INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **14** (2003) 1–18 PII: S0954-898X(03)54125-2

Sensory adaptation as Kalman filtering: theory and illustration with contrast adaptation

Norberto M Grzywacz[1] **and Joaquín de Juan**[2]

[1] Department of Biomedical Engineering and Neuroscience Graduate Program, University of Southern California, University Park, OHE 500, 3650 S McClintock Avenue, Los Angeles, CA 90089-1451, USA
[2] Departamento de Biotecnología, Facultad de Ciencias, Universidad de Alicante, Apartado de Correos 99, 03080 Alicante, Spain

E-mail: nmg@bmsr.usc.edu and jdj@ua.es

Received 1 October 2002, in final form 29 March 2003
Published
Online at stacks.iop.org/Network/14/1

Abstract

Sensory adaptation allows biological systems to adjust to variations in the environment. A recent theoretical work postulated that the goal of adaptation is to minimize errors in the performance of particular tasks. The proposed minimization was Bayesian and required prior knowledge of the environment and of the limitations of the mechanisms processing the information. One problem with that formulation is that the environment changes in time and the theory did not specify how to know what the current state of the environment is. Here, we extend that theory to estimate optimally the environmental state from the temporal stream of responses. We show that such optimal estimation is a generalized form of Kalman filtering. An application of this new Kalman-filtering framework is worked out for retinal contrast adaptation. It is shown that this application can account for surprising features of the data. For example, it accounts for the differences in responses to increases and decreases of mean contrasts in the environment. In addition, it accounts for the two-phase decay of contrast gain when the mean contrast in the environment rises suddenly. The success of this and related theories suggest that sensory adaptation is a form of constrained biological optimization.

1. Introduction

Adaptation is an important property of biological sensory systems, as it allows them to adjust to variations in the environment (Thorson and Biederman-Thorson 1974, Laughlin 1989). Several investigators have suggested that this adjustment optimizes the systems in the performance of particular tasks. Some tasks are generic, such as sending as much information as possible to the rest of the brain (Srinivasan *et al* 1982, Atick and Redlich 1992, Field 1994). Other tasks are

0954-898X/03/000001+18$30.00

Ascii/Word

NET/net154125-xsl/SPE
Printed 26/5/2003

Focal Image
(*Ed: STUART*)

CRC data

File name NE .TEX
Date req.
Issue no.
Artnum

First page
Last page
Total pages
Cover date

more specific, such as optimal localization of edges in a natural image and the determination of their contrast (Balboa and Grzywacz 2000a, 2000b, Grzywacz and Balboa 2002). Regardless of the task, all these theories need knowledge about the statistics of the natural environment. One often refers to these statistics as the prior probabilities, borrowing terminology from Bayesian theory (Berger 1985).

The trouble is that adaptation is only relevant if the environment can have several states. In other words, the statistics of the environment can be different at two or more instants in time. For example, image statistics during the day can be different from image statistics at dusk or at night. As a minimum, the amount of photons reaching the retina is lower at night than during the day, leading to noisier images at night (Rushton 1961, Fuortes and Yeandle 1964, Baylor *et al* 1979). In addition, the colour spectrum of natural images is different (redder) at dusk than at midday. Another important example is that changes of the environment can be caused actively by the organism itself rather than by external changes. For instance, an animal entering a cave will immediately face a darker environment to which it must adapt.

If the environment has several states, then the sensory system must have information about the statistics of each of these states. Furthermore, the system would have to know how to select at each moment which state is the correct one. Theoretically, this is similar to the Bayesian process of prior selection (Berger 1985). In prior selection, one chooses the correct prior distribution to use before performing a task. One denotes the various possible priors by special hyperparameters. The complication is that prior selection by a sensory system must be based on its responses to the environment. These responses have much noise and may use internal parameters that assume the incorrect environment. For instance, consider the example of an animal running into a cave. In the beginning, the animal's retinal parameters are still set to the outside world. The retina has no way to know at first whether the new darker images are just dark objects from the outside environment or whether the environment has changed. It is only after sampling the environment for a while that the retina changes its parameters. Curiously, though, at first the retina samples the dark environment with parameters for the bright environment.

In this paper, we propose a theoretical framework for how sensory systems perform prior selection. This framework starts from our previous work on a Bayesian theory of sensory adaptation (Grzywacz and Balboa 2002). We add to that theory the temporal sampling of the environment and, upon developing the equations, note that the result resembles Kalman filtering (Kalman 1960). The result is tightly connected to a Bayesian generalization of standard Kalman filters (Ho and Lee 1964). Similar generalizations appeared before in motion tracking in artificial vision (Isard and Blake 1996), in sound localization in artificial audition (Datum *et al* 1996), in pruning of artificial neural networks (Sum *et al* 1999) and in modelling of biological vision (Rao and Ballard 1997, Rao 1999, Burgi *et al* 2000).

