Defining statistical perceptions with an empirical Bayesian approach

Satohiro Tajima
Science & Technology Research Laboratories, Japan Broadcasting Corporation, Tokyo, Japan

(Received 16 December 2012; revised manuscript received 18 March 2013; published 12 April 2013)

Extracting statistical structures (including textures or contrasts) from a natural stimulus is a central challenge in both biological and engineering contexts. This study interprets the process of statistical recognition in terms of hyperparameter estimations and free-energy minimization procedures with an empirical Bayesian approach. This mathematical interpretation resulted in a framework for relating physiological insights in animal sensory systems to the functional properties of recognizing stimulus statistics. We applied the present theoretical framework to two typical models of natural images that are encoded by a population of simulated retinal neurons, and demonstrated that the resulting cognitive performances could be quantified with the Fisher information measure. The current enterprise yielded predictions about the properties of human texture perception, suggesting that the perceptual resolution of image statistics depends on visual field angles, internal noise, and neuronal information processing pathways, such as the magnocellular, parvocellular, and koniocellular systems. Furthermore, the two conceptually similar natural-image models were found to yield qualitatively different predictions, striking a note of warning against confusing the two models when describing a natural image.

DOI: 10.1103/PhysRevE.87.042707 PACS number(s): 87.19.ls

I. INTRODUCTION

The findings of a line of studies on natural statistics have shown that stimuli, such as natural scene images, are known to have characteristic statistical structures other than unstructured noise [1–3]. For properly set statistical parameters, we can perceive specific figures, textures, or scenes, which are different from random patterns of light intensities. The processing of natural images based on those statistical regularities has been investigated for its close relationship to the statistical mechanics of lattice systems [4–6]. Generally, statistical values differ among stimuli and vary for each local region within a stimulus. Fluctuations in statistical values are not always random but are often found to have ecological implications for living things. For example, the spatial smoothness of the luminance contrast in a visual stimulus is known to provide cues for recognizing image blur [7], texture [8], and scene category [9]. Furthermore, a line of psychophysical studies has reported that human subjects are actually able to recognize and discriminate the statistics that are related to image smoothness [10–13]. According to these insights, we can naturally expect that animals perceive changes in stimulus statistics in order to obtain valuable information for their survival.

If animals can perceive stimulus statistics, what mechanisms make this possible? What are the factors that determine the resolution of the perception? How do biophysical properties, such as the neuronal receptive field, affect or restrict the perceptual consequences? Although there have been intensive discussions on neuronal receptive-field structures in relation to the efficient encoding of natural images [14–18], few studies have focused on the encoding of the global statistics of images in neural function. However, from a functional viewpoint, the global statistics are not trivial, but rather, they are more important than the pointwise representation of image data, because the statistical structure conveys ecologically meaningful cues for animals, as mentioned above.

In this study, the perceptions of stimulus statistics in the animal sensory system are related to hyperparameter estimations in Bayesian noise reduction. We mathematically interpret the perception of stimulus statistics based on neural population activity, in terms of empirical Bayesian inference. We propose a theoretical framework for quantifying the perceptual resolution of stimulus statistics and deriving the relationship between perceptual resolution and neural receptive-field structures. For a concrete example, we apply the framework to the problem of the retinal information coding of natural-image statistics, by simulating the typical receptive-field structures of the retinal ganglion cells. The aim of the present simulation is first to demonstrate the relationship between the receptive-field structure and the information encoding of global statistics, by using a simple linear model. In addition, we discuss how the current model can be related to more realistic models of neurons, including the nonlinearity of the response functions.

II. METHODS

A. Empirical Bayesian framework

We first derive a mathematical interpretation of the perception of stimulus statistics with an encoding and decoding model of stimuli in the nervous system from the viewpoint of an empirical Bayes framework.In summary, the empirical Bayes framework results from the use of hierarchical generative models, in which prior beliefs about certain parameters are functions of other parameters, which are sometimes referred to as *hyperparameters*. In particular, the empirical Bayes method considers the cases in which a prior belief of the parameter has to be set depending on the observed data. The prior belief on the parameter that is conditioned by a hyperparameter is referred to as an empirical prior. In the aforementioned example of luminance smoothness, the hyperparameters determine the prior beliefs about the second-order statistics of images, such as spatial correlations. Our focus in this paper is on the hyperparameters which, as we see below, correspond to precision or covariance parameters that, in turn, determine the likelihood of a particular visual input. The use of the empirical Bayes method and implicit hierarchical models is important because hierarchical inference is a popular metaphor of perceptual processing in the brain [17,19–22]. When applied to time-varying inputs, it leads to Bayesian filtering schemes, such as the Kalman filter [23]. These schemes can be formulated in terms of predictive coding, and they can be regarded as a biologically plausible implementation of hierarchical Bayesian inference [24].

Consider that the task of the central nervous system is to estimate (decode) the information in the input stimulus ($\boldsymbol{s}$) according to the output ($\boldsymbol{r}$) of early sensory neurons. This corresponds to obtaining $P(\boldsymbol{s}|\boldsymbol{r})$, which is the posterior probability distribution of $\boldsymbol{s}$ under the observed data $\boldsymbol{r}$. The posterior probability is given as follows using Bayes' theorem:

$$P(\boldsymbol{s}|\boldsymbol{r};\boldsymbol{\theta}) \propto P(\boldsymbol{r}|\boldsymbol{s})P(\boldsymbol{s}|\boldsymbol{\theta}). \tag{1}$$

The likelihood $P(\boldsymbol{r}|\boldsymbol{s})$ depends on the noise properties and receptive field structures of early sensory neurons, while the prior $P(\boldsymbol{s}|\boldsymbol{\theta})$ is determined by knowledge of the input stimuli.

Here, $\boldsymbol{\theta} = (\theta_1, \ldots ,\theta_n)$ is a set of hyperparameters. In this study, $\boldsymbol{\theta}$ is considered to reflect knowledge of the stimulus statistics (e.g., image smoothness or contrast). Since statistical values differ between input stimuli, $\boldsymbol{\theta}$ cannot be given *a priori*; for example, if $\boldsymbol{\theta}$ represents image smoothness, its value depends on the properties of the input stimulus, as previously described. This means that the hyperparameter concerning stimulus statistics, $\boldsymbol{\theta}$, needs to be estimated only from the output of the early sensory system, $\boldsymbol{r}$. Determination of the prior according to the observed data needs an empirical Bayes approach, which is in contrast to the conventional Bayesian estimation, in which the prior distribution is given *a priori*.

