Closed-Loop Estimation of Retinal Network Sensitivity by Local Empirical Linearization

Ulisse Ferrari, Christophe Gardella, Olivier Marre, Thierry Mora

eNeuro

New Research

Sensory and Motor Systems

Closed-Loop Estimation of Retinal Network Sensitivity by Local Empirical Linearization

Ulisse Ferrari,[1,*] Christophe Gardella,[1,2,*] Olivier Marre,[1,†] and Thierry Mora[2,†]

DOI:http://dx.doi.org/10.1523/ENEURO.0166-17.2017

[1]Institut de la Vision, Sorbonne Université, INSERM, CNRS, 17 rue Moreau, 75012 Paris, France and [2]Laboratoire de physique statistique, CNRS, Sorbonne Université, Université Paris-Diderot and École normale supérieure (PSL), 24, rue Lhomond, 75005 Paris, France

Abstract

Understanding how sensory systems process information depends crucially on identifying which features of the stimulus drive the response of sensory neurons, and which ones leave their response invariant. This task is made difficult by the many nonlinearities that shape sensory processing. Here, we present a novel perturbative approach to understand information processing by sensory neurons, where we linearize their collective response locally in stimulus space. We added small perturbations to reference stimuli and tested if they triggered visible changes in the responses, adapting their amplitude according to the previous responses with closed-loop experiments. We developed a local linear model that accurately predicts the sensitivity of the neural responses to these perturbations. Applying this approach to the rat retina, we estimated the optimal performance of a neural decoder and showed that the nonlinear sensitivity of the retina is consistent with an efficient encoding of stimulus information. Our approach can be used to characterize experimentally the sensitivity of neural systems to external stimuli locally, quantify experimentally the capacity of neural networks to encode sensory information, and relate their activity to behavior.

Key words: Efficient coding theory; Sensory system; Retina; Closed-loop experiments; Fisher Information

Significance Statement

Understanding how sensory systems process information is an open challenge mostly because these systems have many unknown nonlinearities. A general approach to studying nonlinear systems is to expand their response perturbatively. Here, we apply such a method experimentally to understand how the retina processes visual stimuli. Starting from a reference stimulus, we tested whether small perturbations to that reference (chosen iteratively using closed-loop experiments) triggered visible changes in the retinal responses. We then inferred a local linear model to predict the sensitivity of the retina to these perturbations, and showed that this sensitivity supported an efficient encoding of the stimulus. Our approach is general and could be used in many sensory systems to characterize and understand their local sensitivity to stimuli.

Introduction

An important issue in neuroscience is to understand how sensory systems use their neural resources to represent information. A crucial step toward understanding the sensory processing performed by a given brain area is to characterize its sensitivity (Benichoux et al., 2017), by determining which features of the sensory input are coded in the activity of these sensory neurons, and which features are discarded. If a sensory area extracts a given feature from the sensory scene, any change along that

Received May 12, 2017; accepted October 16, 2017; First published January 16, 2018.

The authors declare no competing financial interests.

Author contributions: O.M. and T.M. designed research; U.F. and C.G. performed research; U.F., C.G., O.M., and T.M. analyzed data; U.F., C.G., O.M., and T.M. wrote the paper.

This work was supported by ANR TRAJECTORY, ANR OPTIMA, ANR IR-REVERSIBLE, the French State program Investissements d'Avenir managed by the Agence Nationale de la Recherche (LIFESENSES: ANR-10-LABX-65), European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 720270, and the National Institutes of Health Grant U01NS090501.

dimension will trigger a noticeable change in the activity of the sensory system. Conversely, if the information about a given feature is discarded by this area, the activity of the area should be left invariant by a change along that feature dimension. To understand which information is extracted by a sensory network, we must determine which changes in the stimulus evoke a significant change in the neural response, and which ones leave the response invariant.

This task is made difficult by the fact that sensory structures process stimuli in a highly nonlinear fashion. At the cortical level, many studies have shown that the response of sensory neurons is shaped by multiple nonlinearities (Machens et al., 2004; Carandini et al., 2005). Models based on the linear receptive field are not able to predict the responses of neurons to complex, natural scenes. This is even true in the retina. While spatially uniform or coarse grained stimuli produce responses that can be predicted by quasi-linear models (Berry and Meister, 1998; Keat et al., 2001; Pillow et al., 2008), stimuli closer to natural scenes (Heitman et al., 2016) or with rich temporal dynamics (Berry et al., 1999; Olveczky et al., 2003) are harder to characterize, as they trigger nonlinear responses in the retinal output. These unknown nonlinearities challenge our ability to model stimulus processing and limit our understanding of how neural networks process information.

Here, we present a novel approach to measure experimentally the local sensitivity of a nonlinear network. Because any nonlinear function can be linearized around a given point, we hypothesized that, even in a sensory network with nonlinear responses, one can still define experimentally a local linear model that can well predict the network response to small perturbations around a given reference stimulus. This local model should only be valid around the reference stimulus, but it is sufficient to predict if small perturbations can be discriminated based on the network response.

This local model allows us to estimate the sensitivity of the recorded network to changes around one stimulus. This local measure characterizes the ability of the network to code different dimensions of the stimulus space, circumventing the impractical task of building a complete accurate nonlinear model of the stimulus-response relationship. Although this characterization is necessarily local and does not generalize to the entire stimulus space, one can hope to use it to reveal general principles that are robust to the chosen reference stimulus.

*U.F. and C.G. contributed equally to this work.
†O.M. and T.M. contributed equally to this work.
Acknowledgements: We thank Stéphane Deny for his help with the experiments and Jean-Pierre Nadal for stimulating discussions and crucial suggestions.
Correspondence should be addressed to either of the following: Olivier Marre at the above address, E-mail: olivier.marre@gmail.com; or Thierry Mora at the above address, E-mail: tmora@lps.ens.fr.
DOI:http://dx.doi.org/10.1523/ENEURO.0166-17.2017

We applied this strategy to the retina. We recorded the activity of a large population of retinal ganglion cells stimulated by a randomly moving bar. We characterized the sensitivity of the retinal population to small stimulus changes, by testing perturbations around a reference stimulus. Because the stimulus space is of high dimension, we designed closed-loop experiments to probe efficiently a perturbation space with many different shapes and amplitudes. This allowed us to build a complete model of the population response in that region of the stimulus space, and to precisely quantify the sensitivity of the neural representation.

We then used this experimental estimation of the network sensitivity to tackle two long-standing issues in sensory neuroscience. First, when trying to decode neural activity to predict the stimulus presented, it is always difficult to know if the decoder is optimal or if it misses some of the available information. We show that our estimation of the network sensitivity gives an upper bound of the decoder performance that should be reachable by an optimal decoder. Second, the efficient coding hypothesis (Attneave, 1954; Barlow, 1961) postulates that neural encoding of stimuli has adapted to represent natural occurring sensory scenes optimally in the presence of limited resources. Testing this hypothesis for sensory structures that perform nonlinear computations on high dimensional stimuli is still an open challenge. Here, we found that the network sensitivity with respect to stimulus perturbations exhibits a peak as a function of the temporal frequency of the perturbation, in agreement with prediction from efficient coding theory. Our method paves the way toward testing efficient coding theory in nonlinear networks.

Materials and Methods

Extracellular recording

Experiments were performed on the adult Long Evans rat of either sex, in accordance with institutional animal care standards. The retina was extracted from the euthanized animal and maintained in an oxygenated Ames' medium (Sigma-Aldrich). The retina was recorded extracellularly on the ganglion cell side with an array of 252 electrodes spaced by 60 μm (Multichannel Systems), as previously described (Marre et al., 2012). Single cells were isolated offline using SpyKING CIRCUS a custom spike sorting algorithm (Yger et al., 2016). We then selected 60 cells that were well separated (no violations of refractory period, i.e., no spikes separated by $<$2 ms), had enough spikes (firing rate larger than 0.5 Hz), had a stable firing rate during the whole experiment, and responded consistently to repetitions of a reference stimulus (see Materials and Methods/Stimulus).

Stimulus

The stimulus was a movie of a white bar on a dark background projected at a refresh rate of 50 Hz with a digital micromirror device. The bar had intensity 7.6×10^{11} photons/cm^{-2}/s^{-1}, and 115-μm width. The bar was horizontal and moved vertically. The bar trajectory consisted in 17034 snippets of 0.9 s consisting in two refer-

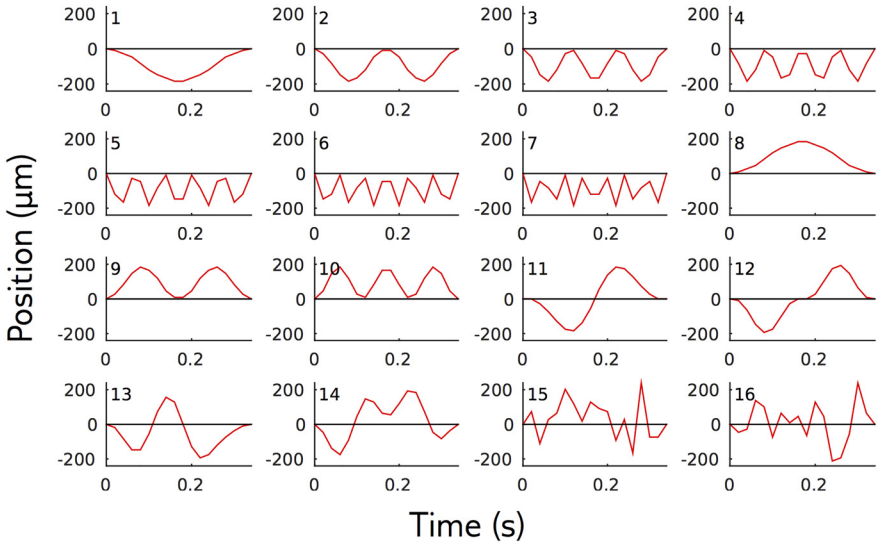

Figure 1. Perturbations shapes. We used the same 16 perturbation shapes for the two reference stimuli. The first 12 perturbation shapes were combinations of two Fourier components, and the last four ones were random combinations of them: $f_k(t) = \cos(2\pi kt/T)$ and $g_k(t) = (1/k) * \sin(2\pi t * k/T)$, with T the duration of the perturbation and $t = 0$ the beginning of the perturbation. The first perturbations $j = 1...7$ were $\mathbf{S}_j = f_j - 1$. For $j = 8,...,10$, they were the opposite of the three first ones: $\mathbf{S}_j = -\mathbf{S}_{j-7}$. For $j = 11, 12$ we used $\mathbf{S}_j = g_{j-10+1} - g_1$. Perturbations 13 and 14 were random combinations of perturbations 1, 2, 3, 11, and 12, constrained to be orthogonal. Perturbations 15 and 16 were random combinations of f_j for $j \in [1,8]$ and g_k for $k \in [1,7]$, allowing higher frequencies than perturbation directions 13 and 14. Perturbation direction 15 and 16 were also constrained to be orthogonal. The largest amplitude for each perturbation we presented was 115 μm. An exception was made for perturbations 15 and 16 applied to the second reference trajectory, as for this amplitude they had a discrimination probability below 70%. They were thus increased by a factor 1.5. The largest amplitude for each perturbation was repeated at least 93 times, with the exception of perturbation 15 (32 times) and 16 (40 times) on the second reference trajectory.

ence trajectories repeated 391 times each, perturbations of these reference trajectories and 6431 random trajectories. Continuity between snippets was ensured by constraining all snippets to start and end in the middle of the screen with velocity 0. Random trajectories followed the statistics of an overdamped stochastic oscillator (Deny et al., 2017). We used a Metropolis-Hastings algorithm to generate random trajectories satisfying the boundary conditions. The two reference trajectories were drawn from that ensemble.

