
Research Article: New Research | Sensory and Motor Systems

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

Ulisse Ferrari¹, Christophe Gardella^{1,2}, Olivier Marre¹ and Thierry Mora²

¹*INSERM and UMPC, Institut De La Vision, 17 Rue Moreau, Paris 75012, France*

²*Laboratoire De Physique Statistique, CNRS, UPMC, UPD, and Ecole Normale Supérieure, (PSL University) 24, Rue Lhomond, Paris 75005, France*

DOI: 10.1523/ENEURO.0166-17.2017

Received: 12 May 2017

Revised: 12 October 2017

Accepted: 16 October 2017

Published: 16 January 2018

Author Contributions: OM and TM designed research; UF and CG performed research; UF, CG, OM and TM analyzed the data; UF, CG, OM and TM wrote the paper.

Funding: <http://doi.org/10.13039/100000002HHS> | National Institutes of Health (NIH) U01NS090501

Funding: <http://doi.org/10.13039/501100000781EC> | European Research Council (ERC) 720270

Funding: <http://doi.org/10.13039/501100001665> Agence Nationale de la Recherche (ANR) OPTIMA
TRAJECTORY
ANR-10-LABX-65
ANR-17-ERC2-0025-01

Conflict of Interest: Authors report no conflict of interest.

This work was supported by ANR TRAJECTORY, ANR OPTIMA, ANR IRREVERSIBLE, 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 National Institutes of Health grant n. U01NS090501.

U.F. and C.G. contributed equally to this work.

O.M. and T.M. contributed equally to this work.

Correspondence should be addressed to either Olivier Marre, olivier.marre@gmail.com or Thierry Mora, tmora@ips.ens.fr

Cite as: eNeuro 2018; 10.1523/ENEURO.0166-17.2017

Alerts: Sign up at eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2018 Ferrari et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1. Manuscript Title:

Closed-loop estimation of retinal network sensitivity by local empirical linearization

2. Abbreviated Title:

Closed-loop estimation of retinal network sensitivity by local empirical linearization

3. Authors:

Ulisse Ferrari, 1, * ;
Christophe Gardella, 2, 1, * ;
Olivier Marre, 1, † ;
Thierry Mora 2, †

Affiliations:

1) Institut de la Vision, INSERM and UMPC, 17 rue Moreau, 75012 Paris, France
2) Laboratoire de physique statistique, CNRS, UPMC, UPD, and Ecole normale supérieure (PSL University), 24, rue Lhomond, 75005 Paris, France
) These authors contributed equally.
†) These authors contributed equally.

4. Author Contributions:

OM and TM designed research;
UF and CG performed research;
UF, CG, OM and TM analyzed the data;
UF, CG, OM and TM wrote the paper.

5. Correspondence should be addressed to

olivier.marre@gmail.com
tmora@lps.ens.fr

6. Number of Figures: 7

7. Number of Tables: 0

8. Number of Multimedia: 0

9. Number of words for Abstract: 153

10. Number of words for Significance Statement: 120

11. Number of words for Introduction: 500

12. Number of words for Discussion: 504

13. Acknowledgements

We thank Stéphane Deny for his help with the experiments, and Jean-Pierre Nadal for

stimulating discussions and crucial suggestions.

14. Conflict of Interest:

‘Authors report no conflict of interest’

15. Funding sources:

This work was supported by ANR TRAJECTORY, ANR OPTIMA, ANR IRREVERSIBLE, 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 National Institutes of Health grant n. U01NS090501.

1 **Closed-loop estimation of retinal network sensitivity by local**
2 **empirical linearization**

3 **Abstract**

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

17 SIGNIFICANCE STATEMENT

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

28

29 INTRODUCTION

30 An important issue in neuroscience is to understand how sensory systems use their neural
31 resources to represent information. A crucial step towards understanding the sensory pro-
32 cessing performed by a given brain area is to characterize its sensitivity (Benichoux *et al.*
33 2017), by determining which features of the sensory input are coded in the activity of these
34 sensory neurons, and which features are discarded. If a sensory area extracts a given feature
35 from the sensory scene, any change along that dimension will trigger a noticeable change in
36 the activity of the sensory system. Conversely, if the information about a given feature is
37 discarded by this area, the activity of the area should be left invariant by a change along
38 that feature dimension. To understand which information is extracted by a sensory network,
39 we must determine which changes in the stimulus evoke a significant change in the neural
40 response, and which ones leave the response invariant.

41 This task is made difficult by the fact that sensory structures process stimuli in a highly
42 non-linear fashion. At the cortical level, many studies have shown that the response of
43 sensory neurons is shaped by multiple non-linearities (Carandini *et al.* 2005, Machens *et al.*
44 2004). Models based on the linear receptive field are not able to predict the responses of
45 neurons to complex, natural scenes. This is even true in the retina. While spatially uniform

46 or coarse grained stimuli produce responses that can be predicted by quasi-linear models
47 (Berry and Meister 1998, Keat *et al.* 2001, Pillow *et al.* 2008), stimuli closer to natural scenes
48 (Heitman *et al.* 2016) or with rich temporal dynamics (Berry *et al.* 1999, Ölveczky *et al.*
49 2003) are harder to characterize, as they trigger non-linear responses in the retinal output.
50 These unknown non-linearities challenge our ability to model stimulus processing and limit
51 our understanding of how neural networks process information.

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

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

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

73 We then used this experimental estimation of the network sensitivity to tackle two long-
74 standing issues in sensory neuroscience. First, when trying to decode neural activity to
75 predict the stimulus presented, it is always difficult to know if the decoder is optimal or if
76 it misses some of the available information. We show that our estimation of the network
77 sensitivity gives an upper bound of the decoder performance that should be reachable by

78 an optimal decoder. Second, the efficient coding hypothesis (Attneave 1954, Barlow 1961)
79 postulates that neural encoding of stimuli has adapted to represent natural occurring sensory
80 scenes optimally in the presence of limited resources. Testing this hypothesis for sensory
81 structures that perform non-linear computations on high dimensional stimuli is still an open
82 challenge. Here we found that the network sensitivity with respect to stimulus perturbations
83 exhibits a peak as a function of the temporal frequency of the perturbation, in agreement
84 with prediction from efficient coding theory. Our method paves the way towards testing
85 efficient coding theory in non-linear networks.

86 MATERIALS AND METHODS

87 **Extracellular recording.** Experiments were performed on the adult Long Evans rat of either sex,
88 in accordance with institutional animal care standards. The retina was extracted from the euthanized animal
89 and maintained in an oxygenated Ames' medium (Sigma-Aldrich). The retina was recorded extracellularly
90 on the ganglion cell side with an array of 252 electrodes spaced by 60 μm (Multichannel Systems), as
91 previously described (Anonymous 2012). Single cells were isolated offline using Anonymous a custom spike
92 sorting algorithm (Anonymous 2016). We then selected 60 cells that were well separated (no violations of
93 refractory period, *i.e.* no spikes separated by less than 2 ms), had enough spikes (firing rate larger than 0.5
94 Hz), had a stable firing rate during the whole experiment, and responded consistently to repetitions of a
95 reference stimulus (see later).

96 **Stimulus.** The stimulus was a movie of a white bar on a dark background projected at a refresh rate
97 of 50 Hz with a digital micromirror device. The bar had intensity $7.6 \cdot 10^{11}$ photons. $\text{cm}^{-2}.\text{s}^{-1}$, and 115 μm
98 width. The bar was horizontal and moved vertically. The bar trajectory consisted in 17034 snippets of 0.9
99 s consisting in 2 reference trajectories repeated 391 times each, perturbations of these reference trajectories
100 and 6431 random trajectories. Continuity between snippets was ensured by constraining all snippets to
101 start and end in the middle of the screen with velocity 0. Random trajectories followed the statistics of an
102 overdamped stochastic oscillator (Anonymous 2015). We used a Metropolis-Hastings algorithm to generate
103 random trajectories satisfying the boundary conditions. The two reference trajectories were drawn from
104 that ensemble.

