

A Neural Computation for Visual Acuity in the Presence of Eye Movements

Xaq Pitkow^{1‡}, Haim Sompolinsky^{2,3}, Markus Meister^{3,4*}

1 Program in Biophysics, Harvard University, Cambridge, Massachusetts, United States of America, **2** Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem, Israel, **3** Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America, **4** Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, United States of America

Humans can distinguish visual stimuli that differ by features the size of only a few photoreceptors. This is possible despite the incessant image motion due to fixational eye movements, which can be many times larger than the features to be distinguished. To perform well, the brain must identify the retinal firing patterns induced by the stimulus while discounting similar patterns caused by spontaneous retinal activity. This is a challenge since the trajectory of the eye movements, and consequently, the stimulus position, are unknown. We derive a decision rule for using retinal spike trains to discriminate between two stimuli, given that their retinal image moves with an unknown random walk trajectory. This algorithm dynamically estimates the probability of the stimulus at different retinal locations, and uses this to modulate the influence of retinal spikes acquired later. Applied to a simple orientation-discrimination task, the algorithm performance is consistent with human acuity, whereas naive strategies that neglect eye movements perform much worse. We then show how a simple, biologically plausible neural network could implement this algorithm using a local, activity-dependent gain and lateral interactions approximately matched to the statistics of eye movements. Finally, we discuss evidence that such a network could be operating in the primary visual cortex.

Citation: Pitkow X, Sompolinsky H, Meister M (2007) A neural computation for visual acuity in the presence of eye movements. *PLoS Biol* 5(12): e331. doi:10.1371/journal.pbio.0050331

Introduction

People with normal visual acuity are able to resolve visual features that subtend a single arc minute of visual angle. For the letters “F” and “P” on a Snellen eye chart, this corresponds to a difference of just a few photoreceptors (Figure 1). As we try to resolve these tiny features, fixational eye movements jitter them across the retina over distances substantially greater than the features themselves (Figure 1). How can we have such fine acuity when our eyes are moving so much?

If the brain knew the complex eye movement trajectory, then it could realign the retinal responses before processing them further. However, central visual circuits probably do not have access to the eye movement trajectory at a sufficiently fine scale. Fixational eye movements arise from imperfect compensation for head and body movements [1,2] and motor noise [3], so it is unlikely that the visual system has a reliable estimate of the resultant image motion. Although there are both efference copies of eye movement signals and proprioceptive feedback, they have a limited accuracy of several degrees [4,5], which is inadequate for tracking the much smaller movements during fixation. Thus, any estimate the brain makes about fine fixational eye movements is probably driven by visual input alone [6,7].

Unfortunately, visual processing in the retina introduces noise, leaving the brain with uncertainty both about the stimulus shape itself and about the precise trajectory the stimulus traces on the retina. The retina’s output neurons—the retinal ganglion cells—are not perfectly reliable in their response to stimulation, and even without stimulation, they fire action potentials at a substantial rate. For brief, small stimuli on a featureless background, the total stimulated

retinal response may consist of just a few tens of spikes. The brain must distinguish these spikes from the many hundreds of spontaneous spikes that reflect only noise. The usual remedy would be to accumulate many spikes over time until the signal emerges from the noise; but this is difficult because the fixational eye movements scatter the desired responses across space.

Thus we recognize a challenge for visual acuity in the presence of eye movements: To identify the stimulus, the brain needs to know the precise stimulus trajectory; yet to track the stimulus trajectory, the brain needs to identify which neural spikes are stimulated and which are only noise.

Presented with this challenge, what strategy could the brain use to achieve the visual acuity that humans exhibit? We will show that naive decodings of retinal spike trains that neglect the eye movements perform poorly at discriminating fine visual features. We derive a significantly better strategy that exploits the fact that eye movements are continuous to estimate the stimulus position on the retina and give greater weight to retinal spikes originating near this position. Surprisingly, we found that this strategy is attainable by a

Academic Editor: David Burr, Istituto di Neurofisiologia, Italy

Received March 27, 2007; **Accepted** November 9, 2007; **Published** December 27, 2007

Copyright: © 2007 Pitkow et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: V1, primary visual cortex

* To whom correspondence should be addressed. E-mail: meister@fas.harvard.edu

‡ Current address: Center for Theoretical Neuroscience, Columbia University, New York, New York, United States of America

Author Summary

Like a camera, the eye projects an image of the world onto our retina. But unlike a camera, the eye continues to execute small, random movements, even when we fix our gaze. Consequently, the projected image jitters over the retina. In a camera, such jitter leads to a blurred image on the film. Interestingly, our visual acuity is many times sharper than expected from the motion blur. Apparently, the brain uses an active process to track the image through its jittering motion across the retina. Here, we propose an algorithm for how this can be accomplished. The algorithm uses realistic spike responses of optic nerve fibers to reconstruct the visual image, and requires no knowledge of the eye movement trajectory. Its performance can account for human visual acuity. Furthermore, we show that this algorithm could be implemented biologically by the neural circuits of primary visual cortex.

simple neural network whose properties are consistent with functional and anatomical features of primary visual cortex.

Results

Psychophysics

For concreteness, we choose a simple task to analyze: An observer is asked to discriminate between two tiny oriented bars that span 1 or 2 arcmin of visual angle. In the retina's fovea, this stimulus affects just a few cone photoreceptors, each collecting light from a region about 0.5 arcmin in diameter. Each cone drives approximately one On-type and one Off-type ganglion cell, and conversely, each ganglion cell receives its input from just one cone [8]. This means that at any given instant, the brain receives information about the stimulus from spiking in a small cluster of retinal ganglion cells, but the identity of those cells changes continually as the stimulus jitters across the retina. We tested human subjects on this discrimination task and found that despite these challenges, many human subjects can actually perform well above chance (Figure 2, see also [9,10]).

It is plausible that the finest human acuity might be limited primarily by the information available in the retina rather than by later constraints or losses. For example, our ability to detect dim lights in absolute darkness is ultimately limited by photon shot noise at the rod photoreceptor. In bright light—the condition considered here—noise introduced by retinal processing greatly exceeds photon shot noise [11–13]. Correspondingly, human thresholds on fine acuity tasks are worse by a factor of ten than expected from ideal processing of photon counts [9,10]. Instead, human performance on simple visual tasks is more compatible with the limitations from noisy retinal ganglion cell spikes [14,15]. If acuity is in fact limited by the retinal spike trains, then the brain must make efficient use of these spikes to extract the relevant information.

Markov Decoder Model

We now present a strategy for accumulating information about position and orientation of the small stimulus bar on the retina. This strategy decodes the observed spike trains from retinal ganglion cells using prior knowledge about the statistics of those spikes and the statistics of eye movements. The output of the decoder is a moment-to-moment estimate of the bar's orientation.

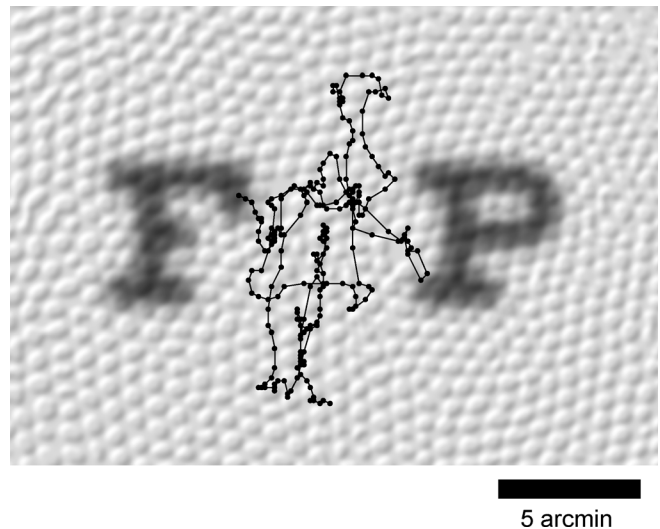


Figure 1. The Neighboring Letters “F” and “P” on the 20/20 Line of the Snellen Eye Chart, Blurred by a Gaussian of Diameter 0.5 arcmin and Projected onto an Image of the Foveal Cone Mosaic (Photoreceptor Image Modified from [92])

The 1-arcmin features that distinguish the letters extend over only a few photoreceptors. Also shown is a sample fixational eye movement trajectory for a standing subject (courtesy of [25]), sampled every 2 ms for a duration of 500 ms and then smoothed with a 4-ms boxcar filter. doi:10.1371/journal.pbio.0050331.g001

The decoder assumes a model of retinal ganglion cell spike generation, shown in Figure 3A, which includes random eye movements, optical blur, spatial receptive fields, temporal filtering, rectification, and probabilistic spiking. Each stimulus is a small, dark, oriented rectangle that jitters across the retina. The eye's optics introduce a spatial blur, implemented by a Gaussian filter with a 0.5 arcmin diameter. We assume this image is sensed by photoreceptors arranged on a square lattice, each activating one Off-type ganglion cell. We neglect the On-type cells because they will generate only a weak response to the small, dark stimulus (see Discussion). For the same reason, we neglect the broad, but shallow, surrounds of Off-cells, which are usually approximately 50 times weaker than the receptive field center [16]. Furthermore, we first assume for simplicity that ganglion cells report on the instantaneous light intensity in their receptive field center; later, we will consider implications of including a temporal filter like that in Figure 3D. Under these assumptions, when a stimulus with orientation S is at position \mathbf{x} , a model retinal ganglion cell at position \mathbf{y} fires action potentials with Poisson statistics at the instantaneous time-dependent rate $r_S(\mathbf{y} - \mathbf{x})$ depicted in Figure 3B, ranging from a peak value r_{\max} at positions near the stimulus to the background firing rate r_0 at large distances. In bright conditions, retinal ganglion cells respond to a contrast of 100% (black on white) with a spike rate of $r_{\max} \sim 100$ Hz [17]. Far from the stimulus, we assume neurons fire spontaneously with rates on the order of $r_0 \sim 10$ Hz [18,19].

In weighting the retinal responses properly, the decoder takes into account the statistics of the trajectories that are traced by the fixed stimulus on the moving retina. Fixational eye movements are classified into three types of motion: microsaccades, drift, and tremor [20]. Microsaccades are not thought to play a role in fine visual tasks [21–23], though they

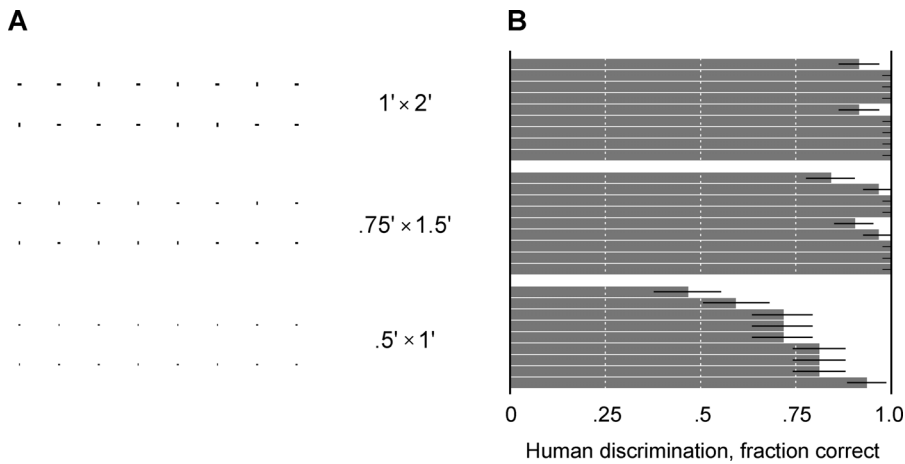


Figure 2. The Discrimination Task

(A) Tiny horizontal and vertical stimuli, sized to subtend 0.5×1 , 0.75×1.5 , and 1×2 arcmin² when viewed at a distance of 88 cm.

(B) Performance of nine human participants on this task, measured by the fraction of correct guesses out of 32 trials. Error bars represent the 68% confidence interval.

doi:10.1371/journal.pbio.0050331.g002

may contribute to peripheral vision [24]. Tremor has very low amplitude, less than a photoreceptor diameter. We therefore concentrate on the drift component, which has the properties of a random walk [3], with modest deviations on short and long timescales [25]. For simplicity, we assume that the fixational eye movements are described by a spatially discrete random walk across the photoreceptor lattice with an effective diffusion constant of $D \sim 100$ arcmin²/s (see Materials and Methods).

