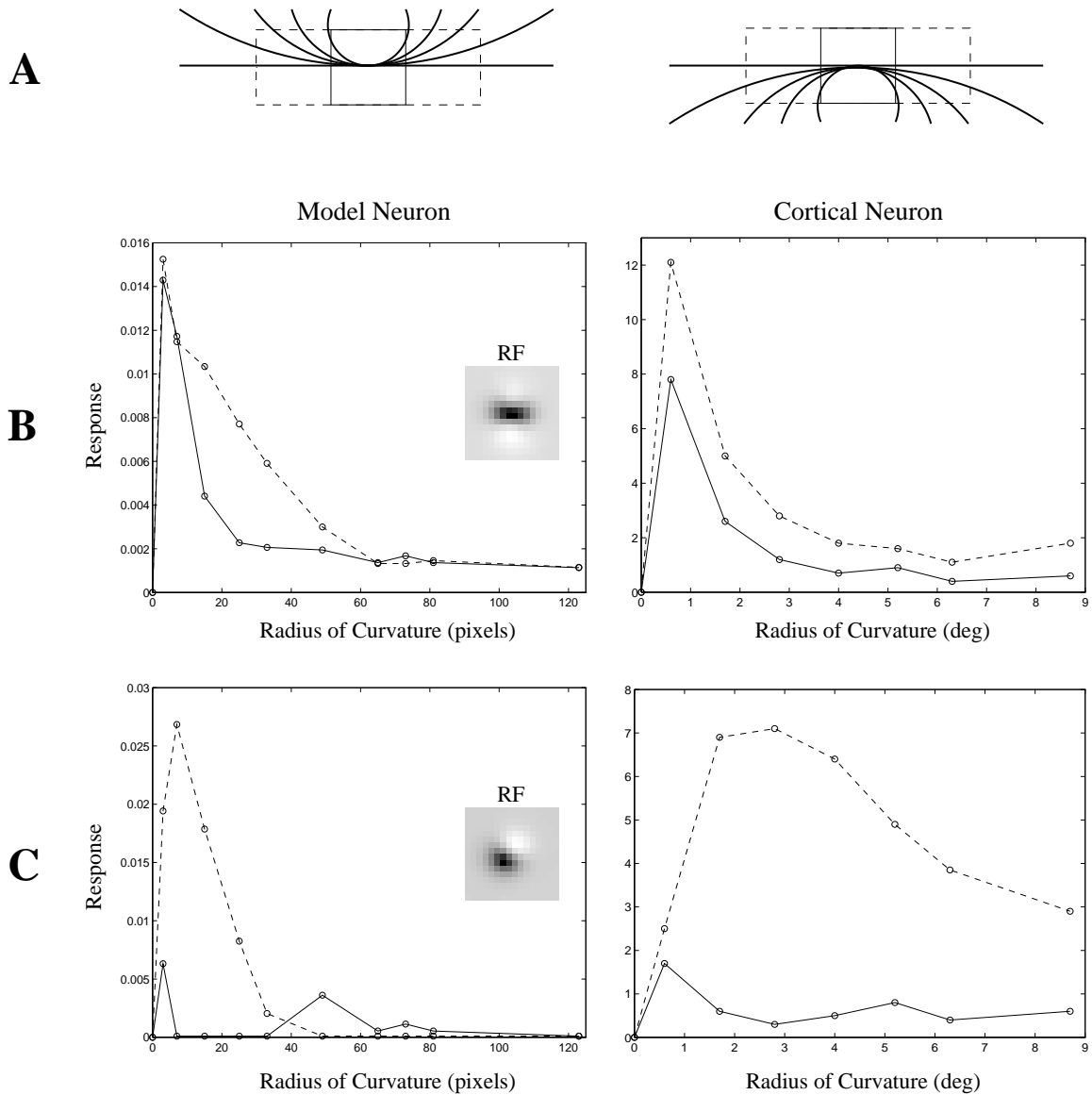


Figure 4: Curvature Selectivity in Endstopped Model Neurons. Numerous authors have ascribed a role for endstopped cortical neurons in representing curvature [3, 21, 23]. To ascertain whether curvature selectivity is subsumed by the Kalman filter model, we tested the responses of layer 2+3 endstopped model neurons to curved edges using stimuli similar to that used by [3]. **(A)** Curvature stimuli of opposite signs and increasing radii of curvature used to test model neurons for curvature selectivity. Semi-circular arcs of radii ranging from 0 to 123 pixels were input to the three-level network from Figure 1C, each vertically and horizontally centered such that the tangent to the curve at mid-arc was parallel to the long axis of the level 2 receptive field (dashed box). Five of these semi-circular arc stimuli are shown superimposed for comparison for each of the two signs of curvature (left and right respectively). **(B)** (Left) Curvature response curves for the two signs of curvature (solid and dashed lines) for the even-symmetric endstopped model neuron from Figure 3A. (Right) Curvature response curves for a neuron in cat striate cortex (V1) reported in [3]. Both show qualitatively similar *symmetrical* responses to curvature stimuli of opposite sign. **(C)** (Left) Response curves for the odd-symmetric endstopped model neuron from Figure 3A. Negative values were rectified to zero as in [3]. (Right) Response curves for a striate cortex neuron reported in [3]. Both show qualitatively similar *asymmetrical* responses to curvature stimuli of opposite sign.