105 **Perturbations.** Stimulus perturbations were small changes in the middle portion of the reference
106 trajectory, between 280 and 600 ms. A perturbation is denoted by its discretized time series with time

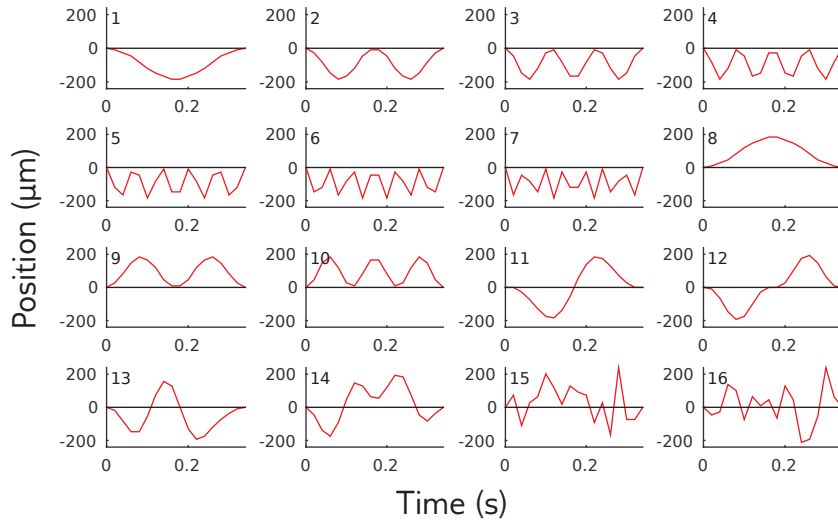


FIG. 1. **Perturbations shapes.** We used the same 16 perturbation shapes for the 2 reference stimuli. The first 12 perturbation shapes were combinations of two Fourier components, and the last 4 ones were random combinations of them: $f_k(t) = \cos(2\pi kt/T)$, $g_k(t) = (1/k) \sin(2\pi kt/T)$, with 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 $j = 8, \dots, 10$ 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.

107 step $\delta t = 20$ ms, $\mathbf{S} = (S_1, \dots, S_L)$, with $L = 16$, over the 320 ms of the perturbation (bold symbols
 108 represent vectors and matrices throughout). Perturbations can be decomposed as $\mathbf{S} = A \times \mathbf{Q}$, where
 109 $A^2 = (1/L) \sum_{t=1}^L S_t^2$ is the amplitude, and $\mathbf{Q} = \mathbf{S}/A$ the shape. Perturbations shapes were chosen to have
 110 zero value and zero derivative at their boundaries. They are represented in Fig. 1.

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

117 was recorded and used to select the next stimulation in real time.

118 **Online spike detection.** During the experiment we detected spikes in real time on each electrode
 119 independently. Each electrode signal was high-pass filtered using a Butterworth filter with a 200 Hz frequency
 120 cutoff. A spike was detected if the electrode potential U was lower than a threshold of 5 times the median
 121 absolute deviation of the voltage (Anonymous 2016).

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

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

128 where $C = 0.74$ is a scaling coefficient that controls the size of steps, and r_n is the number of reversal steps
 129 in the experiment, *i.e.* the number of times when a discrimination D_n larger than D^* was followed by
 130 D_{n+1} smaller than D^* , and vice versa. In order to explore the responses to different ranges of amplitudes
 131 even in the case where the algorithm converged too fast, we also presented amplitudes regularly spaced on
 132 a log-scale. We presented the largest amplitude A_{\max} (value in caption of Fig. 1), and scaled it down by
 133 multiples of 1.4, $A_{\max}/1.4^k$ with $k = 1, \dots, 7$.

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

143 In order to apply linear discrimination on $\mathbf{R}_{\mathbf{S}}$, the response to the perturbation \mathbf{S} , we record multiple
 144 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

145 stimulus shape as \mathbf{S} but at the maximum amplitude that was played during the course of the experiment
 146 (typically 110 μm , see caption Fig. 1). Our goal is to estimate how close $\mathbf{R}_{\mathbf{S}}$ is to the ‘typical’ \mathbf{R}_{ref}
 147 compared to the ‘typical’ $\mathbf{R}_{\mathbf{S}_{\text{max}}}$. To this aim, we compute the mean response to the reference and to the
 148 large perturbation, $\langle \mathbf{R}_{\text{ref}} \rangle$ and $\langle \mathbf{R}_{\mathbf{S}_{\text{max}}} \rangle$, and use their difference as a linear classifier. Specifically we project
 149 $\mathbf{R}_{\mathbf{S}}$ onto the difference between these two mean responses. For a generic response \mathbf{R} (either \mathbf{R}_{ref} , $\mathbf{R}_{\mathbf{S}}$ or
 150 $\mathbf{R}_{\mathbf{S}_{\text{max}}}$), the projection x (respectively, x_{ref} , $x_{\mathbf{S}}$ or $x_{\mathbf{S}_{\text{max}}}$) reads:

$$x = \mathbf{u}^T \cdot \mathbf{R} \quad (2)$$

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

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

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

162 **Offline discrimination and sensitivity.** To measure the discrimination probability as a
 163 function of the perturbation amplitude, we consider the difference of the projections, $\Delta x = x_{\mathbf{S}} - x_{\text{ref}}$. The
 164 response to the stimulation $\mathbf{R}_{\mathbf{S}}$ is noisy, making x and x_{ref} the sum of many random variables (corresponding
 165 to each neuron and time bin combinations), and we can apply the central limit theorem to approximate
 166 their distributions as Gaussian (as verified in the right side in Fig. 3B), for a given perturbation at a given
 167 amplitude. For small perturbations, the mean of Δx grows linearly with the perturbation amplitude A ,
 168 $\mu = \alpha \times A$, and the variances of $x_{\mathbf{S}}$ and x_{ref} are equal at first order, $\text{Var}(x_{\mathbf{S}}) \approx \text{Var}(x_{\text{ref}}) = \sigma^2$, so that
 169 the variance of Δx , $\text{Var}(\Delta x) = \text{Var}(x_{\mathbf{S}}) + \text{Var}(x_{\text{ref}}) = 2\sigma^2$ is independent of A . Then the probability of
 170 discrimination is given by the error function:

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

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

174 **Optimal sensitivity and Fisher information.** We then aimed to find the discrimination
 175 probability for any perturbation. Given the distributions of responses to the reference stimulus, $P(\mathbf{R}|\text{ref})$,
 176 and to a perturbation, $P(\mathbf{R}|\mathbf{S})$, optimal discrimination can be achieved by studying the sign of the response-
 177 specific log-ratio $\mathcal{L}(\mathbf{R}) = \ln[P(\mathbf{R}|\mathbf{S})/P(\mathbf{R}|\text{ref})]$. Note that in the log-ratio, \mathbf{R} represents a stochastic response
 178 and not the independent variable of a probability density. Because it depends on the response \mathbf{R} , this log
 179 ratio is both stimulus dependent and stochastic. Let us define \mathcal{L}_{ref} to be the random variable taking value
 180 $\mathcal{L}(\mathbf{R})$ upon presentation of the reference stimulus, i.e. when \mathbf{R} is a (stochastic) response to the stimulus,
 181 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
 182 the definition given earlier, the probability of successful discrimination is the probability that the log-ratio
 183 calculated from a random response to the perturbed stimulus is larger than the log-ratio calculated from a
 184 random response to the reference, $\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{\text{ref}}$. Using the central limit theorem we assume again that $\mathcal{L}_{\mathbf{S}}$ and
 185 \mathcal{L}_{ref} are Gaussian. We can calculate their mean and variance at small \mathbf{S} (see Mathematical derivations):
 186 $\mu_{\mathcal{L}} = \langle \mathcal{L}_{\mathbf{S}} \rangle - \langle \mathcal{L}_{\text{ref}} \rangle = \mathbf{S}^T \cdot \mathbf{I} \cdot \mathbf{S}$ and $2\sigma_{\mathcal{L}}^2 = \text{Var}(\mathcal{L}_{\mathbf{S}}) + \text{Var}(\mathcal{L}_{\text{ref}}) = 2\mathbf{S}^T \cdot \mathbf{I} \cdot \mathbf{S}$, where