Perturbations

Stimulus perturbations were small changes in the middle portion of the reference trajectory, between 280 and 600 ms. A perturbation is denoted by its discretized time series with time step $\delta t = 20$ ms, $\mathbf{S} = (S_1, ..., S_L)$, with $L = 16$, over the 320 ms of the perturbation (bold symbols represent vectors and matrices throughout). Perturbations can be decomposed as $\mathbf{S} = A \times \mathbf{Q}$, where $A^2 = (1/L)\sum_{i=1}^{L} S_i^2$ is the amplitude, and $\mathbf{Q} = \mathbf{S}/A$ the shape. Perturbations shapes were chosen to have zero value and zero derivative at their boundaries (Fig. 1).

Closed-loop experiments

We aimed to characterize the population discrimination capacity of small perturbations to the reference stimulus. For each perturbation shape (Fig. 1), we searched for the smallest amplitude that will still evoke a detectable change in the retinal response, as we explain below. To do this automatically on the many tested perturbation shapes, we implemented closed-loop experiments (Fig. 3*A*). At each iteration, the retina was stimulated with a perturbed stimulus and the population response was recorded and used to select the next stimulation in real time.

Online spike detection

During the experiment we detected spikes in real time on each electrode independently. Each electrode signal was high-pass filtered using a Butterworth filter with a 200-Hz frequency cutoff. A spike was detected if the electrode potential U was lower than a threshold of five times the median absolute deviation of the voltage (Yger et al., 2016).

Online adaptation of perturbation amplitude

To identify the range of perturbations that were neither too easy nor too hard to discriminate, we adapted perturbation amplitudes so that the linear discrimination probability (see below) converged to target value $D* = 85\%$. For each shape, perturbation amplitudes were adapted using the Accelerated Stochastic Approximation (Kesten, 1958). If an amplitude A_n triggered a response with discrimination probability D_n, then at the next step the perturbation was presented at amplitude A_{n+1} with

$$\ln A_{n+1} = \ln A_n - \frac{C}{r_n + 1}(D_n - D^*), \qquad (1)$$

where $C = 0.74$ is a scaling coefficient that controls the size of steps, and r_n is the number of reversal steps in the

experiment, i.e., the number of times when a discrimination D_n larger than D^* was followed by D_{n+1} smaller than D^*, and vice versa. to explore the responses to different ranges of amplitudes even in the case where the algorithm converged too fast, we also presented amplitudes regularly spaced on a log-scale. We presented the largest amplitude A_{max} (Fig. 1, value), and scaled it down by multiples of 1.4, $A_{max}/1.4^k$ with $k = 1,\ldots,7$.

Online and offline linear discrimination

We applied linear discrimination theory to estimate if perturbed and reference stimuli can be discriminated from the population response they trigger. We applied it twice: online, on the electrode signals to adapt the perturbation amplitude, and offline, on the sorted spikes to estimate the response discrimination capacity. The response $\mathbf{R} = (R_{ib})$ over time of either the $N = 256$ electrodes, or the $N = 60$ cells (the same notation N and $\mathbf{R}$ are used for electrode number and response and cell number and response for mathematical convenience), was binarized into B time bins of size $\delta = 20$ ms: $R_{ib} = 1$ if cell i spiked at least once during the bth time bin, and 0 otherwise. $\mathbf{R}$ is thus a vector of size $N \times B$, labeled by a joint index ib. The response is considered from the start of the perturbation until 280 ms after its end, so that $B = 30$.

To apply linear discrimination on $\mathbf{R_S}$, the response to the perturbation $\mathbf{S}$, we record multiple responses $\mathbf{R}_{ref}$ to the reference, and multiple responses $\mathbf{R}_{\mathbf{S}_{max}}$ to a large perturbation $\mathbf{S}_{max}$, with the same stimulus shape as $\mathbf{S}$ but at the maximum amplitude that was played during the course of the experiment (typically 110 μm; Fig. 1). Our goal is to estimate how close $\mathbf{R_S}$ is to the "typical" $\mathbf{R}_{ref}$ compared to the typical $\mathbf{R}_{\mathbf{S}_{max}}$. To this aim, we compute the mean response to the reference and to the large perturbation, $\langle \mathbf{R}_{ref} \rangle$ and $\langle \mathbf{R}_{\mathbf{S}_{max}} \rangle$, and use their difference as a linear classifier. Specifically, we project $\mathbf{R_S}$ onto the difference between these two mean responses. For a generic response $\mathbf{R}$ (either $\mathbf{R}_{ref}$, $\mathbf{R_S}$ or $\mathbf{R}_{\mathbf{S}_{max}}$), the projection x (respectively, x_{ref}, $x_{\mathbf{S}}$ or $x_{\mathbf{S}_{max}}$) reads:

$$x = \mathbf{u}^{\mathrm{T}} \cdot \mathbf{R} \tag{2}$$

where x is a scalar and $\mathbf{u} = \langle \mathbf{R}_{\mathbf{S}_{max}} \rangle - \langle \mathbf{R}_{ref} \rangle$ is the linear discrimination axis. The computation of x is a projection in our joint index notation, but it can be decomposed in a summation over cells i and consecutive time-bins b of the response: $x = \sum_i \sum_b u_{ib} R_{ib}$. On average, we expect $\langle x_{ref} \rangle < \langle x_{\mathbf{S}} \rangle < \langle x_{\mathbf{S}_{max}} \rangle$. To quantify the discrimination capacity, we compute the probability that $x_{\mathbf{S}} > x_{ref}$, following the classical approach for linear classifiers.

To avoid overfitting, when projecting a response to the reference trajectory, $\mathbf{R}_{ref}$, onto $(\langle \mathbf{R}_{\mathbf{S}_{max}} \rangle - \langle \mathbf{R}_{ref} \rangle)$, we first re-compute $\langle \mathbf{R}_{ref} \rangle$ by leaving out the response of interest. If we did not do this, the discriminability of responses would be overestimated.

In Discussion, Mathematical derivations, we discuss the case of a system with response changes that are linear in the perturbation, or equivalently when the perturbation is small enough so that a linear first order approximation is valid.

Offline discrimination and sensitivity

To measure the discrimination probability as a function of the perturbation amplitude, we consider the difference of the projections, $\Delta x = x_{\mathbf{S}} - x_{ref}$. The response to the stimulation $\mathbf{R_S}$ is noisy, making x and x_{ref} the sum of many random variables (corresponding to each neuron and time bin combinations), and we can apply the central limit theorem to approximate their distributions as Gaussian (Fig. 3*B*, right side), for a given perturbation at a given amplitude. For small perturbations, the mean of Δx grows linearly with the perturbation amplitude A, $\mu = \alpha \times A$, and the variances of $x_{\mathbf{S}}$ and x_{ref} are equal at first order, $\mathrm{Var}(x_{\mathbf{S}}) \approx \mathrm{Var}(x_{ref}) = \sigma^2$, so that the variance of Δx, $\mathrm{Var}(\Delta x) = \mathrm{Var}(x_{\mathbf{S}}) + \mathrm{Var}(x_{ref}) = 2\sigma^2$ is independent of A. Then the probability of discrimination is given by the error function:

$$D = P(x_{ref} < x_{\mathbf{S}}) = \frac{1}{2}(1 + \mathrm{erf}(d'/2)) \tag{3}$$

where $d' = \mu/\sigma = c \times A$ is the standard sensitivity index (Macmillan and Creelman, 2004), and $c = \alpha/\sigma$ is defined as the sensitivity coefficient, which depends on the perturbation shape $\mathbf{Q}$. This coefficient determines the amplitude $A = c^{-1}$ at which discrimination probability is equal to $(1/2)[1 + \mathrm{erf}(1/2)] = 76\%$.

Optimal sensitivity and Fisher information

We then aimed to find the discrimination probability for any perturbation. Given the distributions of responses to the reference stimulus, $P(\mathbf{R}|\mathrm{ref})$, and to a perturbation, $P(\mathbf{R}|\mathbf{S})$, optimal discrimination can be achieved by studying the sign of the response-specific log-ratio $\mathcal{L}(\mathbf{R}) = \ln[P(\mathbf{R}|\mathbf{S})/P(\mathbf{R}|\mathrm{ref})]$. Note that in the log-ratio, $\mathbf{R}$ represents a stochastic response and not the independent variable of a probability density. Because it depends on the response $\mathbf{R}$, this log ratio is both stimulus dependent and stochastic. Let us define $\mathcal{L}_{ref}$ to be the random variable taking value $\mathcal{L}(\mathbf{R})$ on presentation of the reference stimulus, i.e., when $\mathbf{R}$ is a (stochastic) response to the stimulus, and $\mathcal{L}_{\mathbf{S}}$ the random variable taking value $\mathcal{L}(\mathbf{R})$ when $\mathbf{R}$ is a response to the presentation of $\mathbf{S}$. According to the definition given earlier, the probability of successful discrimination is the probability that the log-ratio calculated from a random response to the perturbed stimulus is larger than the log-ratio calculated from a random response to the reference, $\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{ref}$. Using the central limit theorem, we assume again that $\mathcal{L}_{\mathbf{S}}$ and $\mathcal{L}_{ref}$ are Gaussian. We can calculate their mean and variance at small $\mathbf{S}$ (see Discussion, Mathematical derivations): $\mu_{\mathcal{L}} = \langle \mathcal{L}_{\mathbf{S}} \rangle - \langle \mathcal{L}_{ref} \rangle = \mathbf{S}^{\mathrm{T}} \cdot \mathbf{I} \cdot \mathbf{S}$ and $2\sigma_{\mathcal{L}}^2 = \mathrm{Var}(\mathcal{L}_{\mathbf{S}}) + \mathrm{Var}(\mathcal{L}_{ref}) = 2\mathbf{S}^{\mathrm{T}} \cdot \mathbf{I} \cdot \mathbf{S}$, where

$$\mathbf{I} = (I_{tt'}), \quad I_{tt'} = -\sum_{\mathbf{R}} P(\mathbf{R}\,|\,\mathrm{ref}) \frac{\partial^2 \ln P(\mathbf{R}\,|\,\mathbf{S})}{\partial S_t \partial S_{t'}}\Big|_{\mathbf{S}=0} \tag{4}$$

is the Fisher information matrix calculated at the reference stimulus. Following standard discrimination theory (Macmillan and Creelman, 2004; for a derivation in a similar context, see Seung and Sompolinsky, 1993), the discrimination probability is (see Discussion, Mathematical derivations): $D = P(\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{ref}) = (1/2)[1 + \mathrm{erf}(d'/2)]$, with

$$d' = \frac{\mu_{\mathcal{L}}}{\sigma_{\mathcal{L}}} = \sqrt{\mathbf{S}^{\mathsf{T}}\cdot\mathbf{I}\cdot\mathbf{S}}\,. \quad (5)$$

This result generalizes to an arbitrary stimulus dimension the result of Seung and Sompolinsky (1993).