The posterior distribution of $\boldsymbol{\theta}$ under an observation of $\boldsymbol{r}$ is determined, again, using the Bayes' theorem,

$$P(\boldsymbol{\theta}|\boldsymbol{r}) \propto P(\boldsymbol{r}|\boldsymbol{\theta})P(\boldsymbol{\theta}). \tag{2}$$

We then consider the case in which there is no prior knowledge of the stimulus statistics and assume $P(\boldsymbol{\theta})$ is an uninformative prior,

$$P(\boldsymbol{\theta}|\boldsymbol{r}) \propto P(\boldsymbol{r}|\boldsymbol{\theta}). \tag{3}$$

$P(\boldsymbol{r}|\boldsymbol{\theta})$ is called the *marginalized likelihood* [25] or *evidence* [26] (alternatively termed*type-II likelihood* [27,28] or simply called the *likelihood of the statistical model* [29,30]). This value is formally obtained through the following marginalization over the input stimulus:

$$P(\boldsymbol{r}|\boldsymbol{\theta}) = \int d\boldsymbol{s}\, P(\boldsymbol{r}|\boldsymbol{s})P(\boldsymbol{s}|\boldsymbol{\theta}). \tag{4}$$

The logarithm of the marginalized likelihood, $-\ln P(\boldsymbol{r}|\boldsymbol{\theta})$, can be related to the Helmholtz free energyin a precise analogy with statistical mechanics, by expressing the posterior in the form of a Gibbs distribution as described later [20]. If one optimized the hyperparameters with respect to the marginalized likelihood, this would correspond to an exact empirical Bayesian inference. However, in practice, this is not usually a tractable solution, and its variational bound that is computed with an approximated posterior is optimized instead. This bound is also referred as the (variational) *free energy* [20,31]; such that maximizing the variational free energy maximizes the evidence approximately—leading to an approximate Bayesian inference. It can be shown that the expected value of the variational free energy that is defined with the estimated hyperparameter $\hat{\boldsymbol{\theta}}$ is bounded by the true free energy:

$$\begin{aligned}\langle -\ln P(\boldsymbol{r}|\hat{\boldsymbol{\theta}})\rangle_{\boldsymbol{r}|\boldsymbol{\theta}} &= \langle -\ln P(\boldsymbol{r}|\boldsymbol{\theta})\rangle_{\boldsymbol{r}|\boldsymbol{\theta}} \\ &\quad + D_{KL}[P(\boldsymbol{r}|\boldsymbol{\theta})||P(\boldsymbol{r}|\hat{\boldsymbol{\theta}})] \qquad (5)\\ &\geqslant \langle -\ln P(\boldsymbol{r}|\boldsymbol{\theta})\rangle_{\boldsymbol{r}|\boldsymbol{\theta}}, \qquad (6)\end{aligned}$$

and that we have the equality only when $\hat{\boldsymbol{\theta}} \equiv \boldsymbol{\theta}$, except for some special cases [32]. As noted above, there is a large literature on optimizing the free energy in the context of empirical Bayesian filtering and predictive coding.

In an empirical Bayes framework, the optimal prior $P(\boldsymbol{s}|\boldsymbol{\theta})$ is given by choosing a $\boldsymbol{\theta}$ that maximizes $P(\boldsymbol{r}|\boldsymbol{\theta})$. Importantly, in this framework the process of estimating the hyperparameter $\boldsymbol{\theta}$ from the sensory output $\boldsymbol{r}$ can be interpreted as the mathematical interpretation of the perceptual organization of stimulus statistics.

B. Fisher information

We introduce the Fisher information in order to measure how precisely the hyperparameter can be estimated. In a Laplace approximation of empirical Bayesian inference, the Fisher information matrix is simply the posterior precision or confidence about the hyperparameter estimates. The Laplace assumption means that the approximate posterior (or likelihood) distribution is assumed to have a Gaussian form, which is asymptotically equal to the exact distribution at the limit of large sample size. Hereafter, we will refer to the Fisher information as the (upper bound of) precision in the hyperparameter estimates that are related to stimulus statistics. The estimate of the hyperparameter ($\hat{\boldsymbol{\theta}}$) that maximizes the marginalized likelihood is an unbiased estimator. The inverse of the Fisher information gives the lower bound of each hyperparameter estimate $\hat{\theta}_\mu$ through the following Cramér-Rao inequality:

$$\mathrm{Var}[\hat{\theta}_\mu] \geqslant \mathcal{J}_{\theta_\mu}^{-1}, \tag{7}$$

where $\mathcal{J}_{\theta_\mu}$ is the Fisher information of θ_μ,

$$\mathcal{J}_{\theta_\mu} \equiv \mathrm{E}\left[-\frac{\partial^2 \ln P(\boldsymbol{r}|\boldsymbol{\theta})}{\partial\theta_\mu^2} \middle| \boldsymbol{\theta} \right]. \tag{8}$$

Equation (7) yields a measure of the theoretical precision of the hyperparameter estimation. As $\boldsymbol{\theta}$ represents the knowledge of the stimulus statistics, $\mathcal{J}_{\theta_\mu}$ is interpreted as the perceptual resolution of the stimulus statistics.

C. Image encoding by the retina

We will consider the relationship between natural-image statistics and retinal responses. In this situation, $\boldsymbol{s}$ and $\boldsymbol{r}$ correspond to the true visual input and the output of retinal ganglion cells that constitute the information received by the brain, respectively. Bayesian inference then has to recover the true visual input $\boldsymbol{s}$, given retinal output $\boldsymbol{r}$. These are given in vector form, in which the vector elements are arrayed as pixel luminance. The hyperparameter $\boldsymbol{\theta}$ corresponds to the statistical parameters of each stimulus. Of particular interest are the contrast and the smoothness, which are two of the most established statistical parameters. *Contrast* is quantified by the variability of the signal intensity (e.g., pointwise luminance in visual stimuli), while *smoothness* is characterized by the spatiotemporal correlation of light intensities (luminance).

The prior distribution of the input image $P(\boldsymbol{s}|\boldsymbol{\theta})$ is formally written as a form of the Gibbs distribution:

$$P(\boldsymbol{s}|\boldsymbol{\theta}) = \frac{e^{-H(\boldsymbol{s};\boldsymbol{\theta})}}{\int d\boldsymbol{s} e^{-H(\boldsymbol{s};\boldsymbol{\theta})}}. \tag{9}$$

$H(\boldsymbol{s};\boldsymbol{\theta})$ is a value that depends upon the configuration of $\boldsymbol{s}$ and can be thought of as the energy related to the prior distribution, in analogy with statistical mechanics. We will refer to this as the *prior energy*. Here we assume that $H(\boldsymbol{s};\boldsymbol{\theta})$ is represented by the following quadratic form of $\boldsymbol{s}$:

$$H(\boldsymbol{s};\boldsymbol{\theta}) = \boldsymbol{s}^\top \boldsymbol{U}\boldsymbol{s}, \tag{10}$$

where the $\top$ denotes the transposition of the matrix. $\boldsymbol{U}$ corresponds to a precision matrix, such that if the prior beliefs on the images are held with great precision, the prior energy is greater. If the $\boldsymbol{U}$ satisfies the translational symmetry criteria of the image, then $H(\boldsymbol{s};\boldsymbol{\theta})$ is rewritten with a two-dimensional Fourier transform:

$$H(\boldsymbol{s};\boldsymbol{\theta}) = \sum_k \frac{\tilde{U}_k}{N} |\tilde{s}_k|^2, \tag{11}$$

where the tilde denotes a (two-dimensional) Fourier transform. The retinal ganglion cells can be modeled as a communication channel that determines the observation model $P(\boldsymbol{r}|\boldsymbol{s})$ as follows:

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}, \tag{12}$$

$$\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, R^{-1}). \tag{13}$$

The transformation $\boldsymbol{A}$ depends on the receptive field structures of the retinal ganglion cells. The neuronal receptive fields in the early visual system are generally found to be described well by linear filters. For example, the majority of retinal ganglion cells have Mexican-hat-type receptive fields with lateral inhibitions [33–36], which function as bandpass filters in the spatial frequency domain. $\boldsymbol{n}$ denotes the trial-to-trial fluctuation of neural responses (e.g., firing rates), which are assumed to follow independent normal distributions with mean $\boldsymbol{0}$ and variance R^{-1}. For the simulations and results presented below, R is a scalar or a scaled identity matrix. In this case, the likelihood is

$$P(\boldsymbol{r}|\boldsymbol{s}) = \frac{e^{-L(\boldsymbol{s},\boldsymbol{r};\boldsymbol{A},R)}}{\int d\boldsymbol{r} e^{-L(\boldsymbol{s},\boldsymbol{r};\boldsymbol{A},R)}}, \tag{14}$$

where

$$L(\boldsymbol{s},\boldsymbol{r}) \equiv (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{s})^\top R(\boldsymbol{r} - \boldsymbol{A}\boldsymbol{s}). \tag{15}$$

From Eqs. (10) and (14), the posterior is also expressed in the form of a Gibbs distribution $P(\boldsymbol{s}|\boldsymbol{r},\boldsymbol{\theta}) = e^{-E}/\int d\boldsymbol{r} e^{-E}$, where the energy $E \equiv H + L$ comprises a prior energy (H) and a likelihood potential (L). We will consider a simple case in which the shape of the visual receptive field is symmetrical, and $\boldsymbol{s}$ and $\boldsymbol{r}$ have the same dimension (N). This roughly approximates the situation in the fovea (the central region of the visual field) of primate retina [35,36]. Application of a two-dimensional Fourier transform to Eq. (15) yields

$$L(\boldsymbol{s},\boldsymbol{r}) = \sum_k \frac{R}{N} |\tilde{r}_k - \tilde{A}_k \tilde{s}_k|^2. \tag{16}$$

The receptive fields of retinal ganglion cells are approximated by a Gaussian function, or the Laplacian transform of the Gaussian function [Laplacian of Gaussian (LoG)]. Specifically, the cells leading to the magnocellular and the parvocellular pathways have receptive fields that are approximated by a LoG, while the cells that correspond to the koniocellular pathway have Gaussian-like receptive fields. LoG and Gaussian filters are respectively described as follows:

$$\tilde{A}_k = -\frac{4\pi^2\sqrt{2\pi}\gamma^3||\boldsymbol{k}||^2}{N} \exp\left(-\frac{2\pi^2\gamma^2||\boldsymbol{k}||^2}{N}\right), \tag{17}$$

$$\tilde{A}_k = \sqrt{2}\exp\left(-\frac{2\pi^2\gamma^2||\boldsymbol{k}||^2}{N}\right). \tag{18}$$

D. Natural-image priors

Here, we describe the following two conventional statistical models of a natural image: the nearest-neighbor model and the power-law model. The nearest-neighbor model focuses on the luminance differences between every pair of neighboring pixels, formulating the prior energy as follows:

$$H(\boldsymbol{s};\beta,h) = \beta \sum_{\boldsymbol{q}} \sum_{\boldsymbol{q}' \in B(\boldsymbol{q})} (s_{\boldsymbol{q}} - s_{\boldsymbol{q}'})^2 + h \sum_{\boldsymbol{q}} s_{\boldsymbol{q}}^2, \tag{19}$$

where the hyperparameters β and h are positive scalars. A larger value of β means that smoother images are likely to be generated; a larger value of h indicates higher contrast. $B(\boldsymbol{q})$ denotes the neighbor of the pixel that is located at $\boldsymbol{q}$. Considering the four nearest pixels (two horizontal and two vertical) for $B(\boldsymbol{q})$, the first term on the left-hand side of Equation (19) is

$$\sum_{\boldsymbol{q}} \sum_{\boldsymbol{q}' \in B(\boldsymbol{q})} (s_{\boldsymbol{q}} - s_{\boldsymbol{q}'})^2 = \boldsymbol{s}^T \boldsymbol{J}\boldsymbol{s}. \tag{20}$$

All pairs of juxtaposed pixels are related by the matrix $\boldsymbol{J}$, whose component is given as follows:

$$J_{\boldsymbol{q},\boldsymbol{q}'} = 2\delta_{\boldsymbol{q},\boldsymbol{q}'} - \sum_{\boldsymbol{\epsilon}} \delta_{\boldsymbol{q},\boldsymbol{q}'+\boldsymbol{\epsilon}} - \sum_{\boldsymbol{\epsilon}} \delta_{\boldsymbol{q},\boldsymbol{q}'-\boldsymbol{\epsilon}}, \tag{21}$$

where δ is Kronecker's delta, and $\boldsymbol{\epsilon} \in \{(1,0),(0,1)\}$. Introducing a positive definite matrix $\boldsymbol{U} \equiv \beta\boldsymbol{J} + h\boldsymbol{I}$, $H(\boldsymbol{s};\beta,h)$ is expressed as $H(\boldsymbol{s};\beta,h) = \boldsymbol{s}^T\boldsymbol{U}\boldsymbol{s}$, or

$$H(\boldsymbol{s};\beta,h) = \sum_k \frac{\tilde{U}_k}{N} |\tilde{s}_k|^2, \tag{22}$$

$$\tilde{U}_k = \beta\tilde{J}_k + h, \tag{23}$$

$$\tilde{J}_k = 4 - 2\cos\frac{2\pi k_x}{\sqrt{N}} - 2\cos\frac{2\pi k_y}{\sqrt{N}} \tag{24}$$

in the spatial frequency domain.

In the power-law model, the image smoothness is directly defined in the spatial frequency domain. It expects natural images to have a spatial-frequency amplitude spectrum that follows a power law (exponential distribution). In this case,

TABLE I. Natural-image models and variables related to the estimation of stimulus statistics.

	Nearest-neighbor model	Power-law model
Smoothness statistics	β	α
Contrast statistics	h	c
Precision matrix ($\tilde{U}_k$)	$\beta\tilde{J}_k + h$	$\frac{\|\|\boldsymbol{k}\|\|^{2\alpha}}{2c^2}$
Precision of smoothness estimates	$\mathcal{J}_\beta = \sum_{\boldsymbol{k}} \frac{\tilde{J}_k^2 R^2 \|\tilde{A}_k\|^4}{2\tilde{U}_k^2 \tilde{V}_k^2}$	$\mathcal{J}_\alpha = \sum_{\boldsymbol{k}} \frac{2R^2 \|\tilde{A}_k\|^4 (\ln \|\|\boldsymbol{k}\|\|)^2}{\tilde{V}_k^2}$
Precision of contrast estimates	$\mathcal{J}_h = \sum_{\boldsymbol{k}} \frac{R^2 \|\tilde{A}_k\|^4}{2\tilde{U}_k^2 \tilde{V}_k^2}$	$\mathcal{J}_c = \sum_{\boldsymbol{k}} \frac{2R^2 \|\tilde{A}_k\|^4}{c^2 \tilde{V}_k^2}$

the prior energy is given as follows:

$$H(\boldsymbol{s};\alpha,c) = \sum_k \frac{\tilde{U}_k}{N} |\tilde{s_k}|^2, \tag{25}$$

$$\tilde{U}_k \equiv \frac{1}{2(c||\boldsymbol{k}||^{-\alpha})^2} = \frac{||\boldsymbol{k}||^{2\alpha}}{2c^2}, \tag{26}$$

where the hyperparameters α and c are related to the smoothness and the contrast, respectively. A larger value of α indicates a smoother image, while a larger value of c means higher contrast. On the basis of the above formulations, Table I summarizes the hyperparameters and important variables in the two natural-image models, where $\tilde{V}_k \equiv \tilde{U}_k + R|\tilde{A}_k|^2$.