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

187 is the Fisher information matrix calculated at the reference stimulus. Following standard discrimination
 188 theory (see Macmillan and Creelman (2004), and Seung and Sompolinsky (1993) for a derivation in a
 189 similar context), the discrimination probability is (see Mathematical derivations): $D = P(\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{\text{ref}}) =$
 190 $(1/2)[1 + \text{erf}(d'/2)]$, with

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

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

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

198 of the logarithm of response distribution as a function of both stimulus and response:

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

199 where in the integral form, $\{t_i\}$ denotes the set of spiking times of neuron i , and F_i is a stimulus filter
 200 depending on both the stimulus time and spiking time (no time-translation invariance). The second line
 201 is the discretized version adapted to our binary convention for describing spiking activity binned into bins
 202 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
 203 which the change in the response is calculated from the linear projection of the past stimulus. For ease of
 204 notation, hereafter we use matrix multiplications rather than explicit sums over ib and t .

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

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

206 Since the variables R_{ib} are binary, their mean values $\langle R_{ib} \rangle$ upon presentation of the reference completely
 207 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
 208 of the reference stimulus, with a small pseudo-count to avoid zero values.

209 Evaluating the Fisher information matrix, Eq. (18), within the local model, Eq. 6, gives:

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

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

212 **Inference of the local model.** To infer the filters $F_{ib,t}$, we only include perturbations that are
 213 small enough to remain within the linear approximation. We first separated the dataset into a training ($285 \times$
 214 16 perturbations) and testing (20×16 perturbations) sets. We then defined, for each perturbation shape, a
 215 maximum perturbation amplitude above which the linear approximation was no longer considered valid. We
 216 selected this threshold by optimizing the model's ability to predict the changes in firing rates in the testing
 217 set. Model learning was performed for each cell independently by maximum likelihood with an L_2 smoothness
 218 regularization on the shape of the filters, using a pseudo-Newton algorithm. The amplitude threshold
 219 obtained from the optimization varied widely across perturbation shapes. The number of perturbations for

220 each shape used in the inference ranged from 20 (7% of the total) to 260 (91% of the total). Overall only
 221 32% of the perturbations were kept (as we excluded repetitions of perturbations with largest amplitude used
 222 for calibration). Overfitting was limited: when tested on perturbations of similar amplitudes, the prediction
 223 performance on the testing set was never lower than 15% of the performance on the training set.

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

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

227 where the filters K have a time integration windows of $15 \times 20 \text{ ms} = 300 \text{ ms}$, as in the local model.

228 We inferred the linear decoder filters by minimizing the mean square error (Warland *et al.* 1997) ,
 229 $\sum_t [X(t) - \hat{X}(t)]^2$, in the reconstruction of 4000 random trajectories governed by the dynamics of an over-
 230 damped oscillator with noise (see above). The linear decoder is then applied to the perturbed trajectories,
 231 $X(t) = X_0(t) + S(t)$, where $X_0(t)$ denotes the reference trajectory. The linear decoder does not use prior
 232 information about the local structure of the experiment, namely about the fact that the stimulus to decode
 233 consists of perturbations around a reference simulation. However, it implicitly uses prior information about
 234 the statistics of the overdamped oscillator, as it was trained on bar trajectories with those statistics. Tested
 235 on a sequence of ~ 400 repetitions of one of the two reference trajectories, where the first 300 ms of each
 236 have been cut out, we obtain a correlation coefficient of 0.87 between the stimulus and its reconstruction.

237 **Local model Bayesian decoder.** In order to construct a decoder based on the local model, we
 238 use Bayes' rule to infer the presented stimulus given the response:

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

239 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
 240 $P(\mathbf{R})$ is a normalization factor that does not depend on the stimulus. $P(\mathbf{S})$ is taken to be the distribution
 241 of trajectories from an overdamped stochastic oscillator with the same parameters as in the experiment
 242 (Anonymous 2015), to allow for a fair comparison with the linear decoder, which was trained with those
 243 statistics. The stimulus is inferred by maximizing the posterior $P(\mathbf{S}|\mathbf{R})$ numerically, using a pseudo-Newton
 244 iterative algorithm.

245 **Local signal to noise ratio in decoding.** To quantify local decoder performance as a function
 246 of the stimulus frequency, we estimated a local signal-to-noise ratio of the decoding signal, $\text{LSNR}(\mathbf{S})$, which

247 is a function of the reference stimulus. Here we cannot compute SNR as a ratio between total signal power
 248 and noise power, because this would require to integrate over the entire stimulus space, while our approach
 249 only provides a model around the neighbourhood of the reference stimulus.

250 In order to obtain a meaningful comparison between the linear and local decoders, we expand them
 251 at first order in the stimulus perturbation and compute the SNR of these ‘linearized’ decoders. For any
 252 decoder and for stimuli nearby a reference stimulation, the inferred value of the stimulus, $\hat{\mathbf{X}}$, can be written
 253 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
 254 of the response on which the reconstruction relies). Linearizing ϕ for $\mathbf{X} = \mathbf{X}_0 + \mathbf{S}$,

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

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

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

257 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}_ϵ from the ~ 400
 258 reconstructions of the reference stimulation using either of the two decoders, and the values of \mathbf{T} from the
 259 reconstructions of the perturbed trajectories. The inference is done by an iterative algorithm similar to that
 260 used for the inference of the filters \mathbf{F} of the local model. We define the local signal-to-noise ratio (LSNR) in
 261 decoding the perturbation \mathbf{S} as:

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

262 where here $\langle \dots \rangle$ means average with respect to the noise ϵ . In this formula, the signal is defined as the
 263 average predicted perturbation $\langle \hat{\mathbf{S}} \rangle$, from which the systematic bias \mathbf{b} is subtracted, yielding $\mathbf{T} \cdot \mathbf{S}$. The noise
 264 is simply ϵ . Note that here the LSNR is defined for a given perturbation \mathbf{S} . It is the ratio of the squared
 265 signal to the noise variance (summed over the eigendirections of the noise correlator, since we are dealing with
 266 a multidimensional signal). This LSNR gives a measure of decoding performance, through the amplitude
 267 of the decoded signal relative to the noise. To study how this performance depends on the frequency ν of
 268 the input signal, in Fig. 6C we apply Eq. 13 with $S_b = A \exp(2\pi i \nu b \delta t)$, where A is the amplitude of the
 269 perturbation shown in Fig. 5A and b is a time-bin counter. Note that this frequency-dependent LSNR should
 270 not be interpreted as a ratio of signal and noise power spectra, but rather as the dependence of decoding

271 performance on the frequency of the perturbation. It is used rather than the traditional SNR because we
272 are dealing with signals with no time-translation invariance (i.e. $T_{tt'}$ is not just a function of $t - t'$, and
273 neither is $C_{\epsilon, tt'}$). However, our LNSR reduces to the traditional frequency-dependent SNR in the special
274 case of time-translation invariance, i.e. when the decoder is convolutional, and its noise stationary (see
275 Mathematical Derivations)