Local model

Because sampling the full response probability distribution $P(\mathbf{R}|\mathbf{S})$ would require estimating $2^{N\times B}$ numbers (one for each possible response $\mathbf{R}$) for each perturbation $\mathbf{S}$, estimating the Fisher Information Matrix directly is impractical, and requires building a model that can predict how the retina responds to small perturbations of the reference stimulus. We used the data from these closed loop experiments for this purpose. The model, schematized in Figure 4*A*, assumes that a linear correction can account for the response change driven by small perturbations. We introduce the local model as a linear expansion of the logarithm of response distribution as a function of both stimulus and response:

$$\begin{aligned}\ln P(\mathbf{R}|\mathbf{S}) &= \ln P(\mathbf{R}|\text{ref}) + \sum_i \sum_{\{t_i\}} \int \mathrm{d}t F_i(t_i, t) S(t) \\ &+ \text{const} = \ln P(\mathbf{R}|\text{ref}) + \sum_{ib,t} R_{ib} F_{ib,t} S_t + \text{const} \\ &= \ln P(\mathbf{R}|\text{ref}) + \mathbf{R}^{T}\cdot\mathbf{F}\cdot\mathbf{S} + \text{const}\,, \end{aligned} \quad (6)$$

where in the integral form, $\{t_i\}$ denotes the set of spiking times of neuron i, and F_i is a stimulus filter depending on both the stimulus time and spiking time (no time-translation invariance). The second line is the discretized version adapted to our binary convention for describing spiking activity binned into bins indexed by b. The matrix $\mathbf{F} = (F_{ib,t})$ is the discretized version of $F_i(t_i, t)$ and contains the linear filters with which the change in the response is calculated from the linear projection of the past stimulus. For ease of notation, hereafter we use matrix multiplications rather than explicit sums over ib and t.

The distribution of responses to the reference trajectory is assumed to be conditionally independent:

$$\ln P(\mathbf{R}\,|\,\text{ref}) = \sum_{ib} \ln P(R_{ib}\,|\,\text{ref})\,. \quad (7)$$

Since the variables R_{ib} are binary, their mean values $\langle R_{ib}\rangle$ on presentation of the reference completely specify $P(R_{ib}|\text{ref})$: $\langle R_{ib}\rangle = P(R_{ib} = 1|\text{ref})$. They are directly evaluated from the responses to repetitions of the reference stimulus, with a small pseudo-count to avoid zero values.

Evaluating the Fisher information matrix (Eq. 4), within the local model (Eq. 6), gives:

$$\mathbf{I} = \mathbf{F}^{\mathsf{T}}\cdot\mathbf{C}_{\mathbf{R}}\cdot\mathbf{F} \quad (8)$$

where $\mathbf{C}_{\mathbf{R}}$ is the covariance matrix of $\mathbf{R}$, which within the model is diagonal because of the assumption of conditional independence.

Inference of the local model

To infer the filters $F_{ib,t}$, we only include perturbations that are small enough to remain within the linear approximation. We first separated the dataset into a training (285 × 16 perturbations) and testing (20 × 16 perturbations) sets. We then defined, for each perturbation shape, a maximum perturbation amplitude above which the linear approximation was no longer considered valid. We selected this threshold by optimizing the model's ability to predict the changes in firing rates in the testing set. Model learning was performed for each cell independently by maximum likelihood with an L_2 smoothness regularization on the shape of the filters, using a pseudo-Newton algorithm. The amplitude threshold obtained from the optimization varied widely across perturbation shapes. The number of perturbations for each shape used in the inference ranged from 20 (7% of the total) to 260 (91% of the total). Overall only 32% of the perturbations were kept (as we excluded repetitions of perturbations with largest amplitude used for calibration). Overfitting was limited: when tested on perturbations of similar amplitudes, the prediction performance on the testing set was never lower than 15% of the performance on the training set.

Linear decoder

We built a linear decoder of the bar trajectory from the population response. The model takes as input the population response $\mathbf{R}$ to the trajectory $X(t)$ and provides a prediction $\hat{X}(t)$ of the bar position in time:

$$\hat{X}(t) = \sum_{i,\tau} K_{i,\tau} R_{i,t-\tau} + \text{constant} \quad (9)$$

where the filters K have a time integration windows of 15 × 20 ms = 300 ms, as in the local model.

We inferred the linear decoder filters by minimizing the mean square error (Warland et al., 1997), $\sum_t [X(t) - \hat{X}(t)]^2$, in the reconstruction of 4000 random trajectories governed by the dynamics of an overdamped oscillator with noise (see Materials and Methods/Stimulus). The linear decoder is then applied to the perturbed trajectories, $X(t) = X_0(t) + S(t)$, where $X_0(t)$ denotes the reference trajectory. The linear decoder does not use prior information about the local structure of the experiment, namely about the fact that the stimulus to decode consists of perturbations around a reference simulation. However, it implicitly uses prior information about the statistics of the overdamped oscillator, as it was trained on bar trajectories with those statistics. Tested on a sequence of ~400 repetitions of one of the two reference trajectories, where the first 300 ms of each have been cut out, we obtain a correlation coefficient of 0.87 between the stimulus and its reconstruction.

Local model Bayesian decoder

To construct a decoder based on the local model, we use Bayes' rule to infer the presented stimulus given the response:

$$P(\mathbf{S}\,|\,\mathbf{R}) = \frac{P(\mathbf{R}\,|\,\mathbf{S})P(\mathbf{S})}{P(\mathbf{R})} \quad (10)$$

where $P(\mathbf{R}|\mathbf{S})$ is given by the local model (Eq. 6), $P(\mathbf{S})$ is the prior distribution over the stimulus, and $P(\mathbf{R})$ is the prior distribution over the stimulus, and $P(\mathbf{R})$ is a normalization factor that does not depend on the stimulus. $P(\mathbf{S})$ is taken to be the distribution of trajectories from an overdamped stochastic oscillator with the same parameters as in the experiment (Deny et al., 2017), to allow for a fair comparison with the linear decoder, which was trained with those statistics. The stimulus is inferred by maximizing the posterior $P(\mathbf{S}|\mathbf{R})$ numerically, using a pseudo-Newton iterative algorithm.

Local signal to noise ratio in decoding

To quantify local decoder performance as a function of the stimulus frequency, we estimated a local signal-to-noise ratio (LSNR) of the decoding signal, LSNR(**S**), which is a function of the reference stimulus. Here, we cannot compute SNR as a ratio between total signal power and noise power, because this would require to integrate over the entire stimulus space, while our approach only provides a model around the neighborhood of the reference stimulus.

To obtain a meaningful comparison between the linear and local decoders, we expand them at first order in the stimulus perturbation and compute the SNR of these "linearized" decoders. For any decoder and for stimuli nearby a reference stimulation, the inferred value of the stimulus, $\hat{\mathbf{X}}$, can be written as $\hat{\mathbf{X}} = \phi(\mathbf{X})$, where $\mathbf{X}$ is the real bar trajectory, and ϕ has a random component (due to the random nature of the response on which the reconstruction relies). Linearizing ϕ for $\mathbf{X} = \mathbf{X}_0 + \mathbf{S}$,

$$\hat{\mathbf{X}} = \phi(\mathbf{X}_0 + \mathbf{S}) \approx \langle\phi(\mathbf{X}_0)\rangle + \mathbf{T}\cdot\mathbf{S} + \epsilon\,, \quad (11)$$

where $\mathbf{T}$ is a transfer matrix which differs from the identity matrix when decoding is imperfect, and ϵ a Gaussian noise of covariance $\mathbf{C}_\epsilon$. Thus, the reconstructed perturbation $\hat{\mathbf{S}} = \hat{\mathbf{X}} - \mathbf{X}_0$ can be written as:

$$\hat{\mathbf{S}} = \mathbf{T}\cdot\mathbf{S} + \mathbf{b} + \epsilon\,, \quad (12)$$

where $\mathbf{b} = \langle\phi(\mathbf{X}_0)\rangle - \mathbf{X}_0$ is a systematic bias. We inferred the values of $\mathbf{b}$ and $\mathbf{C}_\epsilon$ from the ~400 reconstructions of the reference stimulation using either of the two decoders, and the values of $\mathbf{T}$ from the reconstructions of the perturbed trajectories. The inference is done by an iterative algorithm similar to that used for the inference of the filters $\mathbf{F}$ of the local model. We define the LSNR in decoding the perturbation $\mathbf{S}$ as:

$$\mathrm{LSNR}(\mathbf{S}) = (\langle\hat{\mathbf{S}}\rangle - \mathbf{b})^{\mathsf{T}}\cdot\mathbf{C}_\epsilon^{-1}\cdot(\langle\hat{\mathbf{S}}\rangle - \mathbf{b}) = \mathbf{S}^{\mathsf{T}}\cdot\mathbf{T}^{\mathsf{T}}\cdot\mathbf{C}_\epsilon^{-1}\cdot\mathbf{T}\cdot\mathbf{S}\,. \quad (13)$$

where here $\langle\ldots\rangle$ means average with respect to the noise ϵ. In this formula, the signal is defined as the average predicted perturbation $\langle\hat{S}\rangle$, from which the systematic bias $\mathbf{b}$ is subtracted, yielding $\mathbf{T}\cdot\mathbf{S}$. The noise is simply ϵ. Note that here the LSNR is defined for a given perturbation $\mathbf{S}$. It is the ratio of the squared signal to the noise variance (summed over the eigendirections of the noise correlator, since we are dealing with a multidimensional signal). This LSNR gives a measure of decoding performance, through the amplitude of the decoded signal relative to the noise. To study how this performance depends on the frequency ν of the input signal, in Figure 6*C*, we apply Equation 13 with $S_b = A\exp(2\pi i\nu b\delta t)$, where A is the amplitude of the perturbation (Fig. 5*A*), and b is a time-bin counter. Note that this frequency-dependent LSNR should not be intepreted as a ratio of signal and noise power spectra, but rather as the dependence of decoding performance on the frequency of the perturbation. It is used rather than the traditional SNR because we are dealing with signals with no time-translation invariance (i.e., $T_{tt'}$ is not just a function of $t - t'$, and neither is $C_{\epsilon,tt'}$). However, our LNSR reduces to the traditional frequency-dependent SNR in the special case of time-translation invariance, i.e., when the decoder is convolutional, and its noise stationary (see Discussion, Mathematical derivations)

Fisher information estimation of sensitivity coefficients

In Figures 5*A*,*B*, 7*C*,*D*, we show the Fisher estimations of sensitivity coefficients $c(\mathbf{Q})$ for perturbations of different shapes $\mathbf{Q}$, either those used during the experiment (Fig. 1), or oscillating ones, $S_b = A\exp(2\pi i\nu b\delta t)$. to compute these sensitivity coefficients, we use Equation 14 to compute the sensitivity index d' and then we divide it by the perturbation amplitude, yielding $c(\mathbf{Q}) = d'/A = \sqrt{\mathbf{Q}^{\mathsf{T}}\cdot\mathbf{I}\cdot\mathbf{Q}}$.