For a random walk trajectory, the probability of the current position depends only on its most recent previous position. This attribute, in combination with the assumption that retinal responses are memoryless, allows us to write a differential equation for the probability distribution $P(S, \mathbf{x}, t)$ of the stimulus orientation S and current location \mathbf{x} , given all the spikes observed before time t :

$$\frac{\partial}{\partial t} P(S, \mathbf{x}, t) = \sum_{\mathbf{y}} \lambda_{\mathbf{y}}(t) f_S(\mathbf{y} - \mathbf{x}) P(S, \mathbf{x}, t) - r_S^{tot}(\mathbf{x}) P(S, \mathbf{x}, t) + D \nabla^2 P(S, \mathbf{x}, t) \quad (1)$$

(see Protocol S1 for a derivation). In this equation, $\lambda_{\mathbf{y}}(t) = \sum_{t_y} \delta(t - t_y)$ stands for the observed spike train of the retinal neuron \mathbf{y} at time t ; $f_S(\mathbf{y} - \mathbf{x}) = \ln(r_S(\mathbf{y} - \mathbf{x})/r_0)$ reflects the expected firing-rate profile generated by the stimulus; $r_S^{tot}(\mathbf{x}) = \sum_{\mathbf{y}} r_S(\mathbf{y} - \mathbf{x})$ denotes the total expected firing rate of the retinal ganglion cell array; and ∇^2 represents a discrete version of a second-order spatial derivative operator. On a square lattice, $\nabla^2 P(S, \mathbf{x}, t) = \frac{1}{a^2} (\sum_{\Delta \mathbf{x}} P(S, \mathbf{x} + \Delta \mathbf{x}, t) - 4P(S, \mathbf{x}, t))$, where $\mathbf{x} + \Delta \mathbf{x}$ ranges over the four nearest neighbors of \mathbf{x} on the lattice (Figure 3C), and a is the distance between lattice points.

Equation 1, also known as a Fokker-Planck equation, describes a reaction-diffusion system [26]. There are three sources of changes in the stimulus posterior probabilities $P(S, \mathbf{x}, t)$. The first term,

$$\sum_{\mathbf{y}} \lambda_{\mathbf{y}}(t) f_S(\mathbf{y} - \mathbf{x}) P(S, \mathbf{x}, t), \quad (2)$$

implies that each spike of a retinal neuron \mathbf{y} results in a

multiplicative update of the stimulus posterior probabilities $P(S, \mathbf{x}, t)$ by a factor $r_S(\mathbf{y} - \mathbf{x})/r_0$ (as shown in Materials and Methods), thus increasing the likelihoods of stimulus positions \mathbf{x} near the firing retinal neuron, where $r_S(\mathbf{y} - \mathbf{x})/r_0$ is large. The second term,

$$-r_S^{tot}(\mathbf{x}) P(S, \mathbf{x}, t), \quad (3)$$

represents the “negative” evidence accumulating during quiescent periods. In between retinal spikes, $P(S, \mathbf{x}, t)$ decays exponentially with a decay rate that equals the total expected firing rate of the retinal array with the stimulus S at position \mathbf{x} . In the present case, in which the total activation of the retina is the same regardless of the orientation and position of the stimulus, we ignore this term since it does not affect the relative values of the posterior distribution for different orientations S or positions \mathbf{x} . These first two terms represent the local “reaction” terms. The last term,

$$D \nabla^2 P(S, \mathbf{x}, t), \quad (4)$$

is the “diffusion” term; it describes the lateral spread of the posterior probability across the retina during the time between retinal spikes. This spread accounts for the expected stimulus movements due to the fixational eye movements. The rate of spread is given by D , the diffusion constant of the fixational eye movements. The initial condition for solving Equation 1 is specified by $P(S, \mathbf{x}, 0)$, which is the initial probability distribution of the stimulus orientation and position prior to observing any spikes. We will assume that it is uniform over the entire range of positions and orientations. Finally, we note that Equation 1 technically yields the posterior probability only up to an overall normalization factor (see Protocol S1 for a strictly normalized version). This is unimportant for discrimination, since only the relative values of P for different orientations matter. However, in numerical work, one must supplement Equation 1 by a divisive normalization, periodically dividing all components of P by the sum $\sum_{S, \mathbf{x}} P(S, \mathbf{x}, t)$ over space and orientation (see Materials and Methods).

This decoder of retinal spike trains can be applied to a

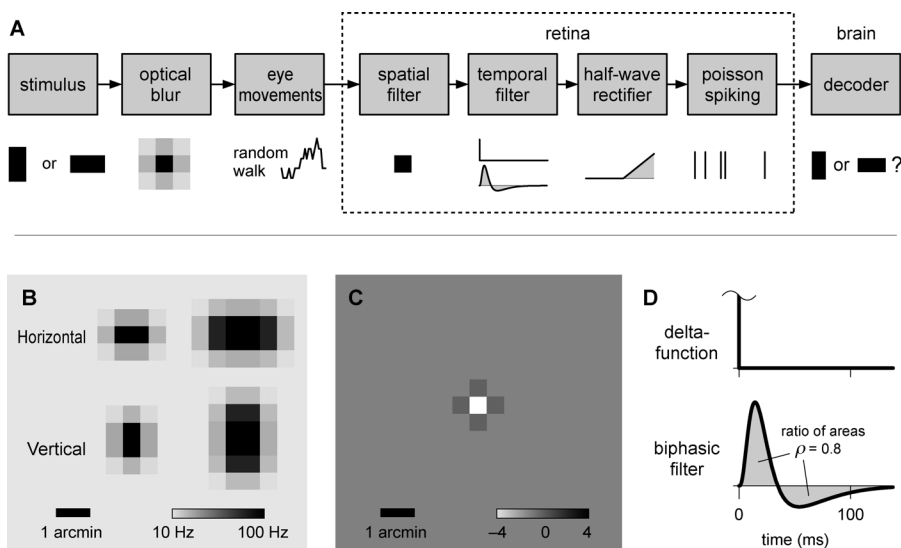


Figure 3. Models of Spike Generation and Decoding

(A) A block diagram of the features in the model visual system; see text for details.

(B) Firing-rate profiles $r_s(\mathbf{y})$ induced by horizontal and vertical stimuli on the model foveal lattice. Left: 0.5×1 arcmin². Right: 1×2 arcmin².

(C) A graphical representation of the discrete second-derivative operator used to calculate diffusion rates.

(D) The temporal filters that model retinal ganglion cells use to convert the time-varying light intensity into the instantaneous firing rate.

doi:10.1371/journal.pbio.0050331.g003

variety of tasks. For instance, in a localization task with a stimulus of known orientation, S , the estimate of the stimulus position \mathbf{x} is given by $\mathbf{x}^{\text{estimate}} = \arg \max P(S, \mathbf{x}, t)$. In a discrimination task in which only the orientation needs to be determined, a sum over the irrelevant position variable yields $S^{\text{estimate}} = \arg \max_S \sum_{\mathbf{x}} P(S, \mathbf{x}, t)$.

The first-order differential equation (Equation 1) implies that the posterior probability can be updated in a way that depends only on the current posterior probability and the current evidence from spikes. This is possible because the assumed process of generating spikes depends only on the current stimulus location. This is an example of what is known as a Markov process, more specifically, a hidden Markov process because the location variable is not observed directly. We will call this decoder of the spike trains the “Markov decoder.” It will yield optimal decisions if the Markov assumptions accurately describe the spike generation process.

Visualizing the Markov Decoder Algorithm

We illustrate the performance of the decoder in Figure 4A–4E, using spike trains from a one-dimensional model retina. In the first task (Figure 4A–4C), the stimulus shape is known, so the only uncertainty is its location. The stimulus follows a random walk trajectory, generating the instantaneous firing-rate pattern (Figure 4A), and eliciting extra spikes for neurons along its path while other neurons produce spontaneous spikes at a lower rate (Figure 4B). The Markov decoder collects all the retinal spikes and solves Equation 1 to estimate the posterior probability distribution over positions (see Materials and Methods for numerical details). The result is displayed in Figure 4C.

In the particular trial depicted, the task of localizing the stimulus appears quite difficult, even with only one spatial dimension: In any given time slice, the evidence provided by retinal spikes is rather weak. Nonetheless, the accumulated

evidence over time provides a good estimate of the stimulus trajectory. As evidence from the spiking neurons accumulates, the decoder locks onto and tracks the true stimulus location.

In a second task, the decoder must discriminate between two possible stimulus shapes moving on a one-dimensional retina (Figure 4D and 4E). Because one dimension does not allow for horizontal and vertical bars, we take the shape variable S to refer to two stimuli related by reflection (Figure 4D, inset). Again, these probabilities evolve according to the reaction-diffusion dynamics of Equation 1, where incoming spikes lead the probability distributions to track the stimulus, but now there is a competition for probability between two stimulus shapes. The Markov decoder may make errors in position, stimulus identity, or both, depending on the particular spike trains it observed, but on average, it discriminates between the two stimulus shapes with an accuracy well above chance.

Non-Markovian Spike Generation with Temporal Filtering

So far, we have assumed that the retinal ganglion cells report on the instantaneous light intensity, but this is not realistic. Primate photoreceptors react slowly, with integration times on the order of 25 ms [27], yet the eye movements’ diffusion constant of 100 arcmin²/s implies that the stimulus typically moves one photoreceptor diameter in under a millisecond. Therefore, the firing of retinal ganglion cells cannot track the light intensity as it fluctuates on this rapid timescale. More realistically, the ganglion cells respond to the light intensity in their receptive field averaged by a biphasic temporal filter like that shown in Figure 3D [28].

This temporal filtering has an important implication: Since the spiking probability depends on an extended history of stimulus positions, the spikes cannot be interpreted optimally by the Markov decoder. One can generalize Equation 1 to derive the optimal decoder in this situation. The posterior

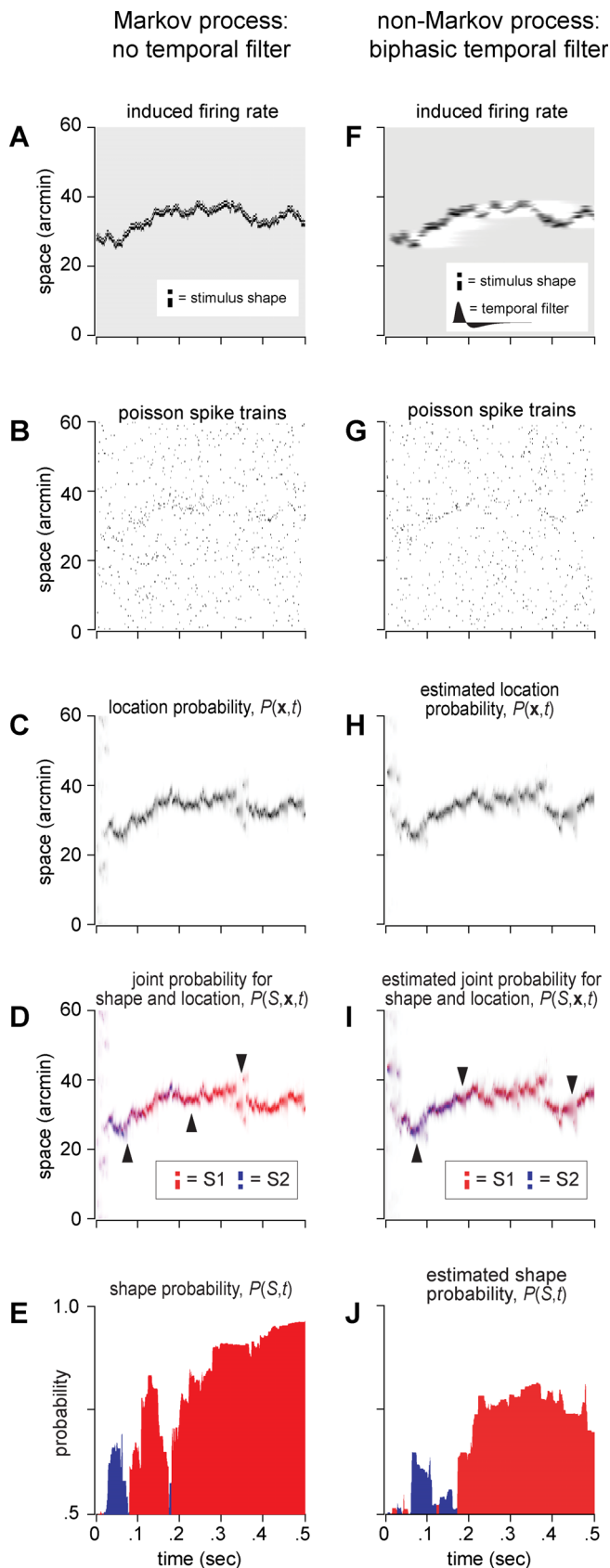


Figure 4. Simulations of the Markov Decoder (Equation 1) for a Small Stimulus Moving on a One-Dimensional Model Retina

(A–E) Spike generation by a Markov process.