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

281 RESULTS

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

294 To assess the sensitivity of the retinal network, we asked how well different perturbations could be
295 discriminated from the reference stimulus based on the population response. We expect the ability to
296 discriminate perturbations to depend on two factors. First, the direction of the perturbation in the stimulus
297 space, called perturbation shape. If we change the reference stimulus by moving along a dimension that is not
298 taken into account by the recorded neurons, we should not see any change in the response. Conversely, if we
299 choose to change the stimulus along a dimension that neurons “care about,” we should quickly see a change

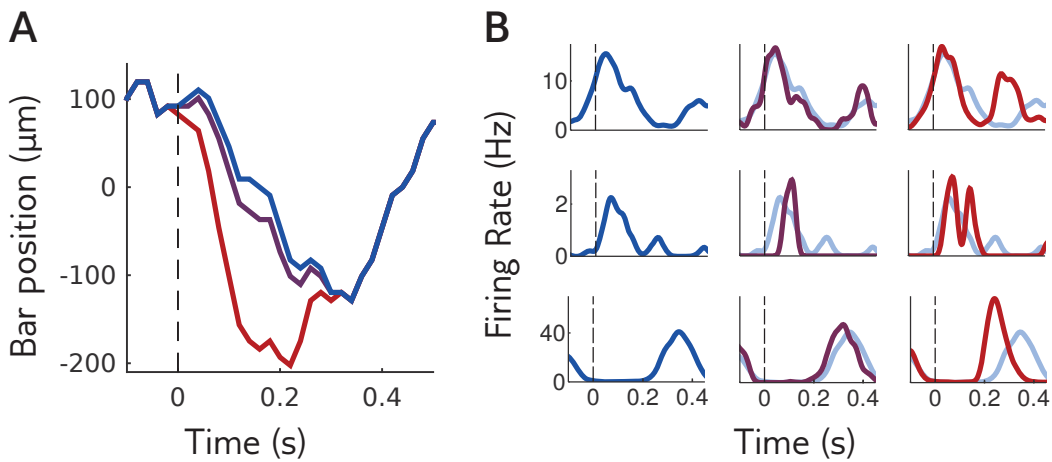


FIG. 2. **Sensitivity of a neural population to visual stimuli.** **A.:** the retina is stimulated with repetitions of a reference stimulus (here the trajectory of a bar, in blue), and with perturbations of this reference stimulus of different shapes and amplitudes. Purple and red trajectories are perturbations 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).

300 in the response. The second factor is the amplitude of the perturbation: responses to small perturbations
 301 should be hardly distinguishable, while large perturbations should elicit easily detectable changes, as can
 302 be seen in Fig. 2B. To assess the sensitivity to perturbations of the reference stimulus we need to explore
 303 many possible directions that these perturbations can take, and for each direction, we need to find a range
 304 of amplitudes that is as small as possible but will still evoke a detectable change in the retinal response. In
 305 other words, we need to find the range of amplitudes for which discrimination is hard but not impossible.
 306 This requires looking for the adequate range of perturbation amplitudes “online,” during the time course of
 307 the experiment.

308 In order to automatically adapt the amplitude of perturbations to the sensitivity of responses for each
 309 of the 16 perturbation shapes and for each reference stimulus, we implemented closed-loop experiments
 310 (Fig. 3A). At each step, the retina was stimulated with a perturbed stimulus and the population response
 311 was recorded. Spikes were detected in real time for each electrode independently by threshold crossing (see
 312 Materials and Methods). This coarse characterization of the response is no substitute for spike sorting,
 313 but it is fast enough to be implemented in real time between two stimulus presentations, and sufficient to
 314 detect changes in the response. This method was used to adaptively select the range of perturbations in
 315 real time during the experiment, and to do it for each direction of the stimulus space independently. Proper

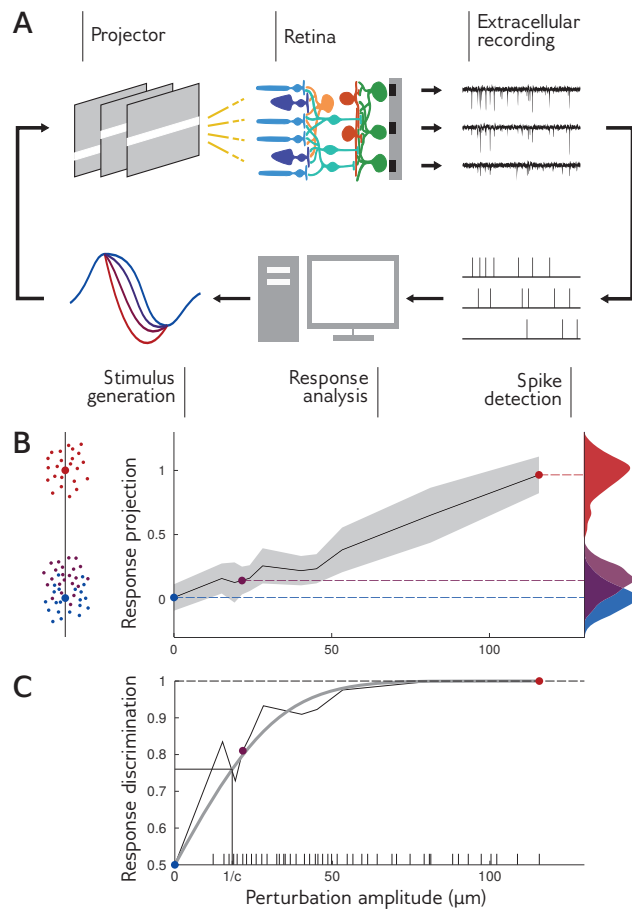


FIG. 3. Closed-loop experiments to probe the range of stimulus sensitivity. **A.** Experimental setup: we stimulated a rat retina with a moving bar. Retinal ganglion cell (RGC) population responses were recorded extracellularly with a multi-electrode array. Electrode signals were high-pass filtered and 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 grey) 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, $(1/2)[1 + \text{erf}(d'/2)]$, with $d' = c \times A$ (grey line: fit). Ticks on the x axis show the amplitudes that have been tested during the closed-loop experiment.

316 spike sorting was performed after the experiment using the procedure described in Anonymous (2012) and
317 Anonymous (2016) and used for all subsequent analyses.

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

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

335 Fig. 3C shows the discrimination probability as a function of the perturbation amplitude for an example
336 perturbation shape. Discrimination grows linearly with small perturbations, and then saturates to 100%
337 for large ones. This behavior is well approximated by an error function (gray line) parametrized by a
338 single coefficient, which we call sensitivity coefficient and denote by c . This coefficient measures how fast
339 the discrimination probability increases with perturbation amplitude: the higher the sensitivity coefficient,
340 the easier it is to discriminate responses to small perturbations. It can be interpreted as the inverse of the
341 amplitude at which discrimination reaches 76%, and is related to the classical sensitivity index d' (Macmillan
342 and Creelman 2004), through $d' = c \times A$, where A denotes the perturbation amplitude (see Materials and
343 Methods).

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

347 bation shapes, with a minimum of $c = 0.028 \mu m^{-1}$ and a maximum of $c = 0.065 \mu m^{-1}$.

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

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

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

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

363 **Local model for predicting sensitivity.** We introduce a local model to describe the stochas-
364 tic population response to small perturbations of the reference stimulus. This model will then be used to
365 estimate the Fisher information matrix, and from it the retina's sensitivity to any perturbation, using Eq. 14.