E. Reconstruction accuracy

This section describes the measure of encoding accuracy in terms of the mean square error (MSE) $\mathcal{E}$ in full-image reconstruction, which will be compared to the Fisher information measure of precision in the hyperparameter estimation. To assess the accuracy, the MSE compares the Bayesian estimate (*reconstruction*) of visual input ($\hat{\boldsymbol{s}}$) to the true value ($\boldsymbol{s}$): $\mathcal{E} \equiv ||\boldsymbol{s} - \hat{\boldsymbol{s}}||^2/N$. Notice that, in order to estimate the visual input, it is necessary to optimize both $\hat{\boldsymbol{s}}$ and the hyperparameter estimate $\hat{\boldsymbol{\theta}}$. The expected value of the MSE is bounded as

$$\langle\mathcal{E}\rangle_{\boldsymbol{s},\boldsymbol{r}|\boldsymbol{\theta}} = \int d\boldsymbol{r} \int d\boldsymbol{s}\, P(\boldsymbol{s},\boldsymbol{r}|\boldsymbol{\theta})\mathcal{E}(\boldsymbol{s},\boldsymbol{r};\boldsymbol{\theta}) \tag{27}$$

$$= \sum_{\boldsymbol{k}} \frac{1}{2N\widetilde{V}_{\boldsymbol{k}}} \left\{1 + \frac{R|\widetilde{A}_{\boldsymbol{k}}|^2}{\widetilde{U}_{\boldsymbol{k}}(\boldsymbol{\theta})} \left(\frac{R}{\widetilde{V}_{\boldsymbol{k}}(\hat{\boldsymbol{\theta}})}\right)^2 \times \left(\frac{\widetilde{U}_{\boldsymbol{k}}(\hat{\boldsymbol{\theta}})}{R} - \frac{\widetilde{U}_{\boldsymbol{k}}(\boldsymbol{\theta})}{R}\right)^2\right\} \tag{28}$$

$$\geqslant \sum_{\boldsymbol{k}} \frac{1}{2N\widetilde{V}_{\boldsymbol{k}}(\boldsymbol{\theta})}, \tag{29}$$

where $\widetilde{V}_{\boldsymbol{k}}$ and $\widetilde{U}_{\boldsymbol{k}}$ are computed as functions of $\hat{\boldsymbol{\theta}}$ or $\boldsymbol{\theta}$. The equality condition is $\widetilde{U}_{\boldsymbol{k}}(\hat{\boldsymbol{\theta}}) = \widetilde{U}_{\boldsymbol{k}}(\boldsymbol{\theta})$ ($\forall\, \boldsymbol{k}$), which is obtained for the hyperparameter estimates $\hat{\boldsymbol{\theta}}$ that are matched to the true values $\boldsymbol{\theta}$. In fact, it is shown that this lower bound quantitatively well approximates the empirically derived MSE of image reconstruction, which is based on the hyperparameters estimated by gradient descent of the variational free energy [32]. In the simulation results appearing in the subsequent section, we used the analytic lower bound $\sum_{\boldsymbol{k}} 1/2N\widetilde{V}_{\boldsymbol{k}}(\boldsymbol{\theta})$ to quantify the reconstruction accuracy.

III. RESULTS

A. Retinal encoding capacity of natural-image statistics

The precision of the hyperparameter estimation can be quantified by the Fisher information that is computed for the marginalized likelihood $P(\boldsymbol{r}|\boldsymbol{\theta})$ (see Sec. II). With the Fisher information, the receptive-field structures of sensory neurons are related to their ability to transfer information about stimulus statistics. For a concrete example, we focus on the relationship between natural-image statistics and the retina. Using the simulation protocol described in the previous section, we evaluated the performance of the two conventional models of natural images (see Sec. II D): one considers the luminance differences between the nearest-neighboring pixels [37,38], while the other is based on the power law of the spatial-frequency spectrum [1,39–43]. Figure 1 shows how the image appearance changes depending on the hyperparameters.

Figure 2 illustrates how the Fisher information of each statistic depends on the parameter that controls the receptive-field size (γ). The case of a LoG filter is shown in the figure. For comparison, it also shows the performance in reconstructing the original image as measured with the MSE. Note that a lower MSE means a more accurate reconstruction, while a larger Fisher information value indicates better performance in transferring the stimulus statistics information. The figure illustrates that an optimal receptive-field size exists for each of the three criteria (MSE, Fisher information of smoothness, and Fisher information of contrast), and that the optimal receptive-field size depends on which criterion is selected.

Figure 3 shows the optimal receptive-field size for each of the three criteria under various neural noise conditions or input stimulus statistics. It illustrates the general tendency that a larger receptive-field size is more advantageous for noisy channel conditions and for a smooth stimulus. Generally, the neural noise level is considered to increase under a lower illumination. With a larger noise level, it has been suggested that a larger filter (weighting signals at lower spatial frequency) becomes advantageous from the viewpoint of efficient and robust image encoding [14]. The present results derived for the MSE criterion are consistent with the previous view. Furthermore, the results for the Fisher information criterion suggest that the noise-to-receptive-field relationship also holds for the encoding of the global statistics of the image. More detailed observation reveals that the choice of the smoothness or the contrast criterion strongly affects the optimal receptive-field size in the nearest-neighbor model,

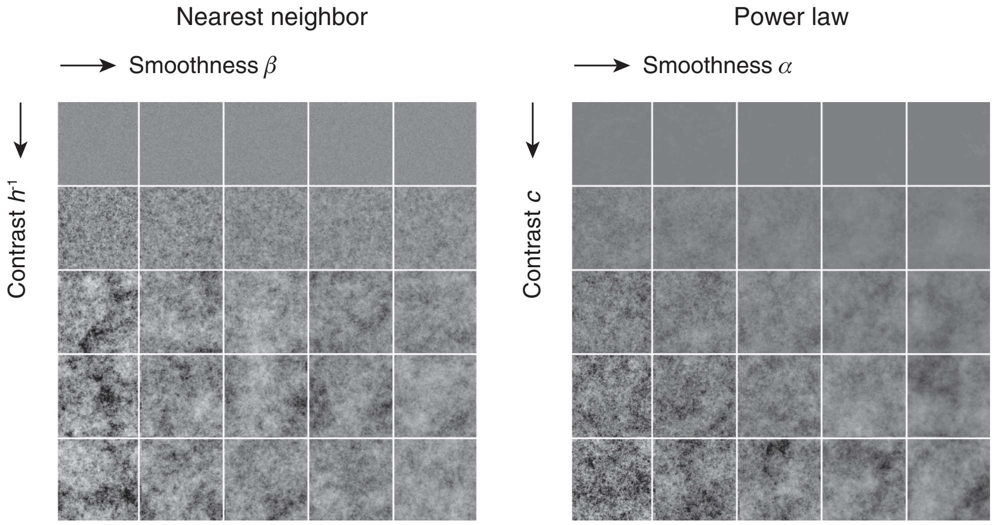

FIG. 1. Changes of the image appearance due to different statistical parameter values. The figure shows examples of artificial noise stimuli that are generated according to the two popular natural-image models (see Sec. II).

while it has little influence in the power-law model. In addition, the two popular natural-image models with power-law-type and nearest-neighbor-type interactions were found to lead to qualitatively different predictions.