Results

Measuring sensitivity using closed-loop experiments

We recorded from a population of 60 ganglion cells in the rat retina using a 252-electrode array while presenting a randomly moving bar (Fig. 2*A*; Materials and Methods). Tracking the position of moving objects is major task that the visual system needs to solve. The performance in this task is constrained by the ability to discriminate different trajectories from the retinal activity. Our aim was to measure how this recorded retinal population responded to different small perturbations around a pre-defined stimulus. We measured the response to many repetitions of a short (0.9 s) reference stimulus, as well as many small perturbations around it. The reference stimulus was the random trajectory of a white bar on a dark background undergoing Brownian motion with a restoring force (see Materials and Methods). Perturbations were small changes affecting that reference trajectory in its middle portion, between 280 and 600 ms. The population response was defined as sequences of spikes and silences in 20-ms time bins for each neuron, independently of the number of spikes (see Materials and Methods).

To assess the sensitivity of the retinal network, we asked how well different perturbations could be discriminated from the reference stimulus based on the population response. We expect the ability to discriminate perturbations to depend on two factors. First, the direction of the perturbation in the stimulus space, called

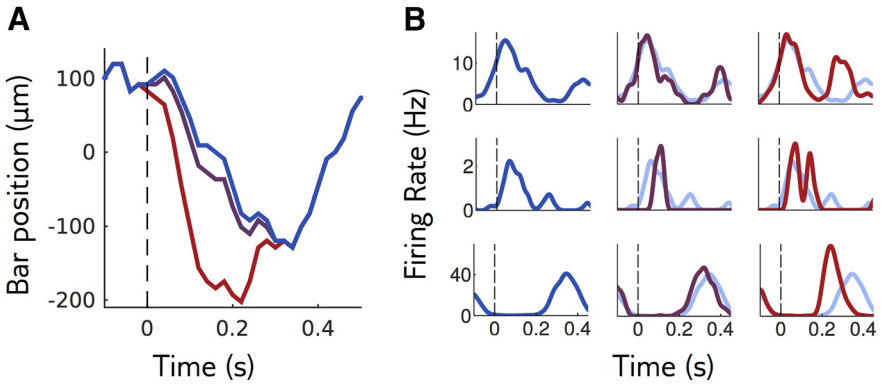

Figure 2. Sensitivity of a neural population to visual stimuli. ***A***, The retina is stimulated with repetitions of a reference stimulus (here the trajectory of a bar, in blue), and with perturbations of this reference stimulus of different shapes and amplitudes. Purple and red trajectories are perturbations with the same shape, of small and large amplitude. ***B***, Mean response of three example cells to the reference stimulus (left column and light blue in middle and right columns) and to perturbations of small and large amplitudes (middle and right columns).

perturbation shape. If we change the reference stimulus by moving along a dimension that is not taken into account by the recorded neurons, we should not see any change in the response. Conversely, if we choose to change the stimulus along a dimension that neurons "care about," we should quickly see a change in the response. The second factor is the amplitude of the perturbation: responses to small perturbations should be hardly distinguishable, while large perturbations should elicit easily detectable changes (Fig. 2*B*). To assess the sensitivity to perturbations of the reference stimulus we need to explore many possible directions that these perturbations can take, and for each direction, we need to find a range of amplitudes that is as small as possible but will still evoke a detectable change in the retinal response. In other words, we need to find the range of amplitudes for which discrimination is hard but not impossible. This requires looking for the adequate range of perturbation amplitudes "online," during the time course of the experiment.

To automatically adapt the amplitude of perturbations to the sensitivity of responses for each of the 16 perturbation shapes and for each reference stimulus, we implemented closed-loop experiments (Fig. 3*A*). At each step, the retina was stimulated with a perturbed stimulus and the population response was recorded. Spikes were detected in real time for each electrode independently by threshold crossing (see Materials and Methods). This coarse characterization of the response is no substitute for spike sorting, but it is fast enough to be implemented in real time between two stimulus presentations, and sufficient to detect changes in the response. This method was used to adaptively select the range of perturbations in real time during the experiment, and to do it for each direction of the stimulus space independently. Proper spike sorting was performed after the experiment using the procedure described in Marre et al., 2012 and Yger et al., (2016) and used for all subsequent analyses.

To test whether a perturbation was detectable from the retinal response, we considered the population response, summarized by a binary vector containing the spiking status of each recorded neuron in each time bin, and projected it onto an axis to obtain a single scalar number. The projection axis was chosen to be the difference between the mean response to a large-amplitude perturbation and the mean response to the reference (Fig. 3*B*). On average, the projected response to a perturbation is larger than the projected response to the reference. However, this may not hold for individual responses, which are noisy and broadly distributed around their mean (for example distributions, see Fig. 3*B*, right). We define the discrimination probability as the probability that the projected response to the perturbation is in fact larger than to the reference. Its value is 100% if the responses to the reference and perturbation are perfectly separable, and 50% if their distributions are identical, in which case the classifier does no better than chance. This discrimination probability is equal to the "area under the curve of the receiver-operating characteristics," which is widely used for measuring the performance of binary discrimination tasks.

During our closed-loop experiment, our purpose was to find the perturbation amplitude with a discrimination probability of 85%. To this end, we computed the discrimination probability online as described above, and then chose the next perturbation amplitude to be displayed using the "accelerated stochastic approximation" method (Kesten, 1958; Faes et al., 2007): when discrimination was above 85%, the amplitude was decreased, otherwise, it was increased (see Materials and Methods).

Figure 3*C* shows the discrimination probability as a function of the perturbation amplitude for an example perturbation shape. Discrimination grows linearly with small perturbations, and then saturates to 100% for large ones. This behavior is well approximated by an error function (gray line) parametrized by a single coefficient, which we call sensitivity coefficient and denote by c. This coefficient measures how fast the discrimination probability increases with perturbation amplitude: the higher the sensitivity coefficient, the easier it is to discriminate responses to small perturbations. It can be interpreted as the inverse of the amplitude at which discrimination

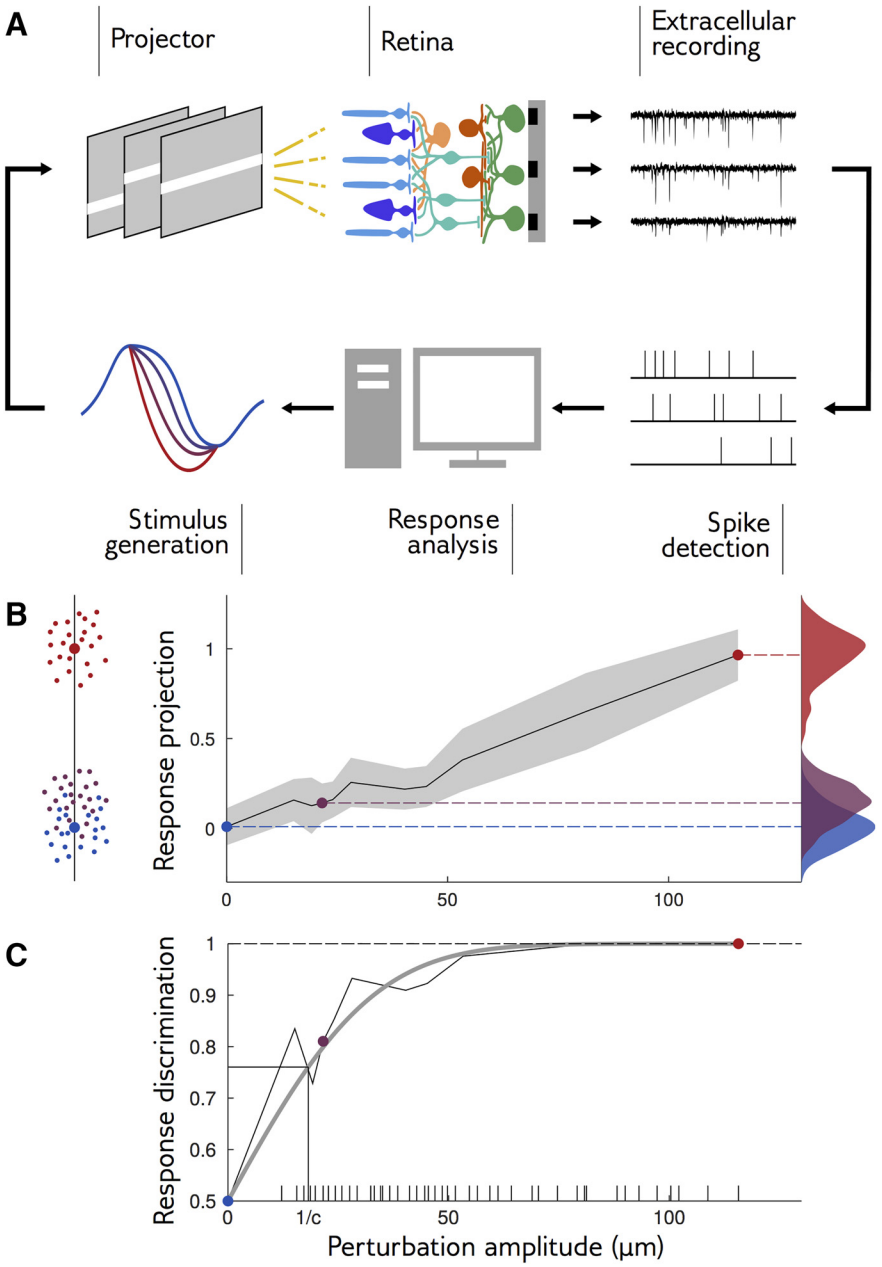

Figure 3. Closed-loop experiments to probe the range of stimulus sensitivity. ***A***, Experimental setup: we stimulated a rat retina with a moving bar. Retinal ganglion cell (RGC) population responses were recorded extracellularly with a multielectrode array. Electrode signals were high-pass filtered and spikes were detected by threshold crossing. We computed the discrimination probability of the population response and adapted the amplitude of the next perturbation. ***B***, left, The neural responses of 60 sorted RGCs are projected along the axis going through the mean response to reference stimulus and the mean response to a large perturbation. Small dots are individual responses, large dots are means. Middle, Mean and standard deviation (in gray) of response projections for different amplitudes of an example perturbation shape. Right, Distributions of the projected responses to the reference (blue), and to small (purple) and large (red) perturbations. Discrimination is high when the distribution of the perturbation is well separated from the distribution of the reference. ***C***, Discrimination probability as a function of amplitude A. The discrimination increases as an error function, $(1/2)[1 + \mathrm{erf}(d'/2)]$, with $d' = c \times A$ (gray line: fit). Ticks on the x-axis show the amplitudes that have been tested during the closed-loop experiment.

reaches 76%, and is related to the classical sensitivity index d' (Macmillan and Creelman, 2004), through $d' = c \times A$, where A denotes the perturbation amplitude (see Materials and Methods).

All 16 different perturbation shapes were displayed, corresponding to 16 different directions in the stimulus space, and the optimal amplitude was searched for each of them independently. We found a mean sensitivity coefficient of $c = 0.0516\ \mu m^{-1}$. However, there were large differences across the different perturbation shapes, with a minimum of $c = 0.028\ \mu m^{-1}$ and a maximum of $c = 0.065\ \mu m^{-1}$.