366 The model, schematized in Fig. 4A, assumes that perturbations are small enough that the response can be
367 linearized around the reference stimulus. First, the response to the reference is described by conditionally
368 independent neurons firing with time-dependent rates estimated from the peristimulus time histograms
369 (PSTH). Second, the response to perturbations is modeled as follows: for each neuron and for each 20
370 ms time bin of the considered response, we use a linear projection of the perturbation trajectory onto a
371 temporal filter to modify the spike rates relative to the reference. These temporal filters were inferred from
372 the responses to all the presented perturbations, varying both in shape and amplitude (but small enough to
373 remain within the linear approximation). Details of the model and its inference are given in Materials and
374 Methods.

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

387 To further validate the local model, we combine it with Eq. 14 to predict the sensitivity c of the network
388 to various perturbations of the bar trajectory, as measured directly by linear discrimination (Fig. 3). The
389 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
390 model's temporal filters (stacked as row vectors), and \mathbf{C}_R is the covariance matrix of the entire response to
391 the reference stimulus across neurons and time. We can then use the Fisher matrix to predict the sensitivity
392 coefficient using Eq. 14, and compare it to the same sensitivity coefficient previously estimated using linear
393 discrimination. Fig. 5A shows that these two quantity are strongly correlated (Pearson correlation: 0.82,
394 $p = 10^{-8}$), although the Fisher prediction is always larger. This difference could be due to two reasons:
395 limited sampling of the responses, or non optimality of the projection axis used for linear discrimination. To
396 evaluate the effect of finite sampling, we repeated the analysis on a synthetic dataset generated using the
397 local model, with the same stimulation protocol as in the actual experiment. The difference in the synthetic
398 data (Fig. 5B) and experiment (Fig. 5A) were consistent, suggesting that finite sampling is indeed the main
399 source of discrepancy. We confirmed this result by checking that using the optimal discrimination axis (see
400 Mathematical Derivations) did not improve performance (data not shown).

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

405 **Linear decoding is not optimal.** When trying to decode the position of random bar trajecto-

406 ries over time using the retinal activity, we found that a linear decoder (Materials and Methods) could reach
407 a satisfying performance, confirming previous results (Warland, Anonymous). Several works have shown
408 that it was challenging to outperform linear decoding on this task in the retina (Warland, Anonymous).
409 From this result we can wonder if the linear decoder is optimal, i.e. makes use of all the information present
410 in the retinal activity, or if this decoder is sub-optimal and could be outperformed by a non-linear decoder.
411 To answer this question, we need to determine an upper bound on the decoding performance reachable by
412 any decoding method. For an encoding model, the lack of reliability of the response sets an upper bound
413 on the encoding model performance, but finding a similar upper bound for decoding is an open challenge.
414 Here we show that our local model can define such an upper bound.

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

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

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

447 To interpret this peak in sensitivity, we studied a minimal theory of retinal function, similar to
448 Van Hateren (1992), to test how maximizing information transmission would reflect on the sensitivity
449 of the retinal response. In this theory, the stimulus is first passed through a low-pass filter, then corrupted
450 by an input white noise. This first stage describes filtering due to the photoreceptors (Ruderman and
451 Bialek 1992). The photoreceptor output is then transformed by a transfer function and corrupted by a
452 second external white noise, which mimics the subsequent stages of retinal processing leading to ganglion
453 cell activity. Here the output is reduced to a single continuous signal (Fig. 7B, see Mathematical Deriva-
454 tions details). Note that this theory is linear: we are not describing the response of the retina to any
455 stimulus, which would be highly non-linear, but rather its linearized response to perturbations around a
456 given stimulus, as in our experimental approach. To apply the efficient coding hypothesis, we assumed that
457 the photoreceptor filter is fixed, and we maximized the transmitted information, measured by Shannon's
458 mutual information, over the transfer function, see Mathematical Derivations, Eq. (31). We constrained
459 the variance of the output to be constant, corresponding to a metabolic constraint on the firing rate of
460 ganglion cells. In this simple and classical setting, this optimal transfer function, and the corresponding
461 sensitivity, can be calculated analytically. Although the power spectrum of the stimulus and photoreceptor
462 output are monotonically decreasing, and the noise spectrum is flat, we found that the optimal sensitivity
463 of the theory is bell shaped (Fig. 7E), in agreement with our experimental findings (Fig. 7C). Recall that
464 in our reasoning, we assumed that the network optimizes information transmission for the statistics of the
465 stimulus used in the experiment. Alternatively, it is possible that the retinal network optimizes information
466 transmission of natural stimuli, which may have slightly different statistics. We also tested our model with
467 natural temporal statistics (power spectrum $\sim 1/\nu^2$ as a function of frequency ν , Dong and Atick (1995))

468 and found the same results (data not shown).

469 One can intuitively understand our result that a bell-shaped sensitivity is desirable from a coding per-
470 spective. On one hand, in the small frequency regime, sensitivity increases with frequency, i.e. decreases
471 with stimulus power. This result is classic: when the input noise is small compared to stimulus, the best
472 coding strategy for maximizing information is to whiten the input signal to obtain a flat output spectrum,
473 which is obtained by having the squared sensitivity be inversely proportional to the stimulus power (Rieke
474 *et al.* 1996, Wei and Stocker 2016). On the other hand, at high frequencies, the input noise is too high
475 (relative to the stimulus power) for the stimulus to be recovered. Allocating sensitivity and output power
476 to those frequencies is therefore a waste of resources, as it is devoted to amplifying noise, and sensitivity
477 should remain low to maximize information. A peak of sensitivity is thus found between the high SNR
478 region, where stimulus dominates noise and whitening is the best strategy, and the low LSNR region, where
479 information is lost into the noise and coding resources should be scarce. A result of this optimization is that
480 the information transferred should monotonically decrease with frequency, just as the input power spectrum
481 does (Fig. 7F). We tested if this prediction was verified in the data. We estimated similarly the information
482 rate against frequency in our data, and found that it was also decreasing monotonically (Fig. 7D). The
483 retinal response has therefore organized its sensitivity across frequencies in a manner that is consistent with
484 an optimization of information transmission across the retinal network.

485 DISCUSSION

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

495 Our approach can be used to test how optimal different decoding methods are. In our case, we found that
496 linear decoding, despite its very good performance, was far from the performance of the Bayesian inversion

497 of our local model, and therefore far from optimal. This result implies that there should exist non-linear
 498 decoding methods that outperform linear decoding (Botella-Soler *et al.* 2016). Testing the optimality of the
 499 decoding method is crucial for brain machine interfaces (Gilja *et al.* 2012): in this case an optimal decoder is
 500 necessary to avoid missing a significant amount of information. Building our local model is a good strategy
 501 for benchmarking different decoding methods.

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

510 More generally, different versions of the efficient coding theory have been proposed to explain the orga-
 511 nization of several areas of the visual system (Bell and Sejnowski 1997, Bialek *et al.* 2006, Dan *et al.* 1996,
 512 Karklin and Simoncelli 2011, Olshausen and Field 1996) and elsewhere (Chechik *et al.* 2006, Kostal *et al.*
 513 2008, Machens *et al.* 2001, Smith and Lewicki 2006). Estimating Fisher information using a local model
 514 could be used in other sensory structures to test the validity of these hypotheses.