(F–J) Spike generation by a non-Markov process that includes the biphasic temporal filter from Figure 3D.

(A and F) Firing rate induced by a stimulus moving on the retina with a random walk diffusion constant of $100 \text{ arcmin}^2/\text{s}$. The stimulus shape activates three neurons in the pattern shown in the inset. The background rate is 10 Hz, and the peak stimulated rate is 100 Hz.

(B and G) Poisson retinal spike trains drawn from this instantaneous firing rate. Each row corresponds to a neuron, spaced every 0.5 arcmin .

(C and H) Evolution of the location probability $P(\mathbf{x}, t)$ for a known stimulus shape S (inset in [A]), but an unknown location \mathbf{x} , derived from the spike trains shown in the previous panel.

(D and I) Decoder behavior when the stimulus can instead take one of two possible shapes, but the true shape is unknown. The two stimuli each activate three retinal neurons, in mirror-image patterns (inset). The spike trains now induce two spatial distributions of the posterior probability $P(S, \mathbf{x}, t)$, plotted in shades of red and blue.

(E and J) Shape probability $P(S, t) = \sum_{\mathbf{x}} P(S, \mathbf{x}, t)$, colored red for the correct stimulus identity and blue for the incorrect one. In these trials, we see that once the decoder coalesces around the stimulus location, it first attributes a greater probability to the wrong stimulus (leftmost arrow in [D] and [I]) before accumulating enough evidence for the correct stimulus (middle arrow). The decoder can lose track of the stimulus briefly (e.g., at rightmost arrow) but continues to favor the correct stimulus until the end of the trial. Note that (E) reflects the true posterior probabilities, whereas in (J), the Markov decoder can only estimate them because the spike generation process includes temporal filtering that the decoder neglects.

doi:10.1371/journal.pbio.0050331.g004

probability distribution now extends over all possible random walk trajectories within the temporal range of the filter. There are approximately 10^8 such trajectories leading up to each stimulus location, and propagating their probability distribution is numerically unwieldy. It also seems improbable that the brain takes such an approach. These arguments apply strictly to the optimal decoder, but there may exist useful and efficient nonoptimal decoders. In fact, we found that the simple Markov decoder still performs well at the discrimination task, despite the mismatch between the encoding process and the decoder's assumptions.

To explore this, we generated retinal ganglion cell spikes (Figure 4F and 4G) with a model that includes a biphasic temporal filter (Figure 3D). The filtering adds a motion smear to the stimulus, which renders the output spike trains more ambiguous. Despite its ignorance of the temporal filtering, the decoder can still track the stimulus location, with a small delay due to the filter (Figure 4H). Furthermore, the decoder successfully accumulates information about the stimulus shape (Figure 4I and 4J).

Performance of the Markov Decoder

We now evaluate the Markov decoder's performance on the original visual task: to discriminate whether a small jittering bar is oriented horizontally or vertically. Here, we modeled the retina and the decoder using two spatial dimensions and simulated many trials of the discrimination task. For every trial, we selected a random stimulus orientation and trajectory, filtered the instantaneous light intensity with a biphasic temporal filter, rectified the result to calculate the expected firing rates for all retinal neurons over time, and generated Poisson spike trains with these firing rates (Figure 3). We then applied the decoder algorithm to these spike trains by numerically solving Equation 1 and selecting the orientation estimated to be more probable. Performance was quantified as the fraction of trials in which the decoder guessed correctly.

The results of these simulations show that the Markov

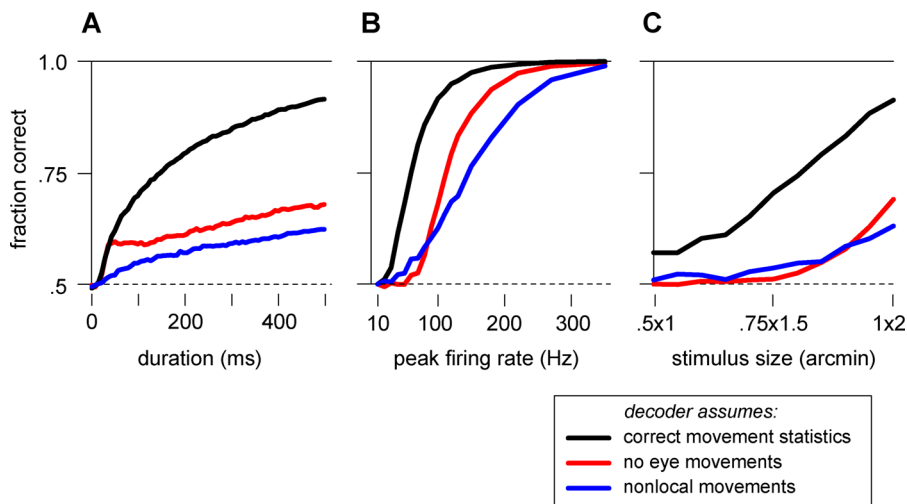


Figure 5. Model Performance on the Horizontal versus Vertical Discrimination Task Shown in Figure 2

Performance is measured by simulating retinal responses, calculating decisions based on those responses, and computing the fraction of correct decisions (see Materials and Methods). When fixational eye movements jitter the stimulus, the Markov decoder is able to perform well on the task by accounting for the eye movement statistics (black curves). Two naive decoders are also applied to this task, one that assumes the stimulus is fixed (red) and one that assumes maximum uncertainty about those movements (blue). Performance is shown as a function of stimulus duration (A), peak stimulated firing rate (B), and stimulus size (C). Where not otherwise specified, the parameters for these simulations are background firing rate of 10 Hz, a peak stimulated rate of 100 Hz, a stimulus of 1×2 arcmin², a duration of 500 ms, and a diffusion constant of 100 arcmin²/s. doi:10.1371/journal.pbio.0050331.g005

decoder's performance is generally compatible with human performance. The decoder is able to reliably discriminate horizontal from vertical within a few hundred milliseconds (Figure 5A) using spikes generated at biologically realistic rates around 100 Hz (Figure 5B). Like humans, the Markov decoder finds discrimination very challenging with the smallest stimuli, and fairly routine for the largest (compare Figures 2B and 5C).

Importance of Accounting for Fixational Eye Movements

The Markov decoder can be used to evaluate the importance of accounting for fixational eye movements in estimating the stimulus shape or orientation. Specifically, we ask the question: how much better does the Markov decoder perform compared to strategies that ignore the eye movement statistics?

Two naive strategies can be proposed: The first assumes that there are no eye movements. This amounts to using a Markov decoder, but setting its presumed diffusion constant to zero. Another strategy recognizes that the eye moves approximately every 0.6 ms (the average time between random walk steps on the square lattice), but is otherwise ignorant of the eye movement statistics; it conservatively assumes that jumps to all stimulus positions are equally likely.

Naturally, the decoder that uses the correct diffusion statistics works best, but simulations reveal that it outperforms the two naive decoders by a large margin (Figure 5). For very brief stimuli of the same duration as the transient retinal response (~ 30 ms), the decoder that assumes a fixed stimulus and the decoder that knows the correct movement statistics perform equally well, because temporal filtering does not allow the responses to track the stimulus movements. Yet, under typical viewing conditions, such a duration is too brief for human subjects to discriminate the stimulus shapes. As the decoder integrates information beyond the temporal filter's persistence time, the movements become relevant and

the naive algorithm essentially blurs the stimulus even more. The decoder giving equal odds to all locations at all times relies only on the rare coincidences when multiple stimulated neurons spike in tight synchrony. Eventually, this naive decoder can manage to discriminate the stimuli, but it requires a much longer time or many more spikes than the Markov decoder.

Robustness

How robust is the algorithm to imperfections in implementation? The key parameter that incorporates the statistics of the eye movements is the assumed diffusion constant. As shown above, if the decoder assumes that the eye movements are much faster or much slower than they really are, then the performance degrades substantially. However, between these two extremes, there is a broad range of assumed diffusion constants that causes only a few percent of extra mistakes (Figure 6A). In fact, the decoder benefits slightly from assuming a lower diffusion constant, probably due to the apparent stimulus persistence caused by temporal blurring. This demonstrates that it is essential to account for eye movements, but the algorithm proposed here is robust to misestimates of the movement statistics.

Every time the decoder receives a retinal spike, the estimated stimulus probability rises locally by a factor proportional to the expected stimulated firing rate divided by the background rate (Materials and Methods, Equation 10), which reflects the confidence in the new information brought by a retinal spike. Changing this factor in the Markov decoder would be expected to alter its performance. However, we found that performance is remarkably insensitive to this variable over a wide range of values (Figure 6B).

Finally, we may ask whether the decoder performance is sensitive to the assumed stimulus shapes. Each retinal spike increases the estimated stimulus probability at all those locations where a stimulus could potentially have caused that

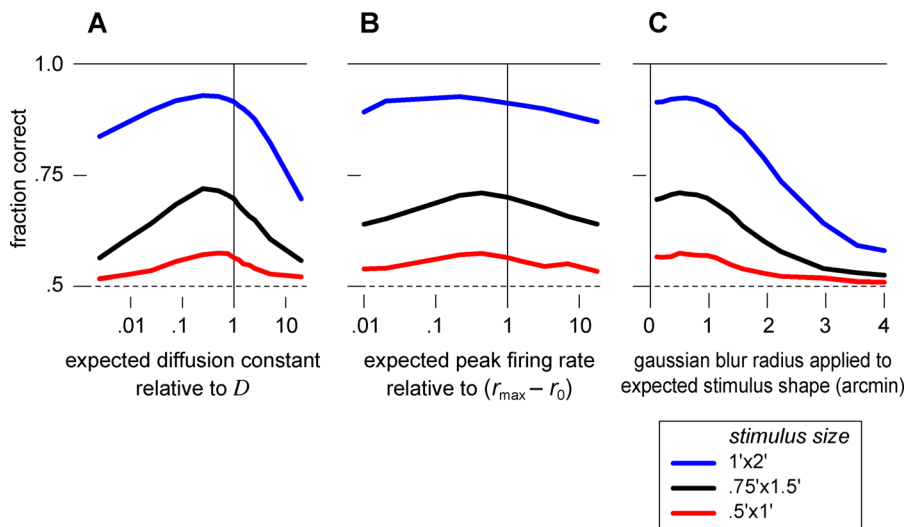


Figure 6. Markov Decoder Robustness to Mismatched Parameters

(A) Discrimination performance when the decoder's estimate for the trajectory statistics is wrong: The stimulus is known to perform a random walk on the retina, but the diffusion constant is misestimated. The performance is optimal for estimated values close to the actual diffusion constant and declines gently on either side.

(B) Performance as a function of the expected stimulated firing rate, parameterized as $(r_{max}^{est} - r_0)/(r_{max} - r_0)$.