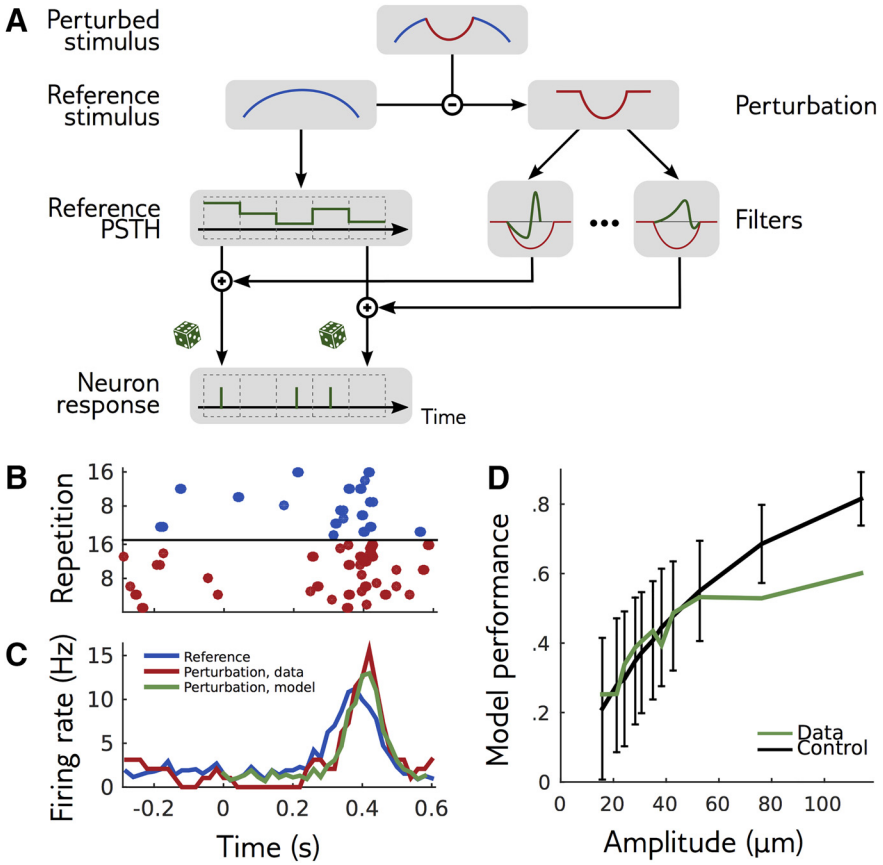

Figure 4. Local model for responses to perturbations. ***A***, The firing rates in response to a perturbation of a reference stimulus are modulated by filters applied to the perturbation. There is a different filter for each cell and each time bin. Because the model is conditionally independent across neurons we show the schema for one example neuron only. ***B***, Raster plot of the responses of an example cell to the reference (blue) and perturbed (red) stimuli for several repetitions. ***C***, PSTH of the same cell in response to the same reference (blue) and perturbation (red). Prediction of the local model for the perturbation is shown in green. ***D***, Performance of the local model at predicting the change in PSTH induced by a perturbation, as measured by Pearson correlation coefficient between data and model, averaged over cells (green). The data PSTH were calculated by grouping perturbations of the same shape and of increasing amplitudes by groups of 20 and computing the mean firing rate at each time over the 20 perturbations of each group. The model PSTH was calculated by mimicking the same procedure. To control for noise from limited sampling, the same performance was calculated from synthetic data of the same size, where the model is known to be exact (black).

Sensitivity and Fisher information

So far, our results have allowed us to estimate the sensitivity of the retina in specific directions of the perturbation space. Can we generalize from these measurements and predict the sensitivity in any direction? The stimulus is the trajectory of a bar and is high dimensional. Under the assumptions of the central limit theorem, we show that the sensitivity can be expressed in matrix form as (see Materials and Methods):

$$d' = \sqrt{\mathbf{S}^{\mathrm{T}} \cdot \mathbf{I} \cdot \mathbf{S}}\,, \qquad (14)$$

where **I** is the Fisher information matrix, of the same dimension as the stimulus, and **S** the perturbation represented as a column vector. This result generalizes that of Seung and Sompolinsky (1993), initially derived for one-dimensional stimuli, to arbitrary dimensions. Thus, the Fisher information is sufficient to predict the code's sensitivity to any perturbation.

Despite the generality of Equation 14, it should be noted that estimating the Fisher information matrix for a highly dimensional stimulus ensemble requires a model of the population response. As already discussed in the introduction, the nonlinearities of the retinal code make the construction of a generic model of responses to arbitrary stimuli a very arduous task, and is still an open problem. However, the Fisher information matrix need only be evaluated locally, around the response to the reference stimulus, and to do so building a local response model is sufficient.

Local model for predicting sensitivity

We introduce a local model to describe the stochastic population response to small perturbations of the reference stimulus. This model will then be used to estimate the Fisher information matrix, and from it the retina's sensitivity to any perturbation, using Equation 14.

The model, schematized in Figure 4*A*, assumes that perturbations are small enough that the response can be linearized around the reference stimulus. First, the response to the reference is described by conditionally independent neurons firing with time-dependent rates es-

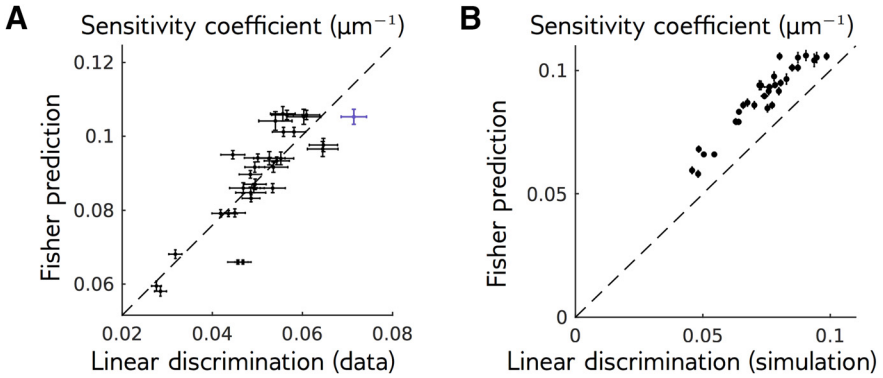

Figure 5. The Fisher information predicts the experimentally measured sensitivity. ***A***, Sensitivity coefficients *c* for the two reference stimuli and 16 perturbation shapes, measured empirically and predicted by the Fisher information (Eq. 14) and the local model. The purple point corresponds to the perturbation shown in Figure 2. Dashed line stands for best linear fit. ***B***, Same as ***A***, but for responses simulated with the local model, with the same amount of data as in experiments. The discriminability of perturbations was measured in the same way than for recorded responses. Dots and error bars stand for mean and SEM over 10 simulations. Dashed line stands for identity.

timated from the peristimulus time histograms (PSTHs). Second, the response to perturbations is modeled as follows: for each neuron and for each 20-ms time bin of the considered response, we use a linear projection of the perturbation trajectory onto a temporal filter to modify the spike rates relative to the reference. These temporal filters were inferred from the responses to all the presented perturbations, varying both in shape and amplitude (but small enough to remain within the linear approximation). Details of the model and its inference are given in Materials and Methods.

We checked the validity of the local model by testing its ability to predict the PSTH of cells in response to perturbations (Fig. 4*B*). To assess model performance, we computed the difference of PSTH between perturbation and reference, and compared it to the model prediction. Figure 4*D* shows the correlation coefficient of this PSTH difference between model and data, averaged over all recorded cells for one perturbation shape. To obtain an upper bound on the attainable performance given the limited amount of data, we computed the same quantity for responses generated by the model (black line). Model performance saturates that bound for amplitudes up to 60 μm, indicating that the local model can accurately predict the statistics of responses to perturbations within that range. For larger amplitudes, the linear approximation breaks down, and the local model fails to accurately predict the response. This failure for large amplitudes is expected if the retinal population responds nonlinearly to the stimulus. We observed the same behavior for all the perturbation shapes that we tested. We have therefore obtained a local model that can predict the response to small enough perturbations in many directions.

To further validate the local model, we combine it with Equation 14 to predict the sensitivity *c* of the network to various perturbations of the bar trajectory, as measured directly by linear discrimination (Fig. 3). The Fisher matrix takes a simple form in the local model: $\mathbf{I} = \mathbf{F}\cdot\mathbf{C_R}\cdot\mathbf{F}^T$, where $\mathbf{F}$ is the matrix containing the model's temporal filters (stacked as row vectors), and $\mathbf{C_R}$ is the covariance matrix of the entire response to the reference stimulus across neurons and time. We can then use the Fisher matrix to predict the sensitivity coefficient using Equation 14, and compare it to the same sensitivity coefficient previously estimated using linear discrimination. Figure 5*A* shows that these two quantities are strongly correlated (Pearson correlation: 0.82, $p = 10^{-8}$), although the Fisher prediction is always larger. This difference could be due to two reasons: limited sampling of the responses, or nonoptimality of the projection axis used for linear discrimination. To evaluate the effect of finite sampling, we repeated the analysis on a synthetic dataset generated using the local model, with the same stimulation protocol as in the actual experiment. The difference in the synthetic data (Fig. 5*B*) and experiment (Fig. 5*A*) were consistent, suggesting that finite sampling is indeed the main source of discrepancy. We confirmed this result by checking that using the optimal discrimination axis (see Discussion, Mathematical derivations) did not improve performance (data not shown).

Summarizing, our estimation of the local model and of the Fisher information matrix can predict the sensitivity of the retinal response to perturbations in many directions of the stimulus space. We now use this estimation of the sensitivity of the retinal response to tackle two important issues in neural coding: the performance of linear decoding and efficient information transmission.

Linear decoding is not optimal

When trying to decode the position of random bar trajectories over time using the retinal activity, we found that a linear decoder (see Materials and Methods) could reach a satisfying performance, confirming previous results (Warland et al., 1997 and Marre et al., 2015). Several works have shown that it was challenging to outperform linear decoding on this task in the retina (Warland et al., 1997 and Marre et al., 2015). From this result, we can wonder whether the linear decoder is optimal, i.e., makes use of all the information present in the retinal activity, or whether this decoder is suboptimal and could be outperformed by a nonlinear decoder. To answer this question, we need to determine an upper bound on the decoding performance reachable by any decoding method. For an

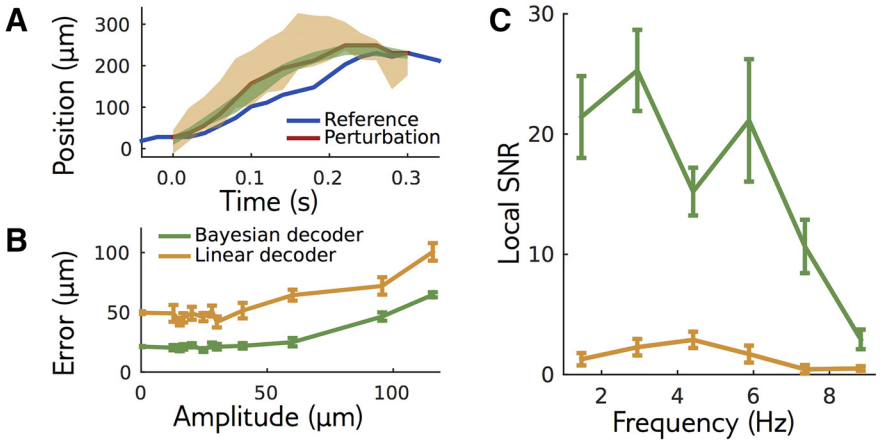

Figure 6. Bayesian decoding of the local model outperforms the linear decoder. ***A***, Responses to a perturbation of the reference stimulus (reference in blue, perturbation in red) are decoded using the local model (green) or a linear decoder (orange). For each decoder, the area shows one standard deviation from the mean. ***B***, Decoding error as a function of amplitude, for an example perturbation shape. ***C***, LSNR for perturbations with different frequencies (differing from the standard SNR definition to deal with locality in stimulus space and in time; Materials And Methods/Local signal to noise ratio in decoding). The performance of both decoders decreases for high frequency stimuli.

encoding model, the lack of reliability of the response sets an upper bound on the encoding model performance, but finding a similar upper bound for decoding is an open challenge. Here, we show that our local model can define such an upper bound.