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

521 MATHEMATICAL DERIVATIONS

522 A. Derivation of discrimination coefficient in arbitrary dimension

523 Here we derive Eq. 5 in detail. Recall that \mathcal{L}_{ref} is a random variable taking value $\mathcal{L}(\mathbf{R}) = \ln[P(\mathbf{R}|\mathbf{S})/P(\mathbf{R}|\text{ref})]$
 524 upon presentation of the reference stimulus and $\mathcal{L}_{\mathbf{S}}$ the random variable taking value $\mathcal{L}(\mathbf{R})$ when \mathbf{R} is a

525 response to the presentation of \mathbf{S} . Then their averages are given by:

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

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

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

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

527 with

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

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

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

529 where we have used the fact that

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

530 Next we assume that $\ln P(\mathbf{R}|\mathbf{S})$ is the sum of weakly correlated variables, meaning that its distribution can

531 be approximated as Gaussian. Thus, the random variable $\mathcal{L}_{\mathbf{S}} - \mathcal{L}_{\text{ref}}$ is also distributed as a Gaussian, with

532 mean $\mu_{\mathcal{L}} = \mathbf{S}^T \cdot \mathbf{I} \cdot \mathbf{S}$ and variance $\sigma_{\mathcal{L}}^2 = 2\mathbf{S}^T \cdot \mathbf{I} \cdot \mathbf{S}$. The discrimination probability is the probability that

533 $\mathcal{L}_{\mathbf{S}} > \mathcal{L}_{\text{ref}}$, i.e.

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

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

535 **B. Fisher and linear discrimination.**

536 There exists a mathematical relation between the Fisher information of Eq. 8 and linear discrimination.
 537 The linear discrimination task described earlier can be generalized by projecting the response difference,
 538 $\mathbf{R}_S - \mathbf{R}_{\text{ref}}$, along an arbitrary direction \mathbf{u} :

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

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

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

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

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

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

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

551 Note that the optimal direction of projection only differs from the direction we used in the experiments,
 552 $\mathbf{u} = \langle \mathbf{R}_S \rangle - \langle \mathbf{R}_{\text{ref}} \rangle$, by an equalization factor \mathbf{C}_R^{-1} . We have checked that applying that factor only improves
 553 discrimination by a few percents (data not shown).

554 **C. Local SNR for a convolutional linear decoder**

555 In this section we show how the local SNR defined in Eq. (13) reduces to standard expression in the
556 simpler case of a convolution decoder ϕ in the linear regime:

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

557 where \star is the convolution symbol, h is a stimulus independent linear filter and $\boldsymbol{\epsilon}$ a Gaussian noise of
558 covariance \mathbf{C}_ϵ and zero mean. Linearizing ϕ for $\mathbf{X} = \mathbf{X}_0 + \mathbf{S}$ as in Eq. (12), we obtain

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

559 but now the transfer matrix $T_{bb'} = h(b - b')$ depends only on the difference between the time-bin indices
560 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
561 have:

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

562 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
563 property, the LSNR takes the following expression (see Eq. (13)):

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

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

564 where $|\tilde{h}(\nu)|^2$ can be interpreted as the signal power at frequency ν for unitary stimulus perturbation. If
565 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
566 (Woyczynski 2010):

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

567 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 .

568 **D. Frequency dependence of sensitivity and information.**

569 To analyze the behavior in frequency of the sensitivity, we compute the sensitivity index for an oscillating
570 perturbation of unitary amplitude. We apply Eq. 14 with $\hat{S}_b(\nu) \equiv \exp(2\pi i\nu b\delta t)$. In order to estimate the

571 spectrum of the information rate we compute its behavior within the linear theory (Van Hateren 1992):

$$\text{MI}(\nu) = \frac{1}{2} \ln [1 + C_S(\nu)I(\nu)/\delta t^2] \quad (31)$$

572 where $C_S(\nu)$ is the power spectrum of the actual stimulus statistics (noisy damped oscillator), and $I(\nu) =$
 573 $(\delta t/L)\hat{\mathbf{S}}^T(\nu) \cdot \mathbf{I} \cdot \hat{\mathbf{S}}(\nu)$. Note that this decomposition in frequency of the transmitted information is valid
 574 because the system is linear and the stimulus is Gaussian distributed (Bernardi and Lindner 2015).

575 E. Efficient coding theory.

576 To build a theory of retinal sensitivity, we follow closely the approach of Van Hateren (Van Hateren
 577 1992). The stimulus is first linearly convolved with a filter f , of power \mathcal{F} , then corrupted by an input white
 578 noise with uniform power H , then convolved with the linear filter r of the retina network of power \mathcal{G} , and
 579 finally corrupted again by an external white noise Γ . The output power spectrum $O(\nu)$ can be expressed as
 580 a function of frequency ν :

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

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

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

586 The photoreceptor filter (Warland *et al.* 1997) was taken to be exponentially decaying in time, $f =$
 587 $\tau^{-1} \exp(-t/\tau)$ (for $t \geq 0$), with $\tau = 100$ ms. The curve $I(\nu)$ only depends on H , Γ and V through
 588 two independent parameters. For the plots in Fig. 7 we chose: $H = 3.38 \mu\text{m}^2\text{s}$, $\Gamma = 0.02$ spikes²s and
 589 $V = 307$ spikes²s, $\delta t = 20$ ms, and $L = 2,500$. In Fig. 7D, we plot the sensitivity to oscillating perturbation
 590 with fixed frequency ν , which results in $\sqrt{I(\nu)L/\delta t}$. In Fig. 7E we plot the spectral density of the transferred
 591 information rate:

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

-
- 592 Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing?
593 *Netw. Comput. Neural Syst.*, **3**(2), 213–251.
- 594 Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.*, **61**(3), 183–
595 193.
- 596 Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. *Sens.*
597 *Commun.*, **6**(2), 57–58.
- 598 Beck, J., Bejjanki, V., and Pouget, A. (2011). *Insights from a Simple Expression for Linear Fisher*
599 *Information in a Recurrently Connected Population of Spiking Neurons*. *Neural Computation* ,
600 **23**(6), 1484–1502.
- 601 Bell, A. J. and Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge
602 filters. *Vision Research*, **37**(23), 3327–3338.
- 603 Benichoux, V., Brown, A. D., Anbuhl, K. L., and Tollin, D. J. (2017). Representation of multi-
604 dimensional stimuli: quantifying the most informative stimulus dimension from neural responses.
605 *Journal of Neuroscience*.
- 606 Bernardi, D. and Lindner, B. (2015). A frequency-resolved mutual information rate and its appli-
607 cation to neural systems. *Journal of neurophysiology*, **113**(5), 1342–1357.
- 608 Berry, M. J. and Meister, M. (1998). Refractoriness and neural precision. *The Journal of neuro-*
609 *science : the official journal of the Society for Neuroscience*, **18**(6), 2200–11.
- 610 Berry, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving
611 stimuli by the retina. *Nature*, **398**(6725), 334–338.
- 612 Bialek, W., De Ruyter Van Steveninck, R. R., and Tishby, N. (2006). Efficient representation
613 as a design principle for neural coding and computation. In *IEEE International Symposium on*
614 *Information Theory - Proceedings*, pages 659–663.
- 615 Borghuis, B. G., Ratliff, C. P., Smith, R. G., Sterling, P., and Balasubramanian, V. (2008). Design
616 of a neuronal array. *The Journal of Neuroscience*, **28**(12), 3178–3189.
- 617 Botella-Soler, V., Deny, S., Marre, O., and Tkačik, G. (2016). Nonlinear decoding of a complex
618 movie from the mammalian retina. *arXiv*, **q-bio**(1605.03373v1), [q-bio.NC].
- 619 Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant,
620 J. L., and Rust, N. C. (2005). Do we know what the early visual system does? *The Journal of*