(C) Performance as a function of the expected stimulus size, obtained by convolving the true stimulus shape with a spatial Gaussian of the specified radius. In each of these plots, parameters are the same as in Figure 5.

doi:10.1371/journal.pbio.0050331.g006

spike. If the expected stimuli differ from the true stimuli, then this probability increases over the wrong set of locations, leading to suboptimal performance. To explore this, we set the decoder's expected stimulus shape to be larger than the true shape by various amounts (Figure 6C). Enlargement up to about 1 arcmin produced no noticeable change in the decoder's performance, but larger discrepancies of about 2 arcmin led to significant decline. This behavior can be understood as follows: a misestimate of the stimulus size effectively leads to excessive smearing of the positional information. This must be compared to the diffusional smearing that occurs as the stimulus moves in the typical time between informative spikes, which amounts to approximately 1 arcmin. Thus the Markov decoder is hardly affected by misestimates in stimulus shape smaller than this amount.

In summary, the Markov decoder is robust to various parameters that encompass its a priori assumptions about the stimulus. If the decoder allows activity to diffuse at an approximately correct rate, and expects shapes not dramatically larger than the true stimuli, then it can achieve good discrimination performance.

Network Implementation

Despite the apparent complexity of the differential equation governing the Markov decoder, its dynamics map directly onto a simple neural network with a structure consistent with many known properties of visual cortex. For clarity, we will first introduce a network that estimates the location probabilities for a given stimulus shape, and then show the extension required for shape discrimination.

Figure 7 depicts a network that implements the Markov decoder algorithm for estimating the location of a stimulus with a known orientation S . The network has three types of neurons: the retinal neurons, a hidden layer of decoder

neurons, and an inhibitory neuron. Each neuron in the hidden layer is associated with a spatial location, \mathbf{x} , and its activity at time t represents the estimated posterior probability (up to a normalization factor) that the stimulus is present at that location, $P(S, \mathbf{x}, t)$. The feedforward input to each hidden layer neuron \mathbf{x} consists of spikes from retinal locations \mathbf{y} , weighted by a spatial receptive field $f_S(\mathbf{y} - \mathbf{x}) = \ln(r_S(\mathbf{y} - \mathbf{x})/r_0)$, which ranges from zero far from the stimulus to a peak of $\ln(r_{max}/r_0)$. The weighted retinal input is then multiplied by a variable gain proportional to the activity of the postsynaptic neuron, $P(S, \mathbf{x}, t)$. This gated retinal input implements the contribution of Equation 2 to the update of the estimated posterior probability. The neurons in the network interact through lateral connections mimicking the diffusion operator (Equation 4 and Figure 3C). Recall that the diffusion operator takes the summed probability of the nearest-neighbors of a given location, $\sum_{\Delta\mathbf{x}} P(S, \mathbf{x} + \Delta\mathbf{x}, t)$, and subtracts $4P(S, \mathbf{x}, t)$ from this in order to conserve probability. In the network, conservation of activity is not required, so the subtraction can be omitted: when the change in P is simply proportional to P , the solution is an exponential decay that scales P uniformly at all locations, leaving the relative values of the activity unaltered. Thus, lateral excitatory connections are sufficient to implement the diffusion term in the network. For the same reason, the network does not need any representation of the local decay term, Equation 3, which also scales all activities equally. Finally, the network includes a global divisive inhibition to maintain network activity at a stable level despite the various excitatory interactions.

To extend this framework to the discrimination task, we need two copies of the network that differ by their orientation tuning (Figure 8). In the "horizontal" network, representing $P(H, \mathbf{x}, t)$, the neurons are tuned to horizontal stimuli, hence their receptive fields are determined by $r_H(\mathbf{y} - \mathbf{x})$ (Figure 3B);

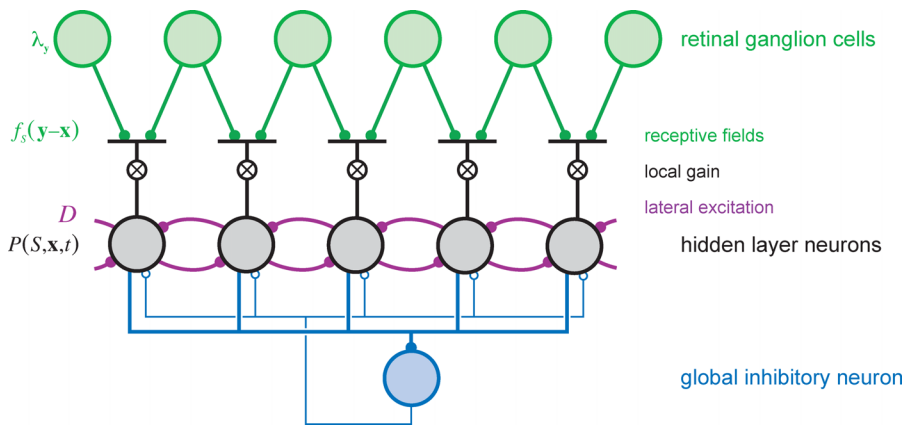


Figure 7. Schematic for a Network Implementation of the Markov Decoder (Equation 1)

Spikes from retinal neurons (green, top layer) are collected by neurons in a hidden layer (black, middle layer) with linear receptive fields $f_S(\mathbf{y} - \mathbf{x})$ and a local gain that is set by activity in the recipient neuron. Global divisive inhibition is driven by the total activity of all neurons in the hidden layer through a pooling neuron (blue, bottom neuron).
doi:10.1371/journal.pbio.0050331.g007

correspondingly, in the “vertical” network, representing $P(V, \mathbf{x}, t)$, the receptive fields are related to $r_V(\mathbf{y} - \mathbf{x})$. For the discrimination task, the retinal position is irrelevant; comparing the pooled activity from each subnetwork is sufficient to discriminate between the stimulus orientations. Note that the lateral excitatory connections in this network architecture are orientation specific because fixational eye movements translate the stimulus, but do not appreciably rotate it: orientation, but not position, is preserved. On the other hand, the stabilizing divisive normalization must be global across orientations to ensure a meaningful comparison between the two orientations.

Discussion

Fixational eye movements pose a major challenge for vision since they scatter weak signals about fine stimulus features across the retina. We addressed this challenge mathematically by deriving an algorithm that guesses the orientation of a stimulus, given spiking responses from a model retina and prior knowledge about its function. It accomplishes this by collecting and sorting the scattered feature information in a systematic way, weighting retinal spikes according to an estimated probability that those spikes reflect stimulus features and not noise.

Biological Implementation

As described above, the decoder algorithm has a direct mapping onto an abstract neural network, and we will argue that primary visual cortex (V1) has many properties well suited to instantiate this network with real neurons. Specifically, we take the hidden layer neurons in Figure 7 to be cortical cells that receive inputs from the retina via the thalamus.

For good performance, these neurons should integrate retinal spikes using linear, oriented receptive fields of the same shape and size as the visual stimuli (Figure 8). We showed that the decoder’s performance was robust to mismatches between the true stimuli and the expected stimuli (Figure 5B and 5C), so these receptive fields need be only approximately tuned to the stimulus size and strength. Linear oriented receptive fields are a well-established

characteristic of cortical simple cells [29]. For stimuli subtending only a few human photoreceptors, we require a receptive field of just 1 or 2 arcmin in size. Receptive fields for cortical neurons dedicated to foveal vision are notoriously difficult to measure, notably due to technical problems associated with fixational eye movements. In macaques, receptive fields have been reported as small as 3 arcmin, slightly larger than the macaque’s cone resolution of about 1.7 arcmin [30,31]. Therefore, cortical neurons are likely to exist with receptive fields of the appropriate size. Although equivalent measurements are unavailable for human cortex, our finest acuity may well be mediated by cortical neurons driven by an oriented set of just a few cones.

To account for fixational eye movements, the neural network must be organized retinotopically so that local stimulus movements correspond to local interactions in cortex. This is, of course, a known property of V1 [32,33]. Because fixational eye movements are largely independent in each eye [34], the fine retinal positioning of the stimulus is also independent for the two eyes: Proper accounting for stimulus movement, therefore, requires that lateral excitation should not cross eyes. Ocular dominance columns [35] are thus seen as a necessary feature if cortex is to accommodate fixational eye movements. Eye movements are best handled before the signals from the two eyes are mixed, favoring a locus in the lateral geniculate nucleus (LGN) or in V1 for the proposed network.

Eye movements are expected to simply translate visual features, but not rotate them, and these expectations should be built into circuitry. Activity in the model decoder network diffuses across space through lateral excitatory connections between nearby neurons, but only those with similar orientation preferences. In the early visual system, the required iso-orientation facilitation has been observed psychophysically [36–38], anatomically [39–42], and physiologically [43–45]. Lateral diffusion of activity has also been directly imaged in visual cortex [46].

As the eye drifts, the retina moves rigidly in world coordinates. But since the size of cortical receptive fields increases with distance from the fovea [30,47,48], fixational eye movements do not move stimuli across many receptive

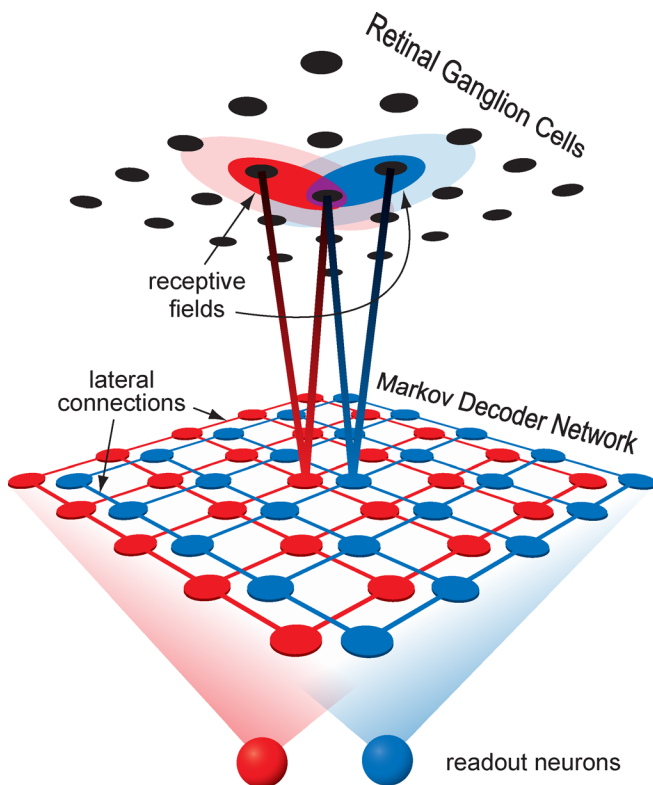


Figure 8. Two Independent but Competing Subnetworks, Each Structured as in Figure 7, Receive Input from the Same Retinal Ganglion Cells, but Use Different Receptive Fields

The total activity in each subnetwork is pooled by two readout neurons. The more active readout neuron indicates the network's estimate of the stimulus orientation.

doi:10.1371/journal.pbio.0050331.g008

fields in the periphery. Accordingly, there is no need to compensate for fixational eye movements in the periphery. We expect, therefore, to see some aspects of the cortical network that are specialized for foveal vision. Consistent with this, more of striate cortex is dedicated to responses from the fovea than can be explained by the density of retinal ganglion cells [49–51], and lateral suppression and facilitation differ between central and peripheral vision [37].

The Markov decoder requires that the lateral facilitatory interactions induce localized changes in the gain for new input spikes. Such multiplicative gain modulations have indeed been observed in the visual cortex [52,53]. A number of neural mechanisms have been invoked to create neural multipliers [54–60]. One potential mechanism involves the postsynaptic NMDA (n-methyl-d-aspartic acid) receptor, a glutamate-gated ion channel with a voltage sensitivity that causes it to open only when the postsynaptic potential is sufficiently large. In the visual cortex, NMDA activation has been shown to produce a multiplicative effect on input gain [61]. Synapses between cortical layers and within layers have different NMDA and AMPA (alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor distributions, so that lateral inputs may be simply additive, whereas feedforward input may experience a variable gain [62], as required by the Markov decoder architecture.

With an accelerating nonlinearity and excitatory interactions, this network has a positive feedback loop that would

cause the activity to quickly diverge. Normalization will maintain stability, but the normalization must be global and orientation independent so that neural activities can be compared on the same scale. Previously described wide-field divisive normalization [63–66] can serve this purpose, although other global homeostatic mechanisms would function as well.