The local model is an encoding model: it predicts the probability of responses given a stimulus. Yet it can be used to create a "Bayesian decoder" using Bayesian inversion (see Materials and Methods): given a response, what is the most likely stimulus that generated this response under the model? Since the local model predicts the retinal response accurately, doing Bayesian inversion of this model should be the best decoding strategy, meaning that other decoders should perform equally or worse. When decoding the bar trajectory, we found that the Bayesian decoder was more precise than the linear decoder, as measured by the variance of the reconstructed stimulus (Fig. 6*A*). The Bayesian decoder had a smaller error than the linear decoder when decoding perturbations of small amplitudes (Fig. 6*B*). For larger amplitudes, where the local model is expected to break down, the performance of the Bayesian decoder decreased.

To quantify decoding performance as a function of the stimulus temporal frequency, we estimated a "LSNR" of the decoding signal for small perturbations of various frequencies (see Materials and Methods). The definition of the LSNR differs from the usual frequency-dependent SNR, as it is defined to deal with signals that are local in stimulus space and in time, i.e., with no invariance to time translations. We verified however that the two are equivalent when time-translation invariance is satisfied (see Discussion, Mathematical derivations). The Bayesian decoder had a much higher LSNR than the linear decoder at all frequencies (Fig. 6*C*), even if both did fairly poorly at high frequencies. This shows that, despite its good performance, linear decoding misses some information about the stimulus present in the retinal activity. This result suggests that inverting the local model, although it does not provide an alternative decoder generalizable to all possible trajectories, sets a gold standard for decoding, and can be used to test whether other decoders miss a significant part of the information present in the neural activity. It also confirms that the local model is an accurate description of the retinal response to small enough perturbations around the reference stimulus.

Signature of efficient coding in the sensitivity

The structure of the Fisher information matrix shows that the retinal population is more sensitive to some directions of the stimulus space than others. Are these differences in the sensitivity optimal for efficient information transmission? We hypothesized that the retinal sensitivity has adapted to the statistics of the bar motion presented throughout the experiment to best transmit information about its position. Figure 7*A* represents the power spectrum of the bar motion, which is maximum at low frequencies, and quickly decays at large frequencies. We used our measure of the Fisher matrix to estimate the retinal sensitivity power as the sensitivity coefficient c to oscillatory perturbations as a function of temporal frequency (see Materials and Methods). Unlike the power spectrum, which depends monotonously on frequency, we found that the sensitivity is bell shaped, with a peak in frequency around 4Hz (Fig. 7*C*).

To interpret this peak in sensitivity, we studied a minimal theory of retinal function, similar to Van Hateren (1992), to test how maximizing information transmission would reflect on the sensitivity of the retinal response. In this theory, the stimulus is first passed through a low-pass filter, then corrupted by an input white noise. This first stage describes filtering due to the photoreceptors (Ruderman and Bialek, 1992). The photoreceptor output is then transformed by a transfer function and corrupted by a second external white noise, which mimics the subsequent stages of retinal processing leading to ganglion cell activity. Here the output is reduced to a single continuous

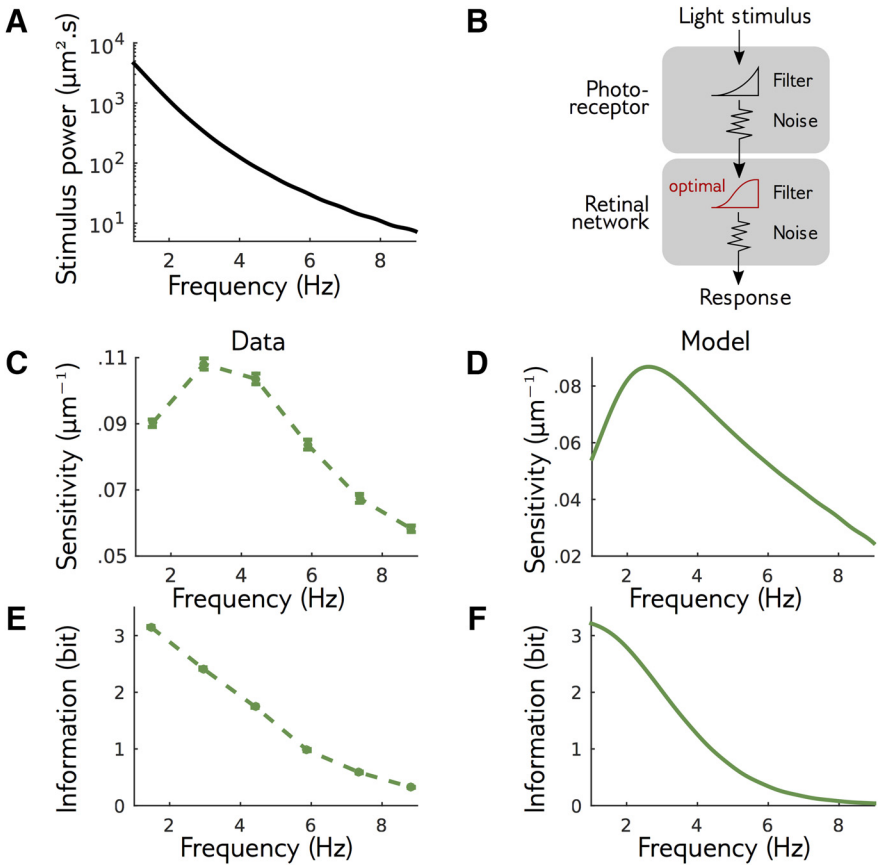

Figure 7. Signature of efficient coding in the sensitivity. ***A***, Spectral density of the stimulus used in experiments, which is monotonically decreasing. ***B***, Simple theory of retinal function: the stimulus is filtered by noisy photoreceptors, whose signal is then filtered by the noisy retinal network. The retinal network filter was optimized to maximize information transfer at constant output power. ***C***, Sensitivity of the recorded retina to perturbations of different frequencies. Note the nonmonotonic behavior. ***D***, Same as ***C***, but for the theory of optimal processing. ***E***, Information transmitted by the retina on the perturbations at different amplitudes. ***F***, Same as ***E***, but for the theory.

signal (Fig. 7*B*; for details, see Discussion, Mathematical derivations). Note that this theory is linear: we are not describing the response of the retina to any stimulus, which would be highly nonlinear, but rather its linearized response to perturbations around a given stimulus, as in our experimental approach. To apply the efficient coding hypothesis, we assumed that the photoreceptor filter is fixed, and we maximized the transmitted information, measured by Shannon's mutual information, over the transfer function (see Discussion, Mathematical derivations; Eq. 31). We constrained the variance of the output to be constant, corresponding to a metabolic constraint on the firing rate of ganglion cells. In this simple and classical setting, this optimal transfer function, and the corresponding sensitivity, can be calculated analytically. Although the power spectrum of the stimulus and photoreceptor output are monotonically decreasing, and the noise spectrum is flat, we found that the optimal sensitivity of the theory is bell shaped (Fig. 7*E*), in agreement with our experimental findings (Fig. 7*C*). Recall that in our reasoning, we assumed that the network optimizes information transmission for the statistics of the stimulus used in the experiment. Alternatively, it is possible that the retinal network optimizes information transmission of natural stimuli, which may have slightly different statistics. We also tested our model with natural temporal statistics (power spectrum $\sim 1/\nu^2$ as a function of frequency ν; Dong and Atick, 1995) and found the same results (data not shown).

One can intuitively understand our result that a bell-shaped sensitivity is desirable from a coding perspective. On one hand, in the small frequency regime, sensitivity increases with frequency, i.e., decreases with stimulus power. This result is classic: when the input noise is small compared to stimulus, the best coding strategy for maximizing information is to whiten the input signal to obtain a flat output spectrum, which is obtained by having the squared sensitivity be inversely proportional to the stimulus power (Rieke et al., 1996; Wei and Stocker, 2016). On the other hand, at high frequencies, the input noise is too high (relative to the stimulus power) for the stimulus to be recovered. Allocating sensitivity and output power to those frequencies is therefore a waste of resources, as it is devoted to amplifying noise, and sensitivity should remain low to maximize information. A peak of sensitivity is thus found between the high SNR region, where stimulus dominates noise and whitening is the best strategy, and the low LSNR region, where information is lost into

the noise and coding resources should be scarce. A result of this optimization is that the information transferred should monotonically decrease with frequency, just as the input power spectrum does (Fig. 7*F*). We tested if this prediction was verified in the data. We estimated similarly the information rate against frequency in our data, and found that it was also decreasing monotonically (Fig. 7*D*). The retinal response has therefore organized its sensitivity across frequencies in a manner that is consistent with an optimization of information transmission across the retinal network.

Discussion

We have developed an approach to characterize experimentally the sensitivity of a sensory network to changes in the stimulus. Our general purpose was to determine which dimensions of the stimulus space most affect the response of a population of neurons, and which ones leave it invariant, a key issue to characterize the selectivity of a neural network to sensory stimuli. We developed a local model to predict how recorded neurons responded to perturbations around a defined stimulus. With this local model we could estimate the sensitivity of the recorded network to changes of the stimulus along several dimensions. We then used this estimation of network sensitivity to show that it can help define an upper bound on the performance of decoders of neural activity. We also showed that the estimated sensitivity was in agreement with the prediction from efficient coding theory.

Our approach can be used to test how optimal different decoding methods are. In our case, we found that linear decoding, despite its very good performance, was far from the performance of the Bayesian inversion of our local model, and therefore far from optimal. This result implies that there should exist nonlinear decoding methods that outperform linear decoding (Botella-Soler et al., 2016). Testing the optimality of the decoding method is crucial for brain machine interfaces (Gilja et al., 2012): in this case, an optimal decoder is necessary to avoid missing a significant amount of information. Building our local model is a good strategy for benchmarking different decoding methods.

In the retina, efficient coding theory had led to key predictions about the shape of the receptive fields, explaining their spatial extent (Atick, 1992; Borghuis et al., 2008), or the details of the overlap between cells of the same type (Liu et al., 2009; Karklin and Simoncelli, 2011; Doi et al., 2012). However, when stimulated with complex stimuli like a fine-grained image, or irregular temporal dynamics, the retina exhibits a nonlinear behavior (Gollisch and Meister, 2010). For this reason, up to now, there was no prediction of the efficient theory for these complex stimuli. Our approach circumvents this barrier, and shows that the sensitivity of the retinal response is compatible with efficient coding. Future works could use a similar approach with more complex perturbations added on top of natural scenes to characterize the sensitivity to natural stimuli.

More generally, different versions of the efficient coding theory have been proposed to explain the organization of several areas of the visual system (Dan et al., 1996; Olshausen and Field, 1996; Bell and Sejnowski, 1997; Bialek et al., 2006; Karklin and Simoncelli, 2011) and elsewhere (Machens et al., 2001; Chechik et al., 2006; Smith and Lewicki, 2006; Kostal et al., 2008). Estimating Fisher information using a local model could be used in other sensory structures to test the validity of these hypotheses.