- 621 *neuroscience : the official journal of the Society for Neuroscience*, **25**(46), 10577–97.
- 622 Chechik, G., Anderson, M. J., Bar-Yosef, O., Young, E. D., Tishby, N., and Nelken, I. (2006).
623 Reduction of Information Redundancy in the Ascending Auditory Pathway. *Neuron*, **51**(3), 359–
624 368.
- 625 Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scenes in the lateral
626 geniculate nucleus: experimental test of a computational theory. *The Journal of neuroscience :*
627 *the official journal of the Society for Neuroscience*, **16**(10), 3351–3362.
- 628 Doi, E., Gauthier, J. L., Field, G. D., Shlens, J., Sher, A., Greschner, M., Machado, T. a., Jepson,
629 L. H., Mathieson, K., Gunning, D. E., Litke, A. M., Paninski, L., Chichilnisky, E. J., and Simoncelli,
630 E. P. (2012). Efficient coding of spatial information in the primate retina. *J. Neurosci.*, **32**(46),
631 16256–64.
- 632 Dong, D. W. and Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Compu-*
633 *tation in Neural Systems*, **6**(3), 345–358.
- 634 Faes, L., Nollo, G., Ravelli, F., Ricci, L., Vescovi, M., Turatto, M., Pavani, F., and Antolini,
635 R. (2007). Small-sample characterization of stochastic approximation staircases in forced-choice
636 adaptive threshold estimation. *Perception & psychophysics*, **69**(2), 254–262.
- 637 Fisher, R. A. (1936). *The Use of Multiple Measurements in Taxonomic Problems. Annals of*
638 *Eugenics*, **7**(2), 179–188.
- 639 Gilja, V., Nuyujukian, P., Chestek, C. A., Cunningham, J. P., Yu, B. M., Fan, J. M., Churchland,
640 M. M., Kaufman, M. T., Kao, J. C., Ryu, S. I., and Shenoy, K. V. (2012). A high-performance
641 neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, **15**(12), 1752–7.
- 642 Gollisch, T. and Meister, M. (2010). Eye smarter than scientists believed: neural computations in
643 circuits of the retina. *Neuron*, **65**(2), 150–64.
- 644 Heitman, A., Brackbill, N., Greschner, M., Sher, A., Litke, A. M., and Chichilnisky, E. (2016).
645 Testing pseudo-linear models of responses to natural scenes in primate retina. *bioRxiv*, page 045336.
- 646 Karklin, Y. and Simoncelli, E. P. (2011). Efficient coding of natural images with a population
647 of noisy Linear-Nonlinear neurons. *Advances in Neural Information Processing Systems (NIPS)*,
648 pages 1–9.
- 649 Keat, J., Reinagel, P., Reid, R. C., and Meister, M. (2001). Predicting every spike: A model for
650 the responses of visual neurons. *Neuron*, **30**(3), 803–817.
- 651 Kesten, H. (1958). Accelerated Stochastic Approximation. *The Annals of Mathematical Statistics*,

- 652 **29**(1), 41–59.
- 653 Kostal, L., Lansky, P., and Rospars, J. P. (2008). Efficient olfactory coding in the pheromone
654 receptor neuron of a moth. *PLoS Computational Biology*, **4**(4).
- 655 Liu, Y. S., Stevens, C. F., and Sharpee, T. (2009). Predictable irregularities in retinal receptive
656 fields. *Proceedings of the National Academy of Sciences*, **106**(38), 16499–16504.
- 657 Machens, C. K., Stemmler, M. B., Prinz, P., Krahe, R., Ronacher, B., and Herz, a. V. (2001).
658 Representation of acoustic communication signals by insect auditory receptor neurons. *Journal of*
659 *Neuroscience*, **21**(9), 3215–3227.
- 660 Machens, C. K., Wehr, M. S., and Zador, A. M. (2004). Linearity of cortical receptive fields
661 measured with natural sounds. *The Journal of neuroscience : the official journal of the Society*
662 *for Neuroscience*, **24**(5), 1089–100.
- 663 Macmillan, N. and Creelman, C. (2004). *Detection Theory: A User's Guide*. Taylor & Francis.
- 664 Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by
665 learning a sparse code for natural images.
- 666 Ölveczky, B. P., Baccus, S. A., and Meister, M. (2003). Segregation of object and background
667 motion in the retina. *Nature*, **423**(6938), 401–408.
- 668 Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli,
669 E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population.
670 *Nature*, **454**(7207), 995–999.
- 671 Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1996). *Spikes: Exploring*
672 *the Neural Code*. MIT Press, Cambridge, MA, USA.
- 673 Ruderman, D. L. and Bialek, W. (1992). Seeing beyond the Nyquist limit. *Neural computation*, **4**,
674 682–690.
- 675 Seung, H. S. and Sompolinsky, H. (1993). Simple models for reading neuronal population codes.
676 *Proc.Natl.Acad.Sci.*, **90**(22), 10749–10753.
- 677 Smith, E. C. and Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, **439**(7079), 978–982.
- 678 Van Hateren, J. (1992). *A theory of maximizing sensory information*. *Biological Cybernetics* ,
679 **68**(1), 23–29.
- 680 Warland, D. K., Reinagel, P., and Meister, M. (1997). Decoding visual information from a popu-
681 lation of retinal ganglion cells. *Journal of Neurophysiology*, **78**(5), 2336–2350.
- 682 Wei, X.-X. and Stocker, A. A. (2016). Mutual Information, Fisher Information, and Efficient

⁶⁸³ Coding. *Neural Computation*, **28**(2), 305–326.

⁶⁸⁴ Woyczynski, W. A. (2010). *A first course in statistics for signal analysis*. Springer Science &

⁶⁸⁵ Business Media.

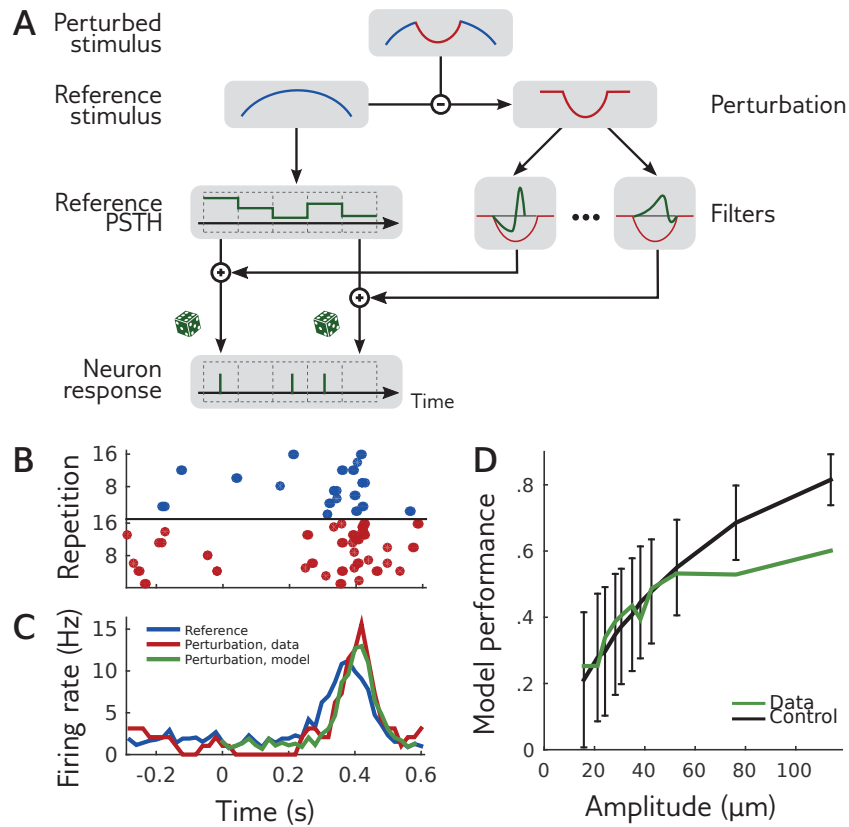


FIG. 4. Local model for responses to perturbations. **A.** The firing rates in response to a perturbation of a reference stimulus are modulated by filters applied to the perturbation. There is a different filter for each cell and each time bin. Because the model is conditionally independent across neurons 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.** Peristimulus time histogram (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's 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).