In our forced-choice task, the accumulated evidence for the horizontal and vertical stimuli must be compared. This can be accomplished downstream by a final winner-take-all computation in which the total activity in each subnetwork is pooled and then compared [67]. This type of computation must take place somewhere in the brain any time a decision must be reached, and various biological implementations have been proposed for this operation [68,69].

Whereas the input to the network consists of discrete spikes, the network units themselves represent the stimulus probability, which is a continuous variable. This variable might be most simply encoded by the collective firing rate of a cluster of neurons [70], especially given that the number of cells representing the visual field expands dramatically from the retina to the visual cortex [71]. Alternatively, the computation might well proceed with discrete spikes: model networks of spiking neurons tend to produce similar behavior as rate models with continuous variables, so long as the spikes are not too strongly correlated [72].

In summary, all the key elements of a Markov decoder for short line segments are present in the neural circuitry of primary visual cortex. One essential feature, namely monocular processing, is no longer available beyond V1. We therefore propose that V1 functions as a dynamic network to accumulate information on fine stimulus features in the face of fixational eye movements.

Human Performance versus Model Performance

We presented psychophysical results indicating that human subjects could reliably discriminate between horizontal and vertical stimuli measuring 1×2 arcmin (100% accuracy; Figure 2), but that the task was barely achievable when the stimulus was half that size (70% accuracy). Using biologically reasonable parameters, a Markov decoder of retinal spike trains attains comparable, but slightly weaker, performance (90% and 60%, respectively; Figure 5). What additional information do humans have that might account for this discrepancy? Here, we consider several aspects of realistic visual processing that were ignored by the Markov decoder.

We treated only Off-type retinal ganglion cells, but there are equally many On-type cells in the fovea, and in principle, they could also contribute to discrimination. An On cell is suppressed when a small, dark stimulus on a light background enters its receptive field, and is then excited when the stimulus exits. These responses are unreliable because the reduction in firing rate from the background of 10 Hz is detectable only after 100 ms of silence, and the excitatory response is slow and weak. We explored this further with explicit simulation of both On and Off cells: The decoder performance improved very little (unpublished data), less than required to fully account for human acuity.

Human fixational eye movements are not exactly random walks. Instead, they exhibit some small persistence of velocity on a timescale of 2 ms and antipersistence on a timescale of 100 ms [25,73]. To explore how these details affect the Markov

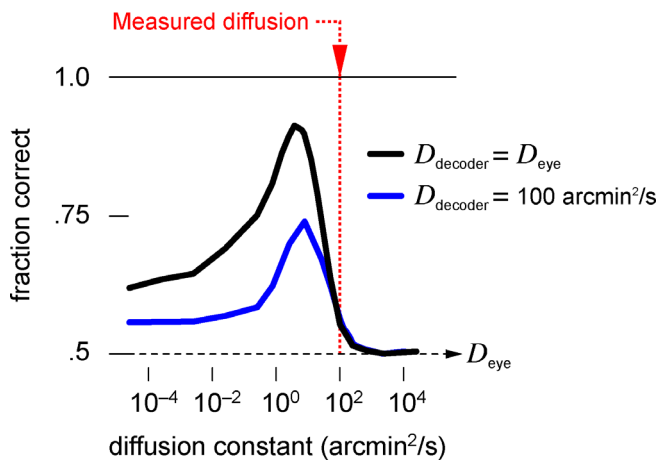


Figure 9. Markov Decoder Discrimination Performance as a Function of Eye Movement Diffusion Constant

The decoder's assumed diffusion constant is either held fixed (blue) or covaried with that of the eye (black). The measured diffusion constant for eye movements is marked in red. These simulations used a biphasic filter with perfectly matched positive and negative lobes, which is the filter that most favors large eye movements. The stimulus measured 0.5×1 arcmin²; otherwise, parameters were as in Figure 5. doi:10.1371/journal.pbio.0050331.g009

decoder's discrimination performance, we performed additional simulations. Antipersistence at long times can be explained by occasional microsaccades that periodically deflect the eye toward its starting position. Such occasional jerks of the image hardly affected a Markov decoder ignorant of microsaccades (unpublished data): After each stimulus jump, there was only a slight delay until the tails of the diffusing posterior distribution encountered the elevated spike rate at the new stimulus location. The persistence of eye movements at short times is consistent with velocity correlations lasting just a few milliseconds [73]. For a given diffusion constant, a persistent random walk lingers longer at each retinal location than a pure Markov random walk, leading to slightly stronger responses. Correspondingly, simulations showed that the Markov decoder's performance improves modestly with the introduction of a short persistence time (unpublished data).

As discussed above, the Markov decoder is suboptimal because of the temporal blurring of the stimulus before spike generation. The optimal decoder must keep track of all possible histories affecting the current firing rate, rather than only the last stimulus position, and the computational effort rapidly becomes prohibitive. Strategies have been proposed to simplify the decoding of such processes [74,75], but these require complicated learning algorithms and do not lend themselves to straightforward neural implementation. A simpler strategy for improving the Markov decoder might be to first process the retinal spike trains with a temporal filter designed to "undo" temporal integration in the retina. This could plausibly take place in the thalamus [76].

Finally, the real visual system enjoys two additional benefits that were not available to the Markov decoder. The first is global image motion: Our human observers viewed the tiny bar stimuli on a white sheet posted within a laboratory scene. As the eye moves, this peripheral background image moves coherently upon the retina, providing additional global motion cues that the brain could perhaps incorporate to

improve perception. Second, our model for retinal responses used the most-random spike pattern for a given firing rate, namely a Poisson process. By contrast, real retinal ganglion cells fire more precisely [11,12] and could thus be more informative, even for a Markov decoder.

Are Fixational Eye Movements Helpful or Harmful?

One commonly held view is that fixational eye movements actually improve vision by preventing the decay of retinal responses that occurs under static stimuli [20]. For example, Rucci and Desbordes have demonstrated that for moderately large, noisy stimuli, orientation discrimination is worse when the image is stabilized on the retina, a result they attribute to a loss of the image motion that would otherwise refresh, and possibly structure, neural activity [77]. In contrast, here we have described these eye movements as a hindrance rather than a help. The transient nature of retinal ganglion cell responses does imply that a fixed stimulus will elicit fewer spikes than a moving stimulus, diminishing the signal that the brain receives. But if the eye movements are too large, then the light intensity is spread thinly over many cells, decreasing each individual response while increasing the positional uncertainty and thus the noise [78]. Between the limits of no eye movement and very large eye movements, an optimum exists. This should occur with eye movements that shift the stimulus to a new set of retinal ganglion cells just as the initial response starts to truncate, and no sooner. For a stimulus area s and transient response duration τ , this occurs when the diffusion constant is $D \sim s/4\tau$. For tiny stimuli ($s = 0.5 \times 1$ arcmin²) and biphasic temporal kernels with $\tau = 35$ ms, the predicted optimum of $D \sim 3$ arcmin²/s is more than a full order of magnitude smaller than the naturally occurring eye movements of approximately 100 arcmin²/s.

To explore this further, we computed the Markov decoder's performance as a function of the eye movement diffusion constant (Figure 9). In one condition, the decoder's assumption about the diffusion constant is held fixed while the eye movement statistics vary; this models a psychophysical experiment in which a viewer's gaze is artificially stabilized. In another condition, the decoder's assumed diffusion constant varies to match the eye movement statistics, approximately optimizing the decoder performance. In both cases, there is an optimum for D near the value predicted above, and the model acuity is dramatically worse than this optimum when the natural diffusion statistics are used. Natural eye movements are therefore substantially larger than optimal for this fine acuity task, implying that they do indeed present a problem for fine visual acuity that the brain must solve.

Predictions

The Markov decoder model yields psychophysical and physiological predictions. We argued that fixational eye movements are unknown to the brain, so using an eye tracker to replace the natural fixational eye movements with exogenous jitter movements, such as eye trajectories recorded from a previous trial, should not affect fine acuity, a prediction supported by recent evidence [79]. We also argued that, for very small stimuli on a featureless background, natural eye movements are larger than optimal: therefore, partially stabilizing the retinal image should improve our finest acuity so long as enough motion remains to avoid

prematurely truncating retinal responses (Figure 9). Although discrimination of larger stimuli does benefit from eye movements [79], there are indications that fine acuity is improved by stabilization [80].

There are two major physiological predictions. First, activity in V1 neurons should locally modulate the gain for feedforward input originating from the retina. Without this modulation, the advantage of using prior expectations is lost. Second, if the neural interactions in V1 are to correctly encode the probabilistic expectations given by random walk eye movement statistics, then the interactions should implement a diffusion operator, which entails that the time delay to reach maximal interaction strength should scale as the square of the interaction distance. This should be observable both directly, as lateral excitatory currents, and indirectly, through the time course of the resulting gain modulation.

The Bayesian Framework

The essential aspect of the Markov decoder we have described is that information of one type attunes the observer to other, related information. In the present context, the decoder expects that responses to oriented line segments are correlated across space and time due to fixational eye movements, and thus these expected responses are enhanced. Other statistical regularities produce expectations as well. For example, strings of line segments often occur together in contours. Correspondingly, collinear iso-orientation facilitation has been hypothesized to subservise contour integration [41,81], and can be viewed as another instance of the principle of enhancing responses to expected signals. More generally, expectations should increase the gain for information that is relevant to the current task, but when that information is irrelevant, then expectations may instead reduce the gain.

The probabilistic processing of information has generated substantial interest as a general framework for neural computation, often designated “Bayesian computation” due to the use of Bayes’ rule in calculating probabilities. Human perception has been shown in several conditions to behave according to this rule [82–84]. Experimental evidence also hints that the cortex may be implementing Bayesian inference on a neural level [85]. Modeling studies have suggested how networks of neurons could make these probabilistic inferences [75,86–89]. One study of particular relevance also describes a neural network for approximately Bayesian decoding of arbitrary hidden Markov processes [90].

Although our mathematical formalism is closely related to previous work, we have made several advances in applying the Bayesian paradigm. First, we identified a concrete biological puzzle of considerable practical importance: how can humans see with high acuity when fixational eye movements rapidly jitter the stimulus over a large area? Second, previous Bayesian computations treated neural signals that were poorly constrained by experiment, so the performance of these computations could be characterized only qualitatively. In contrast, retinal signals are well studied, enabling us to make quantitative comparisons between model and human performance. Third, previous studies predominantly described the formal structure of Bayesian computations, whereas we identified a simple and biologically plausible mapping of the probabilistic calculations onto cortical circuitry.

Outlook

The decoder we have described is optimized for discriminating the orientation of line segments, but human acuity extends to more complex tasks, such as telling “F” from “P.” Within our formalism, optimal discrimination of arbitrary shapes would require receptive fields tuned to those shapes, whereas the early visual system appears to encode oriented edges, with more complex feature selectivity arising only later in higher brain regions. Therefore, this Markov decoder by itself cannot account for discrimination in complex acuity tasks. However, we propose that it functions as a useful preprocessor that reduces the confounding effects of fixational eye movements before passing signals to subsequent cortical regions for high-level processing.

If the stimulus contains several lines of multiple orientations, the decoder’s output will have several peaks that correspond to the individual oriented segments. These peaks will track the stimulus pattern as it is scanned over the retina. This output can then be processed by subsequent networks tuned to more complex patterns. Simulations show that such a pattern detector identifies an arrangement of oriented bars better when it is provided with the output of a Markov decoder than with signals from similar decoders that fail to properly account for eye movements (see figure in Protocol S1). Thus, the Markov decoder elaborates the conventional model of V1 as extracting oriented image elements, and improves over this static model through dynamic processing that partially corrects for eye movements.

In real-world acuity tasks, we do not perceive the incessant motion of the image upon our retinas, but rather perceive a stable image in world coordinates. Nonetheless, our internal representation early in the visual pathway stores visual information in a retinal coordinate system [91]. This moving frame of reference must eventually be superseded before our stable perceptions arise and decisions are reached. The network proposed here could be viewed as creating an intermediate coordinate system: the most current information is represented in retinal coordinates, but the nonlinear operations of the network effectively shift the past retinal coordinates into improved alignment. We may view this neural computation as a step towards invariant world coordinates.