B. Model dependency

Comparing Figs. 3(c) and 3(f), the nearest-neighbor model suggests that a larger receptive field is advantageous for higher contrast, while the power-law model suggests that a smaller receptive field is advantageous for higher contrast. This is due to the difference in parametrizing between the two natural-image models, although they are based on roughly similar concepts. In previous works, many physicists have constructed natural-image models based on Ising spin models, assuming interactions between neighboring pairs of pixels [37,38,44], and sometimes concluded that Ising spin models also lead to scale-free characteristics similar to the power-law model [44]. However, the current results demonstrate that the result of natural-image analysis can depend on the way a natural image is modeled, striking a note of warning against confusing the two models when describing a natural image.

C. Ability of visual pathways

Figure 4 compares the performances of three different simulated retinogeniculate visual processing pathways: the magnocellular, parvocellular, and koniocellular systems,

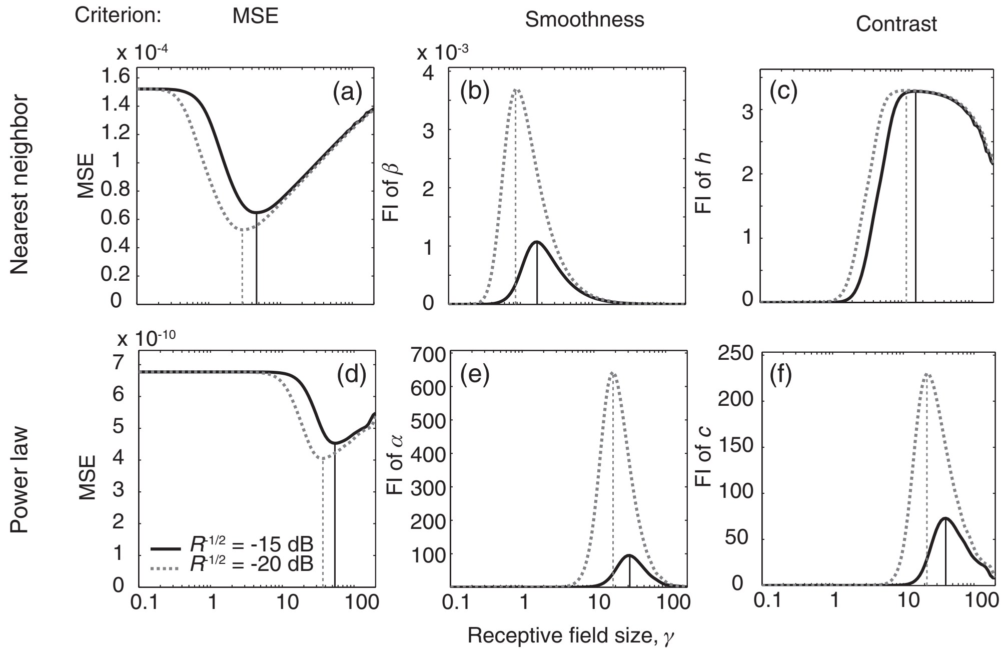

FIG. 2. Optimal filter widths for reconstruction and hyperparameter estimation. (a)–(c) The nearest-neighbor model; $\beta = 3000$, $h = 1$. (d)–(f) The power-law model; $\alpha = 1$, $c = 1$.

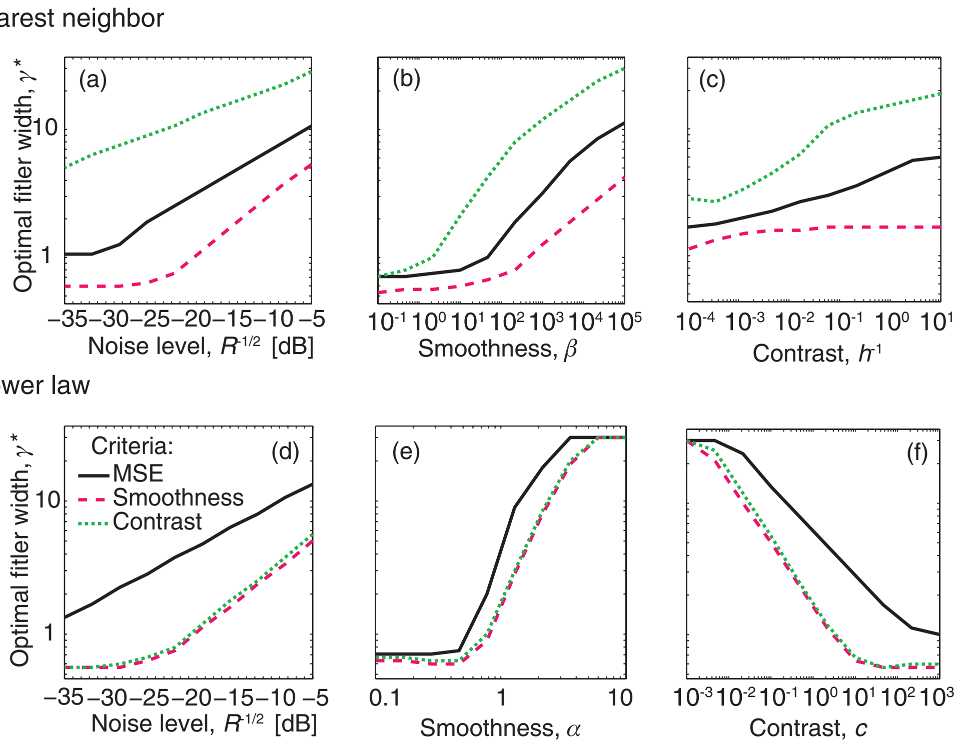

FIG. 3. (Color online) Dependency of optimal filter width on input image statistics and channel noise. (a)–(c) The nearest-neighbor model; $\beta = 3000$, $h = 1$, $R^{-1/2} = -15$ dB. (d)–(f) The power-law model; $\alpha = 1$, $c = 1$, $R^{-1/2} = -15$ dB.

which correspond to the three different types (parasol, midget, and small bistratified) of retinal ganglion cells in the primate retina. We used LoG filters with $\gamma = 10$ and 5 to simulate the magnocellular and the parvocellular pathways, and used a Gaussian with $\gamma = 5$ for the koniocellular pathway, which roughly approximates the scale orders of the receptive-field structures at cone resolution, as has been observed in previous studies [36,45]. Under various conditions of neural noise and stimulus statistics, the present simulation yielded information about the resolution of each visual pathway when encoding the stimulus statistics and how that resolution depends on the neural noise or stimulus statistics. For example, the performance of the magnocellular pathway, which receives signals from parasol ganglion cells, is expected to be more mildly affected by changes in the values of the stimulus statistics, when compared to the other two visual pathways. These results demonstrate the encoding efficiency of each visual pathway, which is normalized by the cell density and response gains. Because the cell distributions and response gains are not uniform in the real nervous system, it should be noted that these factors have to be incorporated for a more realistic analysis. In addition, we also have to be aware of the nonlinearity and contrast adaptation in the cell responses when we try to relate the present results to a real nervous system, as we will discuss in a later section. Nevertheless, the results presented here are expected to provide a useful basis for such an advanced analysis in the future.

IV. DISCUSSION

In this study, we proposed a model of stimulus statistics perception based on encoding by the early sensory system, in terms of empirical Bayes inference. Using the Fisher information as a measure of precision, we analyzed the functional properties of the receptive-field structures in the early visual system in recognizing stimulus statistics.