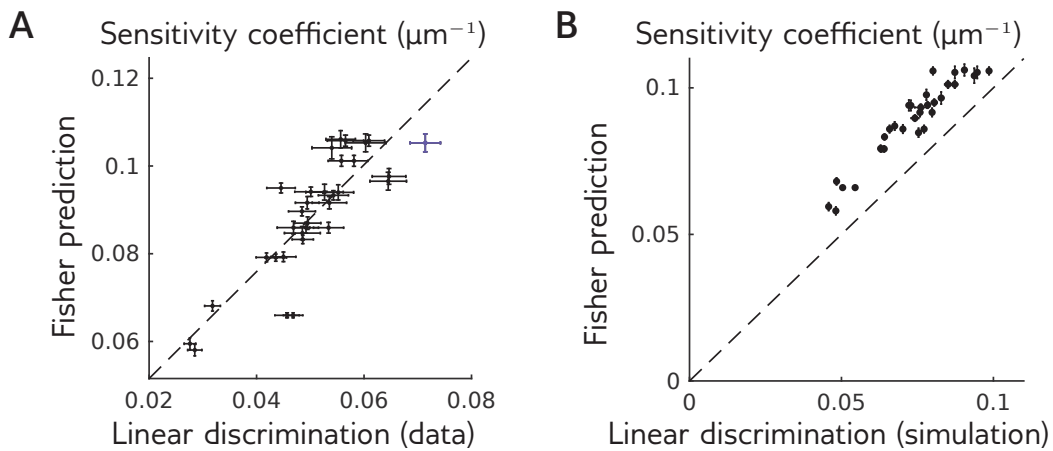


FIG. 5. The Fisher information predicts the experimentally measured sensitivity. A. Sensitivity coefficients c for the two reference stimuli and 16 perturbation shapes, measured empirically and predicted by the Fisher information (Eq. 14) and the local model. The purple point corresponds to the perturbation shown in Fig. 2. Dashed line stands for best linear fit. **B.** Same as B, 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 std over 10 simulations. Dashed line stands for identity.

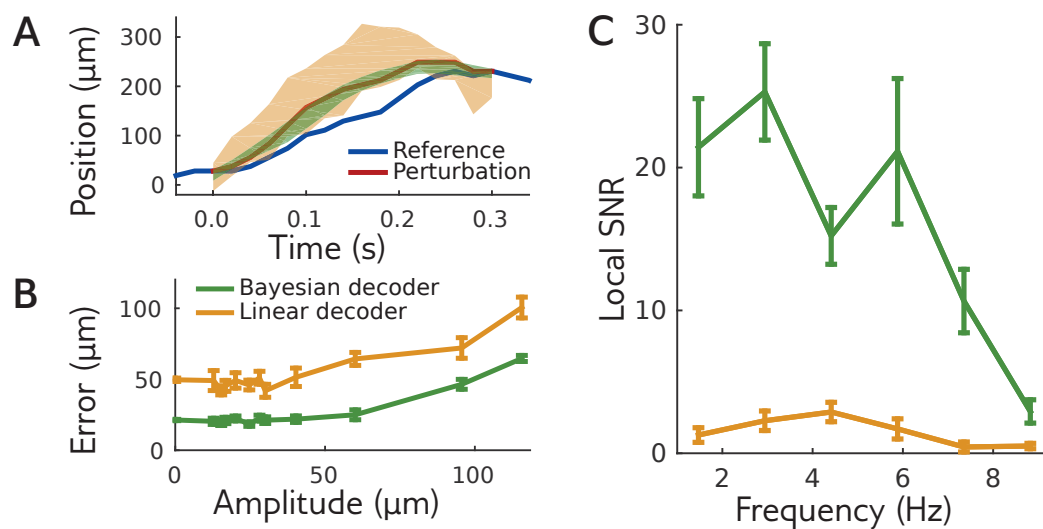


FIG. 6. Bayesian decoding of the local model outperforms the linear decoder. **A.** Responses to a perturbation of the reference stimulus (reference in blue, perturbation in red) are decoded using the local model (green) or a linear decoder (orange). For each decoder, the area shows one standard deviation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

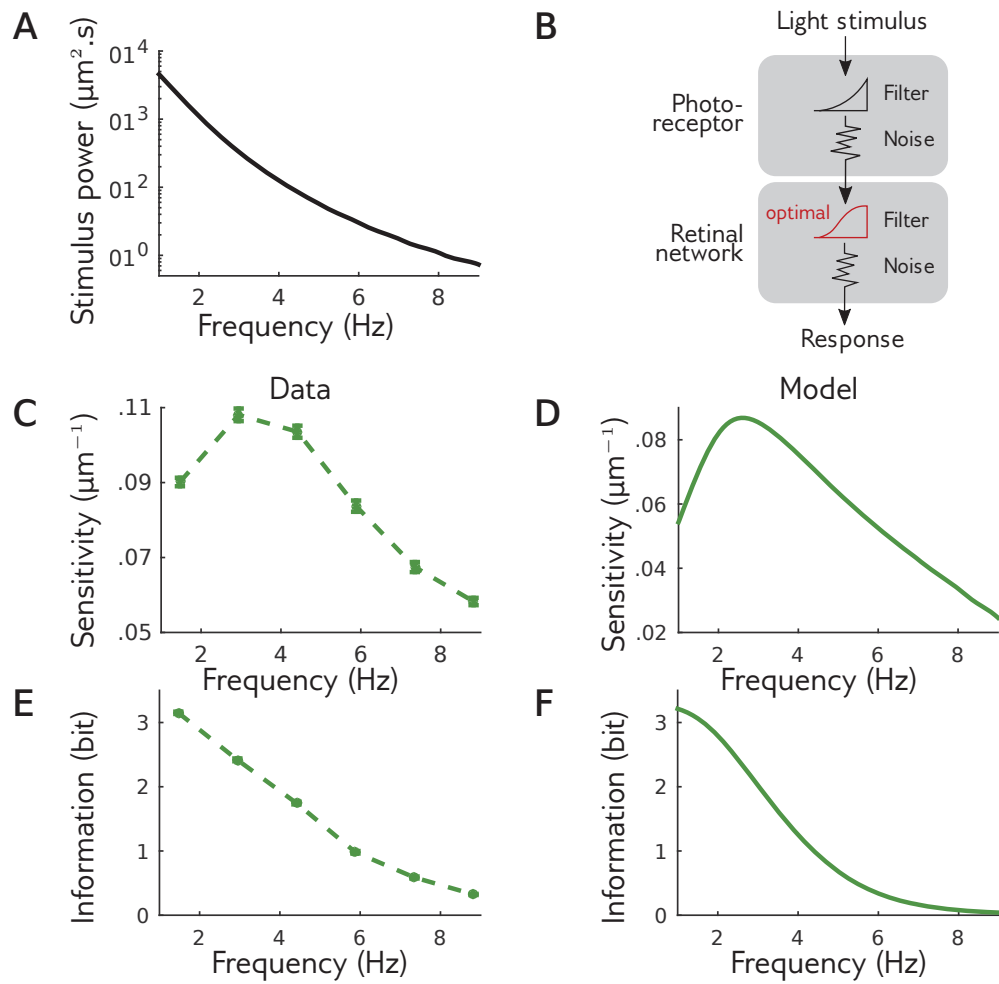


FIG. 7. Signature of efficient coding in the sensitivity **A**. Spectral density of the stimulus used in experiments, which is monotonically decreasing. **B**. Simple theory of retinal function: the stimulus is filtered by noisy photoreceptors, whose signal is then filtered by the noisy retinal network. 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 non monotonic 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.