Materials and Methods

Psychophysics. Three groups of small horizontal and vertical stimuli like those in Figure 2 were presented at a distance of 4 m and were scaled to subtend the angles 0.5×1 , 0.75×1.5 , and 1×2 arcmin². Stimuli were printed in black ink on white paper. Ambient lighting generated a luminance of 86 candelas/m² for the white background, and 20-fold dimmer for the black stimuli. Room features provided global motion cues, which we did not seek to eliminate. Nine subjects were asked to discriminate between the stimuli while standing, and were not provided with error feedback. Subjects were free to view the stimuli as long as they liked, typically taking a few seconds per stimulus. Performance was reported as the fraction of correct answers out of 32 attempts for each condition. Error bars were given as 68% confidence interval around the mean, assuming a binomial distribution of correct guesses and a uniform prior over the fraction correct. Other experiments with briefly flashed stimuli showed that reliable discrimination was already achieved within 500 ms (unpublished data).

Simulations. We generated model retinal responses for the discrimination task in the following steps: the stimulus orientation S was chosen randomly to be either horizontal or vertical, a random walk trajectory was constructed, and the stimulus light intensity profile was moved along this random walk trajectory; the dynamic

light intensity at each retinal position was filtered by a temporal kernel, then passed through a threshold rectifier to yield the instantaneous firing rate; this rate drove an inhomogeneous Poisson generator to produce the spike train for the retinal neuron at that location. We passed these spikes to the Markov decoder implementing Equation 1, which returned a guess of the stimulus identity. These steps are depicted in Figure 3A and described in detail below.

In both the simulations of retinal spike trains and in the Markov decoder, we modeled the fovea as a square lattice of cone photoreceptors. In the human retina, cones are spaced every 0.5 arcmin, and the receptive fields of retinal ganglion cells each consist of a single cone. Correspondingly, the model ganglion cells had square receptive fields separated by 0.5 arcmin. For numerical work, we simulated a 16×16 arcmin² array with wraparound boundary conditions, which was sufficiently large for the relevant values of the diffusion constant and the diffusion time, yet small enough for fast simulations.

The stimulus itself consisted of a rectangle with size z and a 1×2 aspect ratio oriented in either the vertical or horizontal direction. Optical blur was produced by convolving the stimulus with a Gaussian modulation transfer function of diameter $2\sigma = 0.5$ arcmin [10]. The stimulus at location \mathbf{x} induced an instantaneous spatial light absorption profile at retinal positions \mathbf{y} of

$$I_S(\mathbf{y}, \mathbf{x}) = e^{-|\mathbf{y}|^2/2\sigma^2} \circ U_{1,1}(\mathbf{y}) \circ U_{z,2z}(\mathbf{x} - \mathbf{y}), \quad (5)$$

where \circ denotes a convolution operation, and $U_{a,b}(\mathbf{x})$ represents a two-dimensional box profile with dimensions a and b . The resultant stimulus profile is shown in Figure 3B.

We modeled fixational eye movements as a random walk that shifts the stimulus across the retina. The one-sided power spectrum of a one-dimensional random walk is given by $D/\pi^2 f^2$, where f is the temporal frequency. Eizenman et al. [3] reported one-sided power spectra with f^{-2} dependence for the horizontal component of fixational eye movements, from which we inferred a two-dimensional diffusion constant of $D = 100$ arcmin²/s. Corroborating results come from direct measurements of squared eye displacement as a function of time lag [25]; fitting these data with a straight line of slope $4D$ expected from a random walk yielded diffusion constants of the same magnitude, 100 arcmin²/s.

We simulated the trajectory of the stimulus as a random walk on a discrete spatial lattice, but continuous in time. After an infinitesimal time interval dt , the probability of stepping to a nearest neighbor location is $dt \cdot D/a^2$, where D is the diffusion constant, and a is the distance between lattice points. After many such time steps over a finite interval Δt , the probability that the walker has moved a distance Δx horizontally and Δy vertically can be expressed in series form:

$$P(\Delta x, \Delta y, \Delta t) = F(\Delta x, \Delta t) \cdot F(\Delta y, \Delta t)$$

$$F(\Delta x, \Delta t) = \frac{1}{N} \sum_{j=0}^{N-1} \exp\left(i2\pi \frac{j\Delta x}{Na}\right) \exp\left[-\frac{2D\Delta t}{a^2} \left(1 - \cos 2\pi \frac{j}{N}\right)\right], \quad (6)$$

where N is the number of points on a side of the square lattice. For speedy simulations, we chose a constant sampling interval $\Delta t = 0.7$ ms and drew independent random walk steps from this distribution; finer temporal sampling produced nearly identical results (unpublished data).

The spatial stimulus profile was moved around the model retina according to the random walk. This produced a temporal sequence of light intensities within each retinal ganglion cell's receptive field, which was then convolved with the parameterized biphasic temporal filter (Figure 3D)

$$h(t) = \frac{t^n}{\tau_1^{n+1}} e^{-t/\tau_1} - \rho \frac{t^n}{\tau_2^{n+1}} e^{-t/\tau_2} \quad (7)$$

to produce a temporally blurred stimulus (Figure 4F). The parameters were chosen as $\tau_1 = 5$ ms, $\tau_2 = 15$ ms, $n = 3$, and $\rho = 0.8$ for all simulations [28] except Figure 9, for which $\rho = 1$ to maximize the performance improvement attributable to eye movements. Finally, this spatiotemporal profile was offset by the background firing rate r_0 , half-wave rectified to prevent negative firing rates, and scaled so that the maximum possible firing rate was given by r_{\max} . The typical firing-rate parameters we used were $r_0 = 10$ Hz and $r_{\max} = 100$ Hz unless otherwise specified.

The Markov decoder operated on one trial of all ganglion cell spike trains to produce a guess for the stimulus identity, according to the differential equation (Equation 1). This equation can be solved

iteratively, moving from spike to spike. When neuron \mathbf{y} produces a spike at time t_y , the diffusion term (Equation 4) is negligible compared to the spiking term (Equation 2), so we have only

$$\frac{\partial}{\partial t} P(S, \mathbf{x}, t) = \delta(t - t_y) f_S(\mathbf{y} - \mathbf{x}) P(S, \mathbf{x}, t) \quad (8)$$

Dividing both sides by $P(S, \mathbf{x}, t)$ and substituting $f_S(\mathbf{x}) = \ln(r_S(\mathbf{x}/r_0))$, we see that

$$\frac{\partial}{\partial t} \ln P(S, \mathbf{x}, t) = \delta(t - t_y) \ln \frac{r_S(\mathbf{y} - \mathbf{x})}{r_0}. \quad (9)$$

Integrating the delta function over the spike from time t_y^- to time t_y^+ we find that the log-probability jumps at spike times by $\ln(r_S(\mathbf{x}/r_0))$, which means that the probability itself is multiplied:

$$P(S, \mathbf{x}, t_y^+) = \frac{r_S(\mathbf{y} - \mathbf{x})}{r_0} P(S, \mathbf{x}, t_y^-) \quad (10)$$

In the absence of spikes, only the terms of Equations 3 and 4 contribute to the differential equation (Equation 1), so the probability distribution $P(S, \mathbf{x}, t)$ both decays and diffuses laterally across space. Because the two oriented stimuli both produce the same total spike rate from the retinal array regardless of position, the decay term (Equation 3) does not alter the relative probabilities, and we therefore neglect it. The diffusion term (Equation 4) can be implemented most efficiently in the spatial frequency domain $\tilde{P}(S, \mathbf{k}, t)$, where the diffusion operator $D\nabla^2$ simply multiplies its operand. The solution to

$$\frac{\partial}{\partial t} \tilde{P}(S, \mathbf{k}, t) = D\nabla^2 \tilde{P}(S, \mathbf{k}, t) \quad (11)$$

during a spike-free interval $[t, t + \Delta t]$ is

$$\tilde{P}(S, \mathbf{k}, t + \Delta t) = \exp\left[D\Delta t \nabla^2\right] \tilde{P}(S, \mathbf{k}, t). \quad (12)$$

For computational speed, we sampled the decoder's activity every 0.7 ms. Between samples, the probability distribution was multiplied in the Fourier domain according to Equation 12, and at the sample times, the probabilities were multiplied in the spatial domain following Equation 10: once for each spike that occurred since the last sample time. Thus we were able to execute the ideal observer algorithm by multiplication alternately in the spatial domain and the frequency domain. To ensure stability in the absence of the decay term (Equation 3), at every sampling time, we rescaled the posterior probability by its sum, $\sum_{S, \mathbf{x}} P(S, \mathbf{x}, t)$, recovering a properly normalized probability.

These estimated posterior probabilities can be displayed as a function of space and time, as in Figure 4. Or to reach a decision in the discrimination task, we summed the probabilities over all positions after the specified stimulus duration T to obtain the posterior probability for orientation, $P(S, T)$; the orientation with the greatest probability counted as the decoder's guess. By repeating this process many times (10^4 iterations) and calculating the fraction of correct trials, we quantified the performance for this ideal strategy for various parameter sets, as plotted in Figures 5, 6, and 9.

Supporting Information

Protocol S1. The Derivation of the Markov Decoder Equation

Found at doi:10.1371/journal.pbio.0050331.sd001 (1.2 MB PDF).

Acknowledgments

The authors thank Ralf Engbert and Reinhold Kliegl for their eye movement data, and Daniel Fisher, Maneesh Sahani, and an anonymous referee for helpful conversations and suggestions.

Author contributions. XP, HS, and MM conceived and designed the experiments, analyzed the data, and wrote the paper. XP performed the experiments.

Funding. XP and MM were supported by a National Institutes of Health grant. The work of HS was partially supported by a grant of the US-Israel Binational Science Foundation.

Competing interests. The authors have declared that no competing interests exist.