The present study simplifies the response properties of visual nervous cells in several ways. For example, the ganglion cells in the real retina show nonlinear response functions, and, in that sense, the "linear filter + noise" model that was considered in the present study is a simplification of real retinal ganglion cells. However, such a simple linear model has been found to be useful for a basic understanding of the early nervous system [14,46]. In addition, the linear model can be interpreted as approximating a combined output of nonlinear units. Here, we demonstrated the relevance of the present linear model to several nonlinear models that have been proposed in the context of early visual processing (Fig. 5). The most profound nonlinearity in the neuronal response is rectification, which is led by the threshold of spike generation. The rectification process strongly affects the information transmission because it loses information that is conveyed by subthreshold input signals. However, the loss of information can be compensated for by combining the outputs of two different cells that have opposite response polarities (i.e., on- and off-type ganglion cells in the retina), as in a push-pull circuit. We tested three types of rectification nonlinearity: rectification that is followed by additive noise [Fig. 5(b)], rectification followed by additive noise that is limited to positive average response, which roughly simulates the non-negative property of cell spiking [47] [Fig. 5(d)], and rectification after noise perturbation, which was originally proposed to explain the power-function-like response function of neurons in the early visual cortex [48–50] [Fig. 5(f)]. For

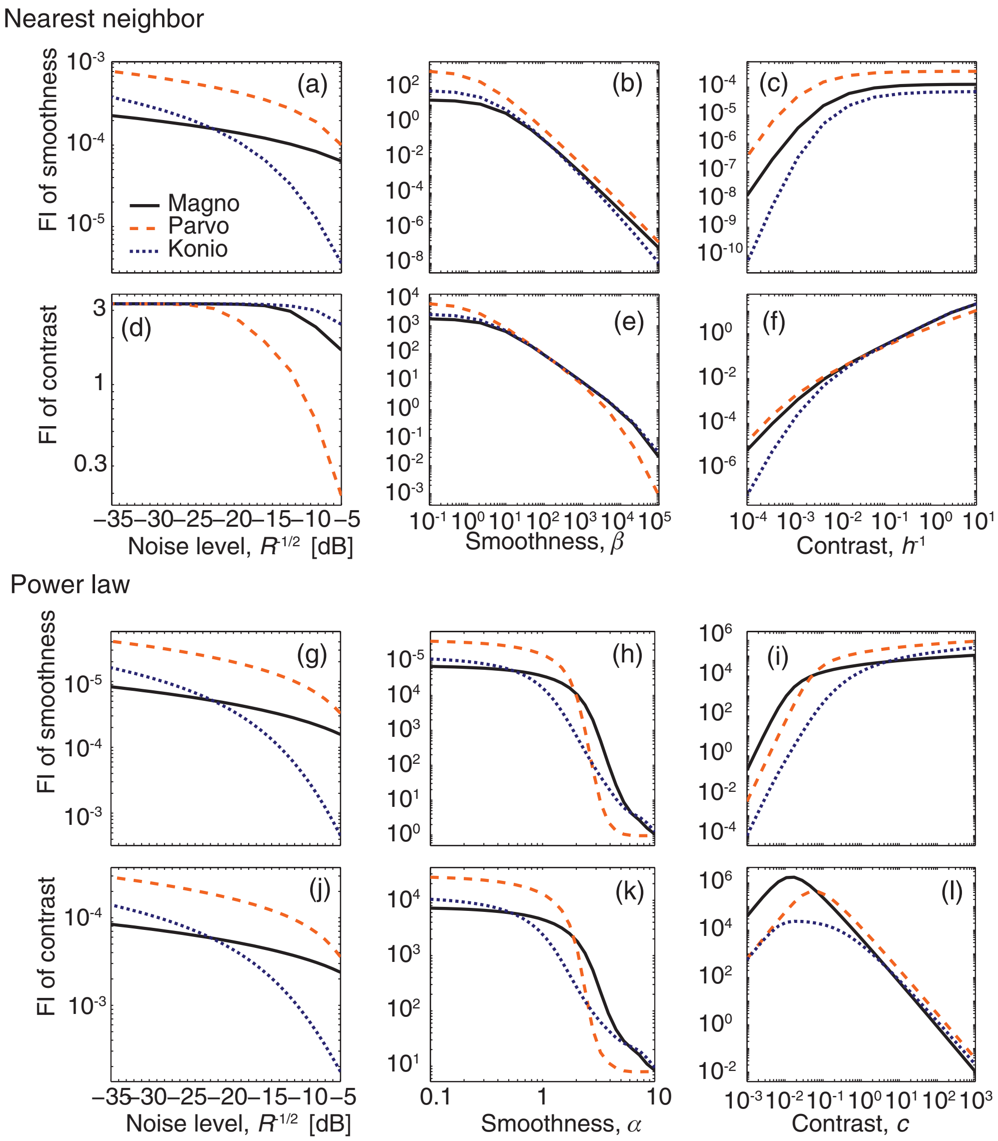

FIG. 4. (Color online) Efficiency of information transmission in the retinogeniculate visual pathways. Three simulated visual pathways are shown: the magnocellular, the parvocellular, and the koniocellular systems. For quantification, Fisher information measures of smoothness (a), (b), (e), (f) and contrast (c), (d), (g), (h) statistics were used. (a)–(d) The nearest-neighbor model. (a), (c) β was varied with fixed $h = 1$; (b), (d) h was varied with fixed $\beta = 3000$. (e)–(h) The power-law model. (e), (g) α was varied with fixed $c = 1$; (f), (h) c was varied with fixed $\alpha = 1$.

all of these three types of nonlinearity, the net output behavior (average and noise variance) of these push-pull circuits is approximated well by the linear + noise model [Figs. 5(c), 5(d), and 5(f)]. This observation supports the idea that the current linear model, at least qualitatively, approximates the information representation by the pairs of on-off cell responses. Nevertheless, it should be noted that the more detailed structure of noise distribution in the last model [Fig. 5(f)] is not exactly identical to the linear + noise model; investigating how explicit modeling of response nonlinearity (after noise addition, in particular) affects the quantitative performance of information transmission is an important issue of future study.

Another simplification in the present model concerns the dynamics in contrast adaptation. The neural encoding in the retina is known to show dynamic changes according to the spatiotemporal context in visual stimuli [51–54]. Although such a dynamic property of encoding is an interesting research subject, it was omitted in the present study as in many other studies, because this simplification was not expected to cause a drastic effect in the current problem setting. First, the present study focuses on the encoding of a single frame of a static natural image. With this assumption, the dynamics of adaptation to the image contrast, which typically has a relatively large time constant on the order of several or decades of seconds, can be ignored. Second, the present study does not assume particular algorithms of decoding, but considers an upper bound of information transmission accuracy by assuming an ideal observer that decodes hyperparameters according to a maximum likelihood estimation. It is worth noting that the adaptation is problematic when one assumes a fixed decoder because it leads to ambiguity for the observer