Finally, the estimation of the sensitivity along several dimensions of the stimulus perturbations allows us to define which changes of the stimulus evoke the strongest change in the sensory network, and which ones should not make a big difference. Similar measures could in principle be performed at the perceptual level, where some pairs of stimuli are perceptually indistinguishable, while others are well discriminated. Comparing the sensitivity of a sensory network to the sensitivity measured at the perceptual level could be a promising way to relate neural activity and perception.

Mathematical derivations

A Derivation of discrimination coefficient in arbitrary dimension

Here, we derive Equation 5 in detail. Recall that $\mathcal{L}_{\text{ref}}$ is a random variable taking value $\mathcal{L}(\mathbf{R}) = \ln[P(\mathbf{R}|\mathbf{S})/P(\mathbf{R}|\text{ref})]$ on presentation of the reference stimulus and $\mathcal{L}_{\mathbf{S}}$ the random variable taking value $\mathcal{L}(\mathbf{R})$ when $\mathbf{R}$ is a response to the presentation of $\mathbf{S}$. Then their averages are given by:

$$\langle \mathcal{L}_{\mathbf{S}} \rangle = \sum_{\mathbf{R}} P(\mathbf{R}|\mathbf{S})[\ln P(\mathbf{R}|\mathbf{S}) - \ln P(\mathbf{R}|\text{ref})] \quad (15)$$

$$\langle \mathcal{L}_{\text{ref}} \rangle = \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref})[\ln P(\mathbf{R}|\mathbf{S}) - \ln P(\mathbf{R}|\text{ref})] \,. \quad (16)$$

Expanding at small $\mathbf{S}$, $P(\mathbf{R}|\mathbf{S}) \approx P(\mathbf{R}|\text{ref})(1 + \partial \ln P(\mathbf{R}|\mathbf{S})/\partial \mathbf{S}^{\mathsf{T}}|_{\mathbf{S}=0} \cdot \mathbf{S})$, one obtains:

$$\langle \mathcal{L}_{\mathbf{S}} \rangle - \langle \mathcal{L}_{\text{ref}} \rangle = \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref}) \left(\frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}^T}\bigg|_{\mathbf{S}=0} \cdot \mathbf{S} \right) \left(\frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}^T}\bigg|_{\mathbf{S}=0} \cdot \mathbf{S} \right) = \mathbf{S}^{\mathsf{T}} \cdot \mathbf{I} \cdot \mathbf{S} + \mathcal{O}(\mathbf{S}^3) \,, \quad (17)$$

with

$$\begin{aligned} \mathbf{I} = (I_{tt'}), I_{tt'} &= \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref}) \frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_t}\bigg|_{\mathbf{S}=0} \frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0} \\ &= \sum_{\mathbf{R}} \frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_t}\bigg|_{\mathbf{S}=0} \frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0} \\ &= \frac{\partial}{\partial S_t} \sum_{\mathbf{R}} P(\mathbf{R}|\mathbf{S}) \frac{\partial \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0} - \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref}) \frac{\partial^2 \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_t \partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0} \\ &= \frac{\partial}{\partial S_t \partial S_{t'}} \sum_{\mathbf{R}} P(\mathbf{R}|\mathbf{S})\bigg|_{\mathbf{S}=0} - \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref}) \frac{\partial^2 \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_t \partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0} \\ &= \sum_{\mathbf{R}} P(\mathbf{R}|\text{ref}) \frac{\partial^2 \ln P(\mathbf{R}|\mathbf{S})}{\partial \mathbf{S}_t \partial \mathbf{S}_{t'}}\bigg|_{\mathbf{S}=0}, \end{aligned} \quad (18)$$

where we have used $\sum_{\mathbf{R}} P(\mathbf{R}|\mathbf{S}) = 1$. Similarly, the variances of these quantities are at leading order:

$$\langle \mathcal{L}_{\text{ref}}^2 \rangle - \langle \mathcal{L}_{\text{ref}} \rangle^2 \approx \langle \mathcal{L}_{\mathbf{S}}^2 \rangle - \langle \mathcal{L}_{\mathbf{S}} \rangle^2 \approx \langle \mathcal{L}_{\mathbf{S}}^2 \rangle \approx \sum_{\mathbf{R}} P(\mathbf{R} \mid \text{ref}) \left(\frac{\partial \ln P(\mathbf{R} \mid \mathbf{S})}{\partial \mathbf{S}^T} \Big|_{\mathbf{S}=0} \cdot \mathbf{S} \right)^2 = \mathbf{S}^{\mathsf{T}} \cdot \mathbf{I} \cdot \mathbf{S} + \mathcal{O}(\mathbf{S}^3) , \quad (19)$$

where we have used the fact that

$$\langle \mathcal{L}_{\mathbf{S}} \rangle = \sum_{\mathbf{R}} P(\mathbf{R} \mid \mathbf{S}) \left(\frac{\partial \ln P(\mathbf{R} \mid \mathbf{S})}{\partial \mathbf{S}^T} \Big|_{\mathbf{S}=0} \cdot \mathbf{S} \right) + \mathcal{O}(\mathbf{S}^2) = \frac{\partial}{\partial \mathbf{S}^T} \sum_{R} P(\mathbf{R} \mid \mathbf{S})_{\mathbf{S}=0} \cdot \mathbf{S} + \mathcal{O}(\mathbf{S}^2) = \mathcal{O}(\mathbf{S}^2) . \quad (20)$$

Next, we assume that $\ln P(\mathbf{R} \mid \mathbf{S})$ is the sum of weakly correlated variables, meaning that its distribution can be approximated as Gaussian. Thus, the random variable $\mathcal{L}_{\mathbf{S}} - \mathcal{L}_{\text{ref}}$ is also distributed as a Gaussian, with mean $\mu_{\mathcal{L}} = \mathbf{S}^{\mathsf{T}} \cdot \mathbf{I} \cdot \mathbf{S}$ and variance $\sigma_{\mathcal{L}}^2 = 2\mathbf{S}^{\mathsf{T}} \cdot \mathbf{I} \cdot \mathbf{S}$. The discrimination probability is the probability that $\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{\text{ref}}$, i.e.,

$$P(\mathcal{L}_{\mathbf{S}} - \mathcal{L}_{\text{ref}} > 0) = \int_0^\infty \frac{dx}{\sqrt{2\pi\sigma_{\mathcal{L}}}} e^{-(x - \mu_{\mathcal{L}})^2 / 2\sigma_{\mathcal{L}}^2} = \frac{1}{2} \left(1 + \text{erf} \left(\frac{\mu_{\mathcal{L}}}{2\sigma_{\mathcal{L}}} \right) \right) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{d'}{2} \right) \right) , \quad (21)$$

with $d' \equiv \mu_{\mathcal{L}} / \sigma_{\mathcal{L}} = \sqrt{\mathbf{S}^{\mathsf{T}} \cdot \mathbf{I} \cdot \mathbf{S}}$.

B Fisher and linear discrimination

There exists a mathematical relation between the Fisher information of Equation 8 and linear discrimination. The linear discrimination task described earlier can be generalized by projecting the response difference, $\mathbf{R}_{\mathbf{S}} - \mathbf{R}_{\text{ref}}$, along an arbitrary direction u:

$$\Delta x = x_{\mathbf{S}} - x_{\text{ref}} = \mathbf{u}^T \cdot (\mathbf{R}_{\mathbf{S}} - \mathbf{R}_{\text{ref}}) . \quad (22)$$

Δx is again assumed to be Gaussian by virtue of the central limit theorem. We further assume that perturbations $\mathbf{S}$ are small, so that $\langle \mathbf{R}_{\mathbf{S}} \rangle - \langle \mathbf{R}_{\text{ref}} \rangle \approx (\partial \langle \mathbf{R}_{\mathbf{S}} \rangle / \partial \mathbf{S}) \cdot \mathbf{S}$, and that $\mathbf{C}_{\mathbf{R}}$ does not depend on $\mathbf{S}$. Calculating the mean and variance of Δx under these assumption gives an explicit expression of d' in Equation 3:

$$d' = \frac{\mathbf{u}^T \cdot \frac{\partial \langle \mathbf{R}_{\mathbf{S}} \rangle}{\partial \mathbf{S}} \cdot \mathbf{S}}{\sqrt{\mathbf{u}^T \cdot \mathbf{C}_{\mathbf{R}} \cdot \mathbf{u}}} . \quad (23)$$

Maximizing this expression of d' over the direction of projection $\mathbf{u}$ yields $\mathbf{u} = \text{const} \times \mathbf{C}_{\mathbf{R}}^{-1} \cdot (\partial \langle \mathbf{R}_{\mathbf{S}} \rangle / \partial \mathbf{S}) \cdot \mathbf{S}$ and

$$d' = \sqrt{\mathbf{S}^{\mathsf{T}} \cdot \mathbf{I}_L \cdot \mathbf{S}} , \quad (24)$$

where $\mathbf{I}_L = (\partial \langle \mathbf{R}_{\mathbf{S}} \rangle / \partial \mathbf{S})^{\mathsf{T}} \cdot \mathbf{C}_{\mathbf{R}}^{-1} \cdot (\partial \langle \mathbf{R}_{\mathbf{S}} \rangle / \partial \mathbf{S})$ is the linear Fisher information (Fisher, 1936; Beck et al., 2011). This expression of the sensitivity corresponds to the best possible discrimination based on a linear projection of the response.

Within the local linear model defined above, one has $\partial \langle \mathbf{R}_{\mathbf{S}} \rangle / \partial \mathbf{S} = \mathbf{F} \cdot \mathbf{C}_{\mathbf{R}}$, and $\mathbf{I}_L = \mathbf{F} \cdot \mathbf{C}_{\mathbf{R}} \cdot \mathbf{F}^T$, which is also equal to the true Fisher information (Eq. 8): $\mathbf{I} = \mathbf{I}_L$. Thus, if the local model (Eq. 6) is correct, discrimination by linear projection of the response is optimal and saturates the bound given by the Fisher information.

Note that the optimal direction of projection only differs from the direction we used in the experiments, $\mathbf{u} = \langle \mathbf{R}_{\mathbf{S}} \rangle - \langle \mathbf{R}_{\text{ref}} \rangle$, by an equalization factor $\mathbf{C}_{\mathbf{R}}^{-1}$. We have checked that applying that factor only improves discrimination by a few percents (data not shown).