References

1. Skavenski AA, Hansen RM, Steinman RM, Winterson BJ (1979) Quality of retinal image stabilization during small natural and artificial body rotations in man. *Vision Res* 19: 675–683.
2. Tomlinson RD (1990) Combined eye-head gaze shifts in the primate. III. Contributions to the accuracy of gaze saccades. *J Neurophysiol* 64: 1873–1891.
3. Eizenman M, Hallett PE, Frecker RC (1985) Power spectra for ocular drift and tremor. *Vision Res* 25: 1635–1640.
4. Guthrie BL, Porter JD, Sparks DL (1983) Corollary discharge provides accurate eye position information to the oculomotor system. *Science* 221: 1193–1195.
5. Donaldson IM (2000) The functions of the proprioceptors of the eye muscles. *Philos Trans R Soc Lond B Biol Sci* 355: 1685–1754.
6. Murakami I, Cavanagh P (1998) A jitter after-effect reveals motion-based stabilization of vision. *Nature* 395: 798–801.
7. Murakami I, Cavanagh P (2001) Visual jitter: evidence for visual-motion-based compensation of retinal slip due to small eye movements. *Vision Res* 41: 173–186.
8. Schein SJ (1988) Anatomy of macaque fovea and spatial densities of neurons in foveal representation. *J Comp Neurol* 269: 479–505.
9. Geisler WS (1984) Physical limits of acuity and hyperacuity. *J Opt Soc Am A* 1: 775–782.
10. Geisler WS, Davila KD (1985) Ideal discriminators in spatial vision: two-point stimuli. *J Opt Soc Am A* 2: 1483–1497.
11. Berry MJ 2nd, Meister M (1998) Refractoriness and neural precision. *J Neurosci* 18: 2200–2211.
12. Uzzell VJ, Chichilnisky EJ (2004) Precision of spike trains in primate retinal ganglion cells. *J Neurophysiol* 92: 780–789.
13. Chichilnisky EJ, Rieke F (2005) Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina. *J Neurosci* 25: 318–330.
14. Hennig MH, Wörgötter F (2004) Eye micro-movements improve stimulus detection beyond the nyquist limit in the peripheral retina. *Adv Neural Inf Process Syst* 16: 1475–1482.
15. Wachtler T, Wehrhahn C, Lee BB (1996) A simple model of human foveal ganglion cell responses to hyperacuity stimuli. *J Comput Neurosci* 3: 73–82.
16. Croner L, Kaplan E (1995) Receptive fields of P and M ganglion cells across the primate retina. *Vision Res* 35: 7–24.
17. Shapley RM, Victor JD (1978) The effect of contrast on the transfer properties of cat retinal ganglion cells. *J Physiol* 285: 275–298.
18. Fréchette ES, Sher A, Grivich MI, Petrusca D, Litke AM, et al. (2005) Fidelity of the ensemble code for visual motion in primate retina. *J Neurophysiol* 94: 119–135.
19. Troy JB, Lee BB (1994) Steady discharges of macaque retinal ganglion cells. *Vis Neurosci* 11: 111–118.
20. Martinez-Conde S, Macknik SL, Hubel DH (2004) The role of fixational eye movements in visual perception. *Nat Rev Neurosci* 5: 229–240.
21. Winterson BJ, Collewijn H (1976) Microsaccades during finely guided visuomotor tasks. *Vision Res* 16: 1387–1390.
22. Kowler E, Steinman RM (1979) Miniature saccades: eye movements that do not count. *Vision Res* 19: 105–108.
23. Bridgeman B, Palca J (1980) The role of microsaccades in high acuity observational tasks. *Vision Res* 20: 813–817.
24. Martinez-Conde S (2006) Fixational eye movements in normal and pathological vision. *Prog Brain Res* 154: 151–176.
25. Engbert R, Kliegl R (2004) Microsaccades keep the eyes' balance during fixation. *Psychol Sci* 15: 431–436.
26. Turing AM (1952) The chemical basis of morphogenesis. *Phil Trans Royal Soc Lon B* 237: 37–72.
27. Schneeweis DM, Schnapf JL (1999) The photovoltage of macaque cone photoreceptors: adaptation, noise, and kinetics. *J Neurosci* 19: 1203–1216.
28. Chichilnisky EJ, Kalmar RS (2002) Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci* 22: 2737–2747.
29. Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148: 574–591.
30. Dow BM, Snyder AZ, Vautin RG, Bauer R (1981) Magnification factor and receptive field size in foveal striate cortex of the monkey. *Exp Brain Res* 44: 213–228.
31. Snodderly DM, Gur M (1995) Organization of striate cortex of alert, trained monkeys (*Macaca fascicularis*): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. *J Neurophysiol* 74: 2100–2125.
32. Talbot SA, Marshall WH (1941) Physiological studies on neural mechanisms of visual localization and discrimination. *Amer J Ophthal* 24: 1255–1263.
33. Tootell RB, Switkes E, Silverman MS, Hamilton SL (1988) Functional anatomy of macaque striate cortex. II. Retinotopic organization. *J Neurosci* 8: 1531–1568.
34. Steinman RM, Collewijn H (1980) Binocular retinal image motion during active head rotation. *Vision Res* 20: 415–429.
35. Hubel DH, Wiesel TN, Stryker MP (1978) Anatomical demonstration of orientation columns in macaque monkey. *J Comp Neurol* 177: 361–380.
36. Polat U, Sagi D (1993) Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Res* 33: 993–999.
37. Xing J, Heeger DJ (2000) Center-surround interactions in foveal and peripheral vision. *Vision Res* 40: 3065–3072.
38. Adini Y, Sagi D, Tsodyks M (1997) Excitatory-inhibitory network in the visual cortex: psychophysical evidence. *Proc Natl Acad Sci U S A* 94: 10426–10431.
39. Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci* 17: 2112–2127.
40. Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J Neurosci* 9: 2432–2442.
41. Sinich LC, Blasdel GG (2001) Oriented axon projections in primary visual cortex of the monkey. *J Neurosci* 21: 4416–4426.
42. Malach R, Amir Y, Harel M, Grinvald A (1993) Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc Natl Acad Sci U S A* 90: 10469–10473.
43. Ts'o DY, Gilbert CD, Wiesel TN (1986) Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *J Neurosci* 6: 1160–1170.
44. Kapadia MK, Ito M, Gilbert CD, 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.
45. Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM (1998) Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature* 391: 580–584.
46. Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J Neurosci* 14: 2545–2568.
47. Hubel DH, Wiesel TN (1974) Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J Comp Neurol* 158: 295–305.
48. Wilson JR, Sherman SM (1976) Receptive-field characteristics of neurons in cat striate cortex: changes with visual field eccentricity. *J Neurophysiol* 39: 512–533.
49. Van Essen DC, Newsome WT, Maunsell JH (1984) The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res* 24: 429–448.
50. Azzopardi P, Cowey A (1993) Preferential representation of the fovea in the primary visual cortex. *Nature* 361: 719–721.
51. Azzopardi P, Cowey A (1996) The overrepresentation of the fovea and adjacent retina in the striate cortex and dorsal lateral geniculate nucleus of the macaque monkey. *Neuroscience* 72: 627–639.
52. McAdams CJ, Maunsell JH (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* 23: 765–773.
53. Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399: 575–579.
54. Koch C, Poggio T (1992) Multiplying with synapses and neurons. In: McKenna T, Davis J, Zornetzer S, editors. *Single neuron computation*. Boston: Academic Press. pp. 315–345.
55. Mel BW (1993) Synaptic integration in an excitable dendritic tree. *J Neurophysiol* 70: 1086–1101.
56. Koch C, Segev I (2000) The role of single neurons in information processing. *Nat Neurosci* 3: 1171–1177.
57. Chance FS, Abbott LF, Reyes AD (2002) Gain modulation from background synaptic input. *Neuron* 35: 773–782.
58. Murphy BK, Miller KD (2003) Multiplicative gain changes are induced by excitation or inhibition alone. *J Neurosci* 23: 10040–10051.
59. Mehaffey WH, Doiron B, Maler L, Turner RW (2005) Deterministic multiplicative gain control with active dendrites. *J Neurosci* 25: 9968–9977.
60. Gabbiani F, Krapp HG, Koch C, Laurent G (2002) Multiplicative computation in a visual neuron sensitive to looming. *Nature* 420: 320–324.
61. Fox K, Sato H, Daw N (1990) The Effect of varying stimulus intensity on NMDA-receptor activity in cat visual cortex. *J Neurophysiol* 64: 1413–1428.
62. Rivadulla C, Sharma J, Sur M (2001) Specific roles of NMDA and AMPA receptors in direction-selective and spatial phase-selective responses in visual cortex. *J Neurosci* 21: 1710–1719.
63. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197.
64. Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17: 8621–8644.
65. Cavanaugh JR, Bair W, Movshon JA (2002) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J Neurophysiol* 88: 2530–2546.
66. Webb BS, Dhruv NT, Solomon SG, Tailby C, Lennie P (2005) Early and late mechanisms of surround suppression in striate cortex of macaque. *J Neurosci* 25: 11666–11675.
67. Lee DK, Itti L, Koch C, Braun J (1999) Attention activates winner-take-all competition among visual filters. *Nat Neurosci* 2: 375–381.
68. Coultrip R, Granger RH, Lynch G (1992) A cortical model of winner-take-all competition via lateral inhibition. *Neural Netw* 5: 47–54.
69. Antón PS, Granger RH, Lynch G (1992) Temporal information processing

- in synapses, cells, and circuits. In: McKenna T, Davis J, Zornetzer S, editors. *Single neuron computation*. Boston: Academic Press. pp. 291–313.
70. Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18: 3870–3896.
 71. Callaway EM (2005) Structure and function of parallel pathways in the primate early visual system. *J Physiol* 566: 13–19.
 72. Shriki O, Hansel D, Sompolinsky H (2003) Rate models for conductance-based cortical neuronal networks. *Neural Comput* 15: 1809–1841.
 73. Mergenthaler K, Engbert R (2007) Modeling the control of fixational eye movements with neurophysiological delays. *Phys Rev Lett* 98: 138104.
 74. Huys QJM, Zemel RS, Natarajan R, Dayan P (2007) Fast population coding. *Neural Comput* 19: 460–497.
 75. Zemel RS, Huys QJM, Natarajan R, Dayan P (2005) Probabilistic computation in spiking populations. In: Saul L, Weiss Y, Bottou L, editors. *Advances in neural information processing systems* 17. Cambridge (Massachusetts): MIT Press. pp. 1609–1616.
 76. Dong DW, Atick JJ (1995) Temporal decorrelation: a theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network* 6: 159–178.
 77. Rucci M, Desbordes G (2003) Contributions of fixational eye movements to the discrimination of briefly presented stimuli. *J Vis* 3: 852–864.
 78. Pelli DG (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. *J Opt Soc Am A* 2: 1508–1532.
 79. Rucci M, Iovin R, Poletti M, Santini F (2007) Miniature eye movements enhance fine spatial detail. *Nature* 447: 851–854.
 80. Riggs LA, Ratliff F, Cornsweet JC, Cornsweet TN (1953) The disappearance of steadily fixated visual test objects. *J Opt Soc Am* 43: 495–501.
 81. Sigman M, Cecchi GA, Gilbert CD, Magnasco MO (2001) On a common circle: natural scenes and Gestalt rules. *Proc Natl Acad Sci U S A* 98: 1935–1940.
 82. Brainard DH, Longere P, Delahunt PB, Freeman WT, Kraft JM, et al. (2006) Bayesian model of human color constancy. *J Vis* 6: 1267–1281.
 83. Kording KP, Ku SP, Wolpert DM (2004) Bayesian integration in force estimation. *J Neurophysiol* 92: 3161–3165.
 84. Stocker AA, Simoncelli EP (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci* 9: 578–585.
 85. Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86: 1916–1936.
 86. Deneve S, Latham PE, Pouget A (2001) Efficient computation and cue integration with noisy population codes. *Nat Neurosci* 4: 826–831.
 87. Deneve S (2005) Bayesian inference in spiking neurons. *Adv Neural Inf Process Syst* 17: 353–360.
 88. Rao RP (2005) Hierarchical Bayesian inference in networks of spiking neurons. *Adv Neural Inf Process Syst* 17: 1113–1120.
 89. Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9: 1432–1438.
 90. Rao RP (2004) Bayesian computation in recurrent neural circuits. *Neural Comput* 16: 1–38.
 91. Gur M, Snodderly DM (1997) Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. *Vision Res* 37: 257–265.
 92. Roorda A, Williams DR (1999) The arrangement of the three cone classes in the living human eye. *Nature* 397: 520–522.

A Neural Computation for Visual Acuity in the Presence of Eye Movements

Xaq Pitkow, Haim Sompolinsky, Markus Meister

– Supporting Material –

Derivation of the Markov decoder equation

In this section we derive the main text's differential equation (1) that describes the Markov decoder for the spike generation process, assuming the retina performs no temporal filtering. We denote by $P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t]}\right)$ the probability that a stimulus of shape S is at location \mathbf{x}_t at time t conditioned on $\mathbf{R}_{[0,t]}$, which represents the retinal response for all neurons over the interval $[0, t]$. We first derive an abstract recursive update equation for the probabilities sampled at a finite time interval dt , then substitute the particular probabilities given by the spike generation model, and finally move to the continuous time limit $dt \rightarrow 0$ that yields Equation (1).