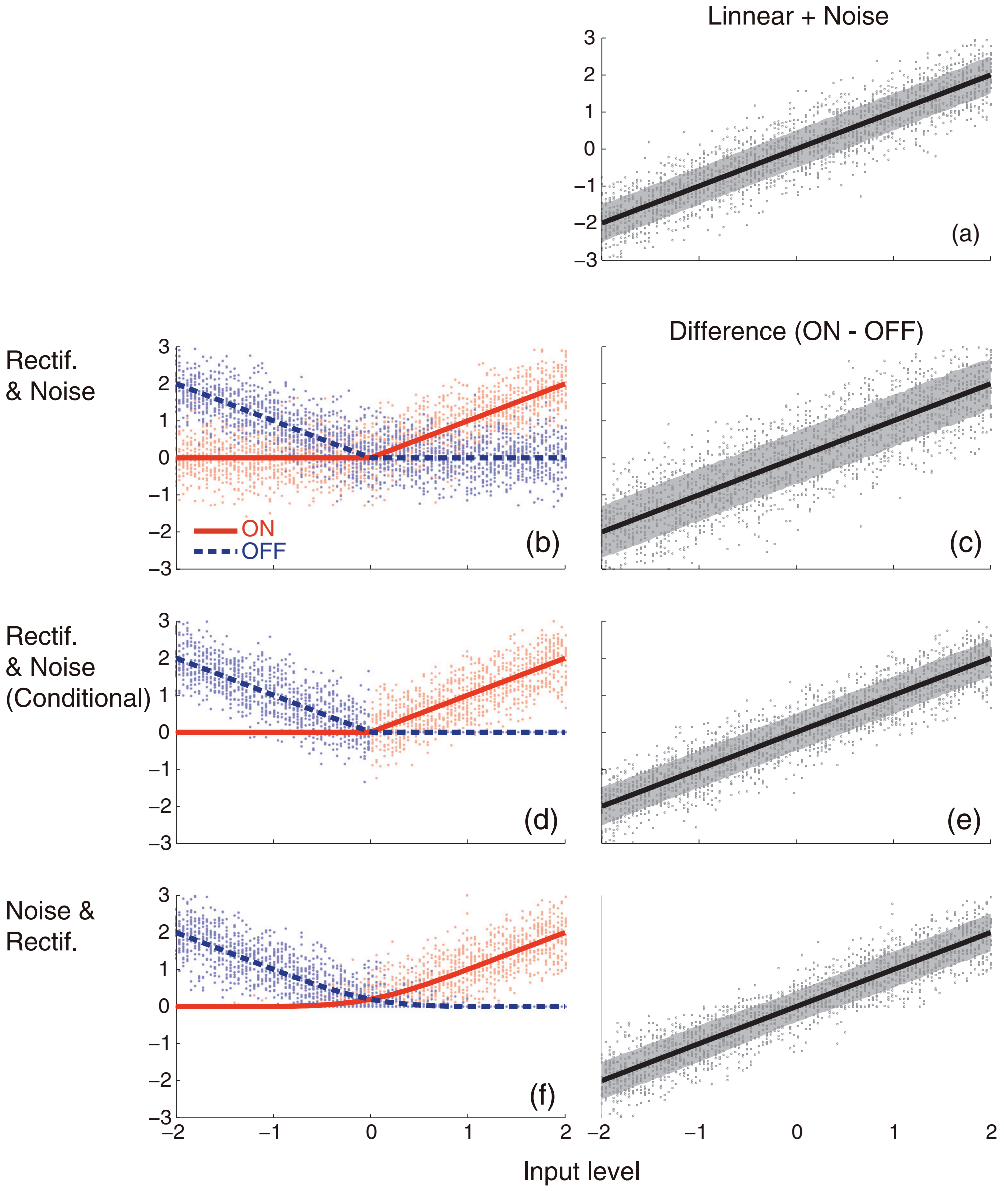

FIG. 5. (Color online) Relevance to nonlinear models. Three typical nonlinear response models are considered: (b), (c) (half-wave) rectification nonlinearity that is followed by additive noise, (d), (e) rectification before conditional noise, which is added only for the input regime eliciting positive average responses [47], and (f), (g) rectification after the noise perturbation process, which yields average responses that are similar to power functions [48–50]. (b), (d), (f) Simulations of nonlinear responses of on and off cells. (c), (e), (g) The differences of on and off responses were computed for each model. The differential responses (on-off) show similar behaviors compared to the linear + noise model (a), which is considered in the present study, as in a push-pull circuit. The solid or dashed lines represent the average outputs. The shaded areas indicate intervals of ±1standard deviation that were derived from 1000 trials of the simulation for each input value. The dots represent trial-to-trial responses (showing 20 trials for each of the input values that were uniformly sampled between −2 and 2). The noise had a Gaussian distribution with a standard deviation of 0.5.

without knowledge of the changes in the encoding process. However, the present study assumes an ideal observer, which provides a complete probabilistic model that relates the stimulus to the encoder output. This indicates that the present study provides a theoretical upper bound of the information transmission performance, whether or not the encoder is in the adaptation state.

In the analysis of the encoding efficiency of natural image statistics, we found that the two conceptually similar natural-image models yielded qualitatively different predictions. It is not likely that such a dependency is only an artifact resulting from the simplification of the model. Considering that the current linear model approximates well the combined outputs of on and off cells with threshold nonlinearity as described above, it is not probable that the dependency in the natural-image model would vanish if more complex nonlinear models were substituted for the linear model. For these reasons, the finding that the method of modeling natural images can qualitatively affect the model predictions should be stated even though the current model is a simplified one. Nevertheless, it should be noted that the quantitative behavior of the model might change if we introduce a more complex model.

In the present study, we focused on the second-order statistics that are related to the local light intensity (pixel luminance). Recent studies have also suggested the human ability to discriminate higher-order local statistics that are defined by the image intensity distribution [55–57] or by filter outputs [58]. Extending the current theoretical framework to those higher-order statistics is an important future issue. We expect that the empirical Bayesian approach that is described in the present study will provide a useful basis for the assessment of higher-order statistical perceptions.

Although there have been a number of previous attempts to relate cortical hierarchy to the hierarchical Bayesian method [17,19–22], few studies have discussed the perceptions of natural statistics in terms of empirical Bayesian inference. In most previous studies,the hyperparameters correspond to the precision (or uncertainty) of predicted estimates in Kalman filtering, which are implemented in neural connectivity [23,59] or activation level [60]. However, it is unclear how the nervous system represents stimulus statistics. Considering that the stimulus statistics have to be dynamically inferred for each stimulus in a setting as in the present study, they are likely to be encoded by the activities of specific neurons in the higher visual area that receives information from a broad visual field, or represented by a population response in early visual neurons [61]. In addition, the gradient of free energy is expressed using the estimated stimulus $\hat{s}$ [32], and the minimum free energy is obtained through an iterative method that alternates between the optimization of $\hat{s}$ and $\hat{\boldsymbol{\theta}}$, which is similar to the expectation-maximization algorithm [62]. This iterative algorithm can be implemented by reciprocal connections between cortical areas.

Finally, in light of the empirical Bayesian interpretation, the estimation of stimulus statistics is found to have at least three ecological functions: (1) to provide an efficient, compressed representation of the external world, which on its own can be utilized as advantageous information for the survival of the animal; (2) to increase the decoding accuracy of neuronal signals by flexibly adapting the prior distribution depending on the input stimulus; and (3) to predict future stimuli according to the generative model of the stimulus, particularly when the statistics have temporal structures (such as the hidden Markov model). Of particular interest, the second and third functions have not been the target of thorough discussions in previous studies (note that there have been recent studies that have linked empirical Bayes and predictive actions; e.g., see [63] for a free-energy interpretation of predictive uncertainty in attentional tasks). We believe that interpreting the perceptual organization of stimulus statistics with a Bayesian framework provides a useful viewpoint for future investigations into the functions of the sensory system and perception.