C Local SNR for a convolutional linear decoder

In this section, we show how the local SNR defined in Equation 13 reduces to standard expression in the simpler case of a convolution decoder ϕ in the linear regime:

$$\hat{\mathbf{X}} = \phi(\mathbf{X}) = \mathbf{h} \bigstar \mathbf{X} + \epsilon \quad (25)$$

where $\bigstar$ is the convolution symbol, h is a stimulus independent linear filter and ϵ a Gaussian noise of covariance $\mathbf{C}_\epsilon$ and zero mean. Linearizing ϕ for $\mathbf{X} = \mathbf{X}_0 + \mathbf{S}$ as in Equation 12, we obtain

$$\hat{\mathbf{S}} = \mathbf{T} \cdot \mathbf{S} + \mathbf{b} + \epsilon , \quad (26)$$

but now the transfer matrix $T_{bb'} = h(b - b')$ depends only on the difference between the time-bin indices b and b' When T is applied to an oscillating perturbation of unitary amplitude $\hat{S}_b(\nu) \equiv \exp(2\pi i \nu b \delta t)$, we have:

$$\mathbf{T} \cdot \mathbf{S}(\nu) = \tilde{h}(\nu) \mathbf{S}(\nu) \quad (27)$$

where $\tilde{h}(\nu) \equiv \sum_\tau h(\tau) \exp(2\pi i \nu \tau \delta t)$ is the Fourier coefficient of filter h. As a consequence of this last property, the LSNR takes the following expression (Eq. 13):

$$\text{LSNR}(\mathbf{S}(\nu)) = \mathbf{S}(\nu)^{\mathsf{T}} \cdot \mathbf{T}^{\mathsf{T}} \cdot \mathbf{C}_\epsilon^{-1} \cdot \mathbf{T} \cdot \mathbf{S}(\nu) \quad (28)$$

$$= |\tilde{h}(\nu)|^2 \mathbf{S}(\nu)^{\mathsf{T}} \cdot \mathbf{C}_\epsilon^{-1} \cdot \mathbf{S}(\nu) , \quad (29)$$

where $|\tilde{h}(\nu)|^2$ can be interpreted as the signal power at frequency ν for unitary stimulus perturbation. If furthermore $C_{\epsilon, bb'} \equiv \langle \epsilon_b \epsilon_{b'} \rangle = C_\epsilon(b - b')$, then $\text{LSNR}(\mathbf{S}(\nu))$ reduces to the standard expression of SNR (Woyczynski, 2010):

$$\text{LSNR}(\mathbf{S}(\nu)) = \frac{|\tilde{h}(\nu)|^2}{\tilde{C}(\nu)} \quad (30)$$

where $\tilde{C}_\epsilon(\nu) \equiv \sum_\tau C_\epsilon(\tau) \exp(2\pi i \nu \tau \delta t)$ is the noise power at frequency ν.

D frequency dependence of sensitivity and information

To analyze the behavior in frequency of the sensitivity, we compute the sensitivity index for an oscillating perturbation of unitary amplitude. We apply Equation 14 with $\hat{S}_b(\nu) \equiv \exp(2\pi i \nu b \delta t)$. to estimate the spectrum of the information rate we compute its behavior within the linear theory (Van Hateren, 1992):

$$\mathrm{MI}(\nu) = \frac{1}{2}\ln\left[1 + C_{\mathbf{s}}(\nu)I(\nu)/\delta t^2\right] \quad (31)$$

where $C_{\mathbf{s}}(\nu)$ is the power spectrum of the actual stimulus statistics (noisy damped oscillator), and $I(\nu) = (\delta t/L)\widehat{\mathbf{S}}^{\dagger}(\nu)\cdot\mathbf{I}\cdot\widehat{\mathbf{S}}(\nu)$. Note that this decomposition in frequency of the transmitted information is valid because the system is linear and the stimulus is Gaussian distributed (Bernardi and Lindner, 2015).

E efficient coding theory

To build a theory of retinal sensitivity, we follow closely the approach of Van Hateren (1992). The stimulus is first linearly convolved with a filter f, of power $\mathcal{F}$, then corrupted by an input white noise with uniform power H, then convolved with the linear filter r of the retina network of power $\mathcal{F}$, and finally corrupted again by an external white noise Γ. The output power spectrum $O(\nu)$ can be expressed as a function of frequency ν:

$$O(\nu) = (\delta tL)\mathcal{G}(\nu)[(\delta tL)\mathcal{F}(\nu)C_S(\nu) + H] + \Gamma \quad (32)$$

where $C_S(\nu)$ is the power spectrum of the input. The information capacity of such a noisy input-output channel is limited by the allowed total output power $V = \sum_{\nu} O(\nu)$, which can be interpreted as a constraint on the metabolic cost. The efficient coding hypothesis consists in finding the input-output relationship g^*, of power $\mathcal{G}^*(\nu)$, that maximizes the information transmission under a constraint on the total power of the output. The optimal Fisher information matrix can be computed in the frequency domain as:

$$I(\nu) = \frac{\delta t^4 L^2 \mathcal{G}^*(\nu)\mathcal{F}(\nu)}{\Gamma + L\delta t\mathcal{G}^*(\nu)H}. \quad (33)$$

The photoreceptor filter (Warland et al., 1997) was taken to be exponentially decaying in time, $f = \tau^{-1}\exp(-t/\tau)$ (for $t \geq 0$), with $\tau = 100$ ms. The curve $I(\nu)$ only depends on H, Γ, and V through two independent parameters. For the plots in Figure 7, we chose: $H = 3.38\ \mu\mathrm{m}^2/\mathrm{s}$, $\Gamma = 0.02$ spikes$^2 s$ and $V = 307$ spikes$^2 s$, $\delta t = 20\ ms$, and $L = 2, 500$. In Figure 7*D*, we plot the sensitivity to oscillating perturbation with fixed frequency ν, which results in $\sqrt{I(\nu)L/\delta t}$. In Figure 7*E*, we plot the spectral density of the transferred information rate:

$$\mathrm{MI}(\nu) = \frac{1}{2}\ln\left[1 + \frac{(\delta tL)^2\mathcal{G}(\nu)\mathcal{F}(\nu)C_S(\nu)}{\Gamma + (\delta tL)\mathcal{G}(\nu)H}\right]. \quad (34)$$

References

Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? Netw Comput Neural Syst 3:213–251. CrossRef

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193. CrossRef

Barlow H (1961) Possible principles underlying the transformations of sensory messages. Sens Commun 6:57–58.

Beck J, Bejjanki V, Pouget A (2011) Insights from a simple expression for linear Fisher information in a recurrently connected population of spiking neurons. Neural Comput 23:1484–1502. CrossRef

Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. Vision Res 37:3327–3338. Medline

Benichoux V, Brown AD, Anbuhl KL, Tollin DJ (2017) Representation of multidimensional stimuli: quantifying the most informative stimulus dimension from neural responses. J Neurosci 37:7332–7346.

Bernardi D, Lindner B (2015) A frequency-resolved mutual information rate and its application to neural systems. J Neurophysiol 113:1342–1357. CrossRef

Berry MJ, Meister M (1998) Refractoriness and neural precision. J Neurosci 18:2200–2211. Medline

Berry MJ, Brivanlou IH, Jordan TA, Meister M (1999) Anticipation of moving stimuli by the retina. Nature 398:334–338. CrossRef Medline

Bialek W, De Ruyter Van Steveninck RR, Tishby N (2006) Efficient representation as a design principle for neural coding and computation. Proc IEEE Int Symp Info Theory 659–663.

Borghuis BG, Ratliff CP, Smith RG, Sterling P, Balasubramanian V (2008) Design of a neuronal array. J Neurosci 28:3178–3189. CrossRef

Botella-Soler V, Deny S, Marre O, Tkačik G (2016) Nonlinear decoding of a complex movie from the mammalian retina. arXiv **q-bio**(1605.03373v1), [q–bio.NC].

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? J Neurosci 25:10577–10597. CrossRef Medline

Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I (2006) Reduction of information redundancy in the ascending auditory pathway. Neuron 51:359–368. CrossRef

Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16:3351–3362.

Deny S, Ferrari U, Macé E, Yger P, Caplette R, Picaud S, Tkačik G, Marre O (2017) Multiplexed computations in retinal ganglion cells of a single type Nature Communications 8, Article number: 1964 CrossRef Medline

Doi E, Gauthier JL, Field GD, Shlens J, Sher A, Greschner M, Machado T. a, Jepson LH, Mathieson K, Gunning DE, Litke AM, Paninski L, Chichilnisky EJ, Simoncelli EP (2012) Efficient coding of spatial information in the primate retina. J Neurosci 32:16256–16264. CrossRef

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Network 6:345–358. CrossRef

Faes L, Nollo G, Ravelli F, Ricci L, Vescovi M, Turatto M, Pavani F, Antolini R (2007) Small-sample characterization of stochastic approximation staircases in forced-choice adaptive threshold estimation. Percept Psychophys 69:254–262. CrossRef

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7:179–188. CrossRef

Gilja V, Nuyujukian P, Chestek CA, Cunningham JP, Yu BM, Fan JM, Churchland MM, Kaufman MT, Kao JC, Ryu SI, Shenoy KV (2012) A high-performance neural prosthesis enabled by control algorithm design. Nat Neurosci 15:1752–1757. CrossRef

Gollisch T, Meister M (2010) Eye smarter than scientists believed: neural computations in circuits of the retina. Neuron 65:150–164. CrossRef

Heitman A, Brackbill N, Greschner M, Sher A, Litke AM, Chichilnisky E (2016) Testing pseudo-linear models of responses to natural scenes in primate retina. bioRxiv 045336.

Karklin Y, Simoncelli EP (2011) Efficient coding of natural images with a population of noisy linear-nonlinear neurons. Adv Neural Inf Process Syst 24:999–1007. [CrossRef]

Keat J, Reinagel P, Reid RC, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. Neuron 30:803–817. CrossRef

Kesten H (1958) Accelerated stochastic approximation. Ann Math Stat 29:41–59. CrossRef

Kostal L, Lansky P, Rospars JP (2008) Efficient olfactory coding in the pheromone receptor neuron of a moth. PLoS Comput Biol 4:e1000053. CrossRef

Liu YS, Stevens CF, Sharpee T (2009) Predictable irregularities in retinal receptive fields. Proc Natl Acad Sci USA 106:16499–16504. CrossRef

Machens CK, Stemmler MB, Prinz P, Krahe R, Ronacher B, Herz AV (2001) Representation of acoustic communication signals by insect auditory receptor neurons. J Neurosci 21:3215–3227.
Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. J Neurosci 24:1089–1100. CrossRef
Macmillan N, Creelman C (2004) Detection theory: a user's guide. London: Taylor & Francis.
Marre O, Amodei D, Deshmukh N, Sadeghi K, Soo F, Holy TE, Berry MJ (2012) Mapping a complete neural population in the retina. Journal of Neuroscience, 32(43), 14859–14873.
Marre O, Botella-Soler V, Simmons KD, Mora T, Tkačik G, Berry II MJ (2015) High accuracy decoding of dynamical motion from a large retinal population. PLoS computational biology, 11(7), e1004304. CrossRef Medline
Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.
Olveczky BP, Baccus SA, Meister M (2003) Segregation of object and background motion in the retina. Nature 423:401–408. CrossRef Medline
Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454:995–999. CrossRef
Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1996) Spikes: exploring the neural code. Cambridge: MIT Press.
Ruderman DL, Bialek W (1992) Seeing beyond the Nyquist limit. Neural Comput 4:682–690. CrossRef
Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci USA 90:10749–10753. Medline
Smith EC, Lewicki MS (2006) Efficient auditory coding. Nature 439:978–982. CrossRef
Van Hateren J (1992) A theory of maximizing sensory information. Biol Cybern 68:23–29. CrossRef
Yger Pierre, Giulia LB, Spampinato Elric, Esposito, Baptiste, Lefebvre Stephane, Deny Christophe, Gardella Marcel, Stimberg et al. Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. bioRxiv (2016): 067843.
Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. J Neurophysiol 78:2336–2350. CrossRef Medline
Wei X-X, Stocker AA (2016) Mutual information, Fisher information, and efficient coding. Neural Comput 28:305–326. CrossRef
Woyczynski WA (2010) A first course in statistics for signal analysis. New York: Springer Science & Business Media.