The total response $\mathbf{R}_{[0,t]}$ can be divided into the response \mathbf{R}_t in the current time step and the response history $\mathbf{R}_{[0,t-dt]}$, so that the desired probability is written as

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t]}\right) = P\left(S, \mathbf{x}_t \mid \mathbf{R}_t, \mathbf{R}_{[0,t-dt]}\right) \quad (13)$$

We can then use Bayes' rule,

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t]}\right) = \frac{P\left(\mathbf{R}_t \mid S, \mathbf{x}_t, \mathbf{R}_{[0,t-dt]}\right) P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t-dt]}\right)}{P\left(\mathbf{R}_t \mid \mathbf{R}_{[0,t-dt]}\right)} \quad (14)$$

Since the current responses are assumed to depend only on the present stimulus location, this can be simplified to

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t]}\right) = \frac{P\left(\mathbf{R}_t \mid S, \mathbf{x}_t\right) P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t-dt]}\right)}{Z_t} \quad (15)$$

where $Z_t = P\left(\mathbf{R}_t \mid \mathbf{R}_{[0,t-dt]}\right)$ ensures the proper normalization of the left hand side. The current stimulus position depends only on the previous position because the underlying process is a random walk, and the stimulus does not change its orientation. Therefore,

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t-dt]}\right) = \sum_{\mathbf{x}_{t-dt}} P\left(\mathbf{x}_t \mid \mathbf{x}_{t-dt}\right) P\left(S, \mathbf{x}_{t-dt} \mid \mathbf{R}_{[0,t-dt]}\right) \quad (16)$$

Combining equations (13)–(16), we obtain the recursive update equation

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0,t]}\right) = \frac{1}{Z_t} P\left(\mathbf{R}_t \mid S, \mathbf{x}_t\right) \sum_{\mathbf{x}_{t-dt}} P\left(\mathbf{x}_t \mid \mathbf{x}_{t-dt}\right) P\left(S, \mathbf{x}_{t-dt} \mid \mathbf{R}_{[0,t-dt]}\right) \quad (17)$$

This equation expresses the current posterior stimulus probability as a spatially averaged version of the past stimulus probability, weighted by the current response probabilities and then properly normalized.

Next we substitute the various general probabilities in Equation (17) with the particular probabilities determined by the spike generation model. The factor $P\left(\mathbf{x}_t \mid \mathbf{x}_{t-dt}\right)$ represents the

probability that the stimulus moves from position \mathbf{x}_{t-dt} to position \mathbf{x}_t , which is the probability of a random walk step. This is independent of starting position, so that $P(\mathbf{x}_t | \mathbf{x}_{t-dt}) = P(\mathbf{x}_t - \mathbf{x}_{t-dt})$ and the weighted sum over positions $\sum_{\mathbf{x}_{t-dt}} P(\mathbf{x}_t | \mathbf{x}_{t-dt}) P(S, \mathbf{x}_{t-dt} | \mathbf{R}_{[0, t-dt]})$ becomes a convolution.

In an infinitesimal interval dt , the probability of moving one step to each of the four nearest neighbor locations is $dt \cdot D/a^2$ and the probability of staying in the same location is $1 - 4dt \cdot D/a^2$, where D is the diffusion constant, and a is the distance between lattice points. We can use this to express the convolution as

$$\sum_{\mathbf{x}_{t-dt}} P(\mathbf{x}_t | \mathbf{x}_{t-dt}) P(S, \mathbf{x}_{t-dt} | \mathbf{R}_{[0, t-dt]}) = (1 + D dt \nabla^2) P(S, \mathbf{x}_{t-dt} | \mathbf{R}_{[0, t-dt]}) \quad (18)$$

where ∇^2 denotes the discrete second derivative operator and represents a convolution with the kernel

$$k(\Delta \mathbf{x}) = \frac{1}{a^2} \begin{cases} -4 & \Delta \mathbf{x} = 0 \\ 1 & |\Delta \mathbf{x}| = a \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

The factor $P(\mathbf{R}_t | S, \mathbf{x}_t)$ is the likelihood of the stimulus attributes S and \mathbf{x}_t based on the response \mathbf{R}_t observed in the current time step. Note that \mathbf{R}_t represents the response from the entire retina. Specifically, we take this to be the vector of spike counts produced by each retinal neuron in the time interval dt . With the assumption that all neurons are conditionally independent given the

stimulus orientation and position, the likelihood based on the total retinal response is a product over the likelihoods from each individual spike count $R_{t,y}$:

$$P(\mathbf{R}_t | S, \mathbf{x}_t) = \prod_y P(R_{t,y} | S, \mathbf{x}_t) \quad (20)$$

A retinal neuron at position \mathbf{y} fires at a rate $r_S(\mathbf{y} - \mathbf{x})$ when stimulus S is centered on position \mathbf{x} .

Poisson response statistics give the resultant likelihood as

$$P(\mathbf{R}_t | S, \mathbf{x}_t) = e^{-r_S^{tot}(\mathbf{x}_t) dt} \prod_y \frac{1}{R_{t,y}!} (r_S(\mathbf{y} - \mathbf{x}_t) dt)^{R_{t,y}} \quad (21)$$

where $r_S^{tot}(\mathbf{x}) = \sum_y r_S(\mathbf{y} - \mathbf{x})$ is the total retinal mean firing rate induced by a stimulus S at location \mathbf{x} .

For the final step in this derivation, we consider the continuous time limit, $dt \rightarrow 0$. When the sampling interval dt is small enough, most neurons are silent and at most one retinal neuron at position \mathbf{y} will spike, so that to first order in dt

$$P(\mathbf{R}_t | S, \mathbf{x}_t) = \begin{cases} r_S(\mathbf{y} - \mathbf{x}_t) dt & \text{neuron } \mathbf{y} \text{ fires} \\ 1 - r_S^{tot}(\mathbf{x}_t) dt & \text{no spikes at time } t \end{cases} \quad (22)$$

Let us first consider the case where no neuron fires at time t , combining Equations (17), (18) and (22) to find that

$$P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]}) = \frac{1}{Z_t} (1 - r_S^{tot}(\mathbf{x}_t) dt) (1 + D dt \nabla^2) P(S, \mathbf{x}_{t-dt} | \mathbf{R}_{[0,t-dt]}) \quad (23)$$

The normalization factor Z_t is the sum over all possible states,

$$Z_t = \sum_{S, \mathbf{x}_t} \left(1 - r_S^{tot}(\mathbf{x}_t) dt\right) \left(1 + D dt \nabla^2\right) P\left(S, \mathbf{x}_{t-dt} \mid \mathbf{R}_{[0, t-dt]}\right) \quad (24)$$

We can simplify this sum by defining $\bar{r}^{tot} = \sum_{S, \mathbf{x}_t} r_S^{tot}(\mathbf{x}_t) P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0, t]}\right)$ and noting that the diffusion kernel totals to zero. Thus to first order in dt ,

$$\frac{1}{Z_t} = 1 + \bar{r}_S^{tot}(\mathbf{x}_t) dt \quad (25)$$

Substituting all relevant terms,

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0, t]}\right) = \left(1 + \bar{r}^{tot} dt - r_S^{tot}(\mathbf{x}_t) dt + dt D \nabla^2\right) P\left(S, \mathbf{x}_{t-dt} \mid \mathbf{R}_{[0, t-dt]}\right) \quad (26)$$

so that finally, writing $P_{t+dt} = P_t + dt \frac{dP}{dt}$ we obtain

$$\frac{d}{dt} P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0, t]}\right) = \left(\bar{r}^{tot} - r_S^{tot}(\mathbf{x}_t) + D \nabla^2\right) P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0, t]}\right) \quad (27)$$

Now let us consider what happens when there is a spike at time t by neuron y . In this case,

Equations (17), (18) and (22) imply that

$$P\left(S, \mathbf{x}_t \mid \mathbf{R}_{[0, t]}\right) = \frac{1}{Z_t} r_S(\mathbf{y} - \mathbf{x}_t) dt P\left(S, \mathbf{x}_{t-dt} \mid \mathbf{R}_{[0, t-dt]}\right) \quad (28)$$

where the diffusion term contributes only to second-order in dt and can thus be neglected. The

factor Z_t also changes discontinuously, $Z_t = \bar{r}(\mathbf{y}) dt$ where we have defined

$\bar{r}(\mathbf{y}) = \sum_{S, \mathbf{x}_t} r_S(\mathbf{y} - \mathbf{x}_t) P(S, \mathbf{x}_t | \mathbf{R}_{[0, t-dt]})$. Substituting this normalization into Equation (28) we

find that at spike times there is a discontinuous change in P ,

$$P(S, \mathbf{x}_t | \mathbf{R}_{[0, t]}) = P(S, \mathbf{x}_{t-dt} | \mathbf{R}_{[0, t-dt]}) \frac{r_S(\mathbf{y} - \mathbf{x}_t)}{\bar{r}(\mathbf{y})} \quad (29)$$

We can convert Equation (29) to a differential equation by writing it as

$$\frac{d}{dt} \log P(S, \mathbf{x}_t | \mathbf{R}_{[0, t]}) = \delta(t - t_y) \log \frac{r_S(\mathbf{y} - \mathbf{x}_t)}{\bar{r}(\mathbf{y})} \quad (30)$$

where t_y is the time of firing of retinal neuron \mathbf{y} and $\delta(t - t_y)$ is a Dirac delta-function centered on the spike time.

Combining Equations (27) and (30) we find that, for all times t ,

$$\frac{d}{dt} P(S, \mathbf{x}_t | \mathbf{R}_{[0, t]}) = \sum_{\mathbf{y}} \lambda_{\mathbf{y}}(t) \log \frac{r_S(\mathbf{y} - \mathbf{x}_t)}{\bar{r}(\mathbf{y})} P(S, \mathbf{x}_t | \mathbf{R}_{[0, t]}) + (\bar{r}^{tot} - r_S^{tot}(\mathbf{x}_t) + D\nabla^2) P(S, \mathbf{x}_t | \mathbf{R}_{[0, t]}) \quad (31)$$

where $\lambda_{\mathbf{y}}(t) = \sum_{t_y} \delta(t - t_y)$ is the instantaneous firing rate of neuron \mathbf{y} .

Using this equation, the probability is always properly normalized. However, we may modify two terms to produce an unnormalized version with slightly nicer properties. In particular, we omit the constant decay term proportional to \bar{r}^{tot} , and substitute the background firing rate r_0 in place of $\bar{r}(\mathbf{y})$. This is possible since neither \bar{r}^{tot} nor $\bar{r}(\mathbf{y})$ depend on S or \mathbf{x}_t , so their

modification does not alter the relative probabilities $P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]})$. Defining a weighting for new input spikes,

$$f_S(\mathbf{y} - \mathbf{x}) = \log \frac{r_S(\mathbf{y} - \mathbf{x})}{r_0} \quad (32)$$

which is both invariant over time and spatially localized, we obtain the Markov decoder equation presented in the main text as Equation (1),

$$\frac{d}{dt} P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]}) = \sum_{\mathbf{y}} \lambda_{\mathbf{y}}(t) f_S(\mathbf{y} - \mathbf{x}_t) P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]}) - r_S^{tot}(\mathbf{x}_t) P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]}) + D \nabla^2 P(S, \mathbf{x}_t | \mathbf{R}_{[0,t]}). \quad (33)$$

Figure S1.

The Markov Decoder Is a Useful Preprocessing Step for Identifying More Complex Shapes Than Single Oriented Bars

Here, we demonstrate this with an example task: discrimination of the two “letters” shown in (A). Each letter is a different configuration of two bars. In a first stage, the stimulus is processed by the coupled network of bar detectors discussed in the text, which produces an array of output signals $P(S, \mathbf{x}, t)$. The second stage is a simple translation-invariant pattern detector. It matches the output pattern of the first stage to templates for the two letters, and chooses the letter with the better match. Specifically, we compute

$$Q(L, T) = \prod_{t=1}^T \sum_{\mathbf{x}'} \exp \left(\beta \sum_{\mathbf{x}, S} f_L(S, \mathbf{x} - \mathbf{x}') P(S, \mathbf{x}, t) \right).$$

Here, $P(S, \mathbf{x}, t)$ is the output of the Markov decoder; $f_L(S, \mathbf{x} - \mathbf{x}')$ is the template of bars with orientations S at positions \mathbf{x} if the letter L is located at position \mathbf{x}' , as illustrated in (B); β is a weighting factor that determines how accurately the template needs to be matched. After performing the template match over T time frames, we compare $Q(L_1, T)$ with $Q(L_2, T)$ and choose the letter with the larger value. The performance of this pattern detector depends strongly on how the retinal inputs are processed by the first stage (C). Here, we vary the effective diffusion constant D of the Markov decoder, as in Figure 5: Performance is much better when D is adjusted to the eye movement statistics than when D is zero (red) or very large (blue). For this plot, the simulation parameters are: background firing rate of 10 Hz, peak firing rate of 150 Hz, stimulus shapes with two 2×0.5 arcmin bars, templates that are zero everywhere except at two

points taking value 1, and weighting factor $\beta = 6$. Remaining parameters are the same as in Figure 5.

Figure S1