ACKNOWLEDGMENT

We deeply thank the two anonymous reviewers for their helpful comments on the original version of the manuscript.

[1] D. L. Ruderman and W. Bialek, Phys. Rev. Lett. **73**, 814 (1994).

[2] E. P. Simoncelli and B. A. Olshausen, Annu. Rev. Neurosci. **24**, 1193 (2001).

[3] A. Srivastava, A. Lee, E. Simoncelli, and S. Zhu, J. Math. Imaging Vision **18**, 17 (2003).

[4] K. Tanaka, J. Phys. A **35**, R81 (2002).

[5] H. Nishimori and K. Y. Michael Wong, Phys. Rev. E **60**, 132 (1999).

[6] J.-i. Inoue and D. Carlucci, Phys. Rev. E **64**, 036121 (2001).

[7] R. Liu, Z. Li, and J. Jia, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2008), p. 1.

[8] B. C. Hansen, E. A. Essock, Y. Zheng, and J. K. DeFord, Netw., Comput. Neural Syst. **14**, 501 (2003).

[9] A. Torralba and A. Oliva, Netw., Comput. Neural Syst. **14**, 391 (2003).

[10] D. C. Knill, D. Field, and D. Kersten, J. Opt. Soc. Am. A **7**, 1113 (1990).

[11] D. J. Tolhurst and Y. Tadmor, Perception **26**, 1011 (1997).

[12] C. A. Párraga and D. J. Tolhurst, Perception **29**, 1101 (2000).

[13] B. C. Hansen and R. F. Hess, J. Vision **6**, 696 (2006).

[14] J. J. Atick and A. N. Redlich, Neural Comput. **4**, 196 (1992).

[15] B. A. Olshausen and D. J. Field, Nature (London) **381**, 607 (1996).

[16] A. J. Bell and T. J. Sejnowski, Vision Res. **37**, 3327 (1997).

[17] R. P. N. Rao and D. Ballard, Nat. Neurosci. **2**, 79 (1999).

[18] O. Schwartz and E. P. Simoncelli, Nat. Neurosci. **4**, 819 (2001).

[19] M. Kawato, H. Hayakawa, and T. Inui, Netw., Comput. Neural Syst. **4**, 415 (1993).

[20] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel, Neural Comput. **7**, 889 (1995).

[21] T. S. Lee and D. Mumford, J. Opt. Soc. Am. A **20**, 1434 (2003).

[22] K. J. Friston, PLoS Comput. Biol. **4**, e1000211 (2008).

[23] R. P. N. Rao and D. H. Ballard, Neural Comput. **9**, 721 (1997).

[24] R. P. N. Rao,' in *Advances in Neural Information Processing Systems*, edited by L. Saul, Y. Weiss, and L. Bottou (MIT Press, Cambridge, MA, 2005), Vol. 17, pp. 1113–1120.

[25] Y. Iba, J. Phys. A **32**, 3875 (1999).

[26] D. J. C. MacKay, Neural Comput. **4**, 720 (1992).

[27] I. J. Good, *The Estimation of Probabilities* (MIT Press, Cambridge, MA, 1965).

[28] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. (Springer-Verlag, New York, 1985).

[29] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*, Lecture Notes in Statistics No. 116 (Springer-Verlag, New York, 1996).

[30] P. A. Devijver and M. M. Dekesel, in *Pattern Recognition Theory and Applications*, edited by P. A. Devijver and J. Kittler, NATO Advanced Studies Institute, Series F: Computer and Systems Sciences (Springer-Verlag, Berlin, 1987), Vol. 30.

[31] K. Friston, Nat. Rev. Neurosci. **11**, 127 (2010).

[32] S. Tajima, M. Inoue, and M. Okada, J. Phys. Soc. Jpn. **77**, 054803 (2008).

[33] S. W. Kuffler, J. Neurophysiol. **16**, 37 (1953).
[34] H. Barlow, J. Physiol. **119**, 69 (1953).
[35] R. H. Masland, Nat. Neurosci. **4**, 877 (2001).
[36] G. G. D. Field, J. J. L. Gauthier, A. Sher, M. Greschner, T. A. T. A. Machado, L. L. H. Jepson, J. Shlens, D. D. E. Gunning, K. Mathieson, W. Dabrowski, L. Paninski, A. M. Litke, and E. J. Chichilnisky, Nature (London) **467**, 673 (2010).
[37] S. Geman and D. Geman, IEEE Trans. Pattern Anal. Machine Intell. **6**, 721 (1984).
[38] J. Besag, J. R. Stat. Soc. B **48**, 259 (1986).
[39] G. J. Burton and I. R. Moorhead, Appl. Opt. **26**, 157 (1987).
[40] D. J. Field, J. Opt. Soc. Am. A **4**, 2379 (1987).
[41] D. J. Tolhurst, Y. Tadmor, and T. Chao, Ophthalm. Physiol. Opt. **12**, 229 (1992).
[42] A. van der Schaaf and J. H. van Hateren, Vision Res. **36**, 2759 (1996).
[43] V. Billock, Physica D **137**, 379 (2000).
[44] G. J. Stephens, T. Mora, G. Tkačik, and W. Bialek, Phys. Rev. Lett. **110**, 018701 (2013).
[45] L. C. Sincich, Y. Zhang, P. Tiruveedhula, J. C. Horton, and A. Roorda, Nat. Neurosci. **12**, 967 (2009).
[46] E. Doi, D. C. Balcan, and M. S. Lewicki, IEEE Trans. Image Process. **16**, 442 (2007).
[47] S. Wu, H. Nakahara, and S. Amari, Neural Comput. **13**, 775 (2001).
[48] J. S. Anderson, I. Lampl, D. C. Gillespie, and D. Ferster, Science **290**, 1968 (2000).
[49] K. D. Miller and T. W. Troyer, J. Neurophysiol. **87**, 653 (2002).
[50] D. Hansel and C. van Vreeswijk, J. Neurosci. **22**, 5118 (2002).
[51] S. M. Smirnakis, M. J. Berry, D. K. Warland, W. Bialek, and M. Meister, Nature (London) **386**, 69 (1997).
[52] D. Chander and E. J. Chichilnisky, J. Neurosci. **21**, 9904 (2001).
[53] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. R. de Ruyter Van Steveninck, Nature (London) **412**, 787 (2001).
[54] T. Hosoya, S. A. Baccus, and M. Meister, Nature (London) **436**, 71 (2005).
[55] J. D. Victor and M. M. Conte, Vision Res. **44**, 541 (2004).
[56] G. Tkacik, J. S. Prentice, J. D. Victor, and V. Balasubramanian, Proc. Natl. Acad. Sci. USA **107**, 18149 (2010).
[57] J. D. Victor and M. M. Conte, J. Opt. Soc. Am. A **29**, 1313 (2012).
[58] H. E. Gerhard, F. a. Wichmann, and M. Bethge, PLoS Comput. Biol. **9**, e1002873 (2013).
[59] S. Denève, J.-R. Duhamel, and A. Pouget, J. Neurosci. **27**, 5744 (2007).
[60] J. M. Beck, P. E. Latham, and A. Pouget, J. Neurosci. **31**, 15310 (2011).
[61] S. Tajima and M. Okada, PLoS One **5**, e9704 (2010).
[62] A. Dempster, N. Laird, and D. Rubin, J. R. Stat. Soc. B **39**, 1 (1977).
[63] H. Feldman and K. J. Friston, Frontiers Human Neurosci. **4**, 215 (2010).