Finally, we use examples from retinal research to provide mechanistic examples of the elements of the new theoretical framework.

2. Theory

In this section, we give a physical description of the theory (section 2.1), present its general equations (section 2.2) and simplify these equations through biologically reasonable assumptions (section 2.3).

2.1. Physical description of the theory

Grzywacz and Balboa (2002) described how a sensory system could adapt optimally to perform particular tasks if it knew the statistics of the environment. Their model consisted of the four left

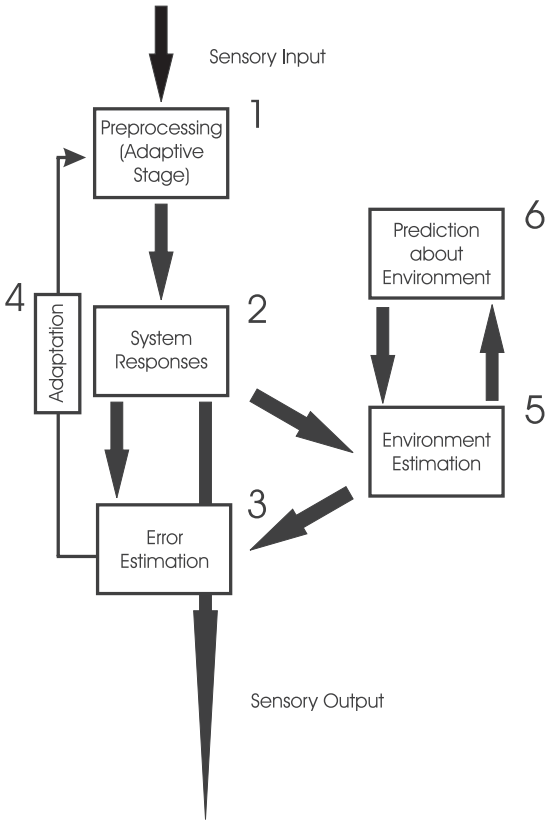

Figure 1. Schematics of the new Kalman theory of sensory adaptation. The sensory input is preprocessed by an adaptive stage (box 1), whose output is further processed to give rise to the system responses (box 2). These responses are both the output of the system and contribute to an estimate of the state of the environment (box 5). This estimate compares these responses with predictions about the state of the environment based on past responses (box 6). The new estimate of the environment then aids an error-computing stage (box 3) to estimate an adaptation signal (box 4). This signal minimizes the error of the contribution of these responses to the performance of pre-specified tasks.

boxes in figure 1. In this model, the sensory input is processed by an adaptive stage (box 1), whose output is further processed to yield the system responses (box 2). The system then estimates with how much error these responses code important attributes from the environment (box 3). Next, the system adapts the preprocessing stage such as to minimize this error (box 4). To estimate the error, the system needs to have knowledge about the environment. The trouble that we that address in this paper is that, as the environment changes, its parameters must be estimated for the system to know what knowledge to apply. We propose that the system estimates them from its own responses (box 5) and from predictions that it makes about the environment based on past responses (box 6). If current responses are statistically consistent with these predictions, then the estimates do not change. Otherwise, they change, but do so slowly to take into account the tendency of the environment to remain stable.

To facilitate the reading of the next theoretical section, we now link its equations to the above physical description of the new adaptation theory. The theoretical results of Grzywacz and Balboa (2002) are summarized by equations (1)–(3) below. Next, we parametrize the

environment and then consider the probability that we have the correct parameters, given the recent past responses of the system (equation (4)). This probability can be broken up into two terms. The first term tells us how much to trust the new responses (equations (4), (10) and (11)). The second term tells us how much to trust the predictions from past responses (equations (4) and (6)). If the environment changes, then trust in the past responses keeps the estimated parameters of the environment unchanged for a while. However, if the responses continue to tell us that something is wrong with the estimated environment, then the parameters of the environment change towards the most probable values (equation (5)).

2.2. General equations

We use modifications of the notations introduced by Grzywacz and Balboa (2002) and Burgi *et al* (2000). Vectors are denoted by bold symbols, time t_k is denoted by subscript k, and time series from t_0 on are denoted by an inverted hat accent. In other words, we define $\check{\boldsymbol{z}}_k = \{\boldsymbol{z}_k, \boldsymbol{z}_{k-1}, \ldots, \boldsymbol{z}_0\}$. Let $\boldsymbol{I}$ be the input to the sensory system and $\boldsymbol{H}^1(\boldsymbol{I}), \boldsymbol{H}^2(\boldsymbol{I}), \ldots, \boldsymbol{H}^N(\boldsymbol{I})$ the N relevant task-input attributes to extract from $\boldsymbol{I}$. For instance, these attributes could be things like contrasts and positions of occluding borders in a visual image (for examples, see section 3 and Grzywacz and Balboa 2002). The input is first preprocessed by an adaptive stage (box 1 in figure 1), whose output $\boldsymbol{O}$ is transformed by the task-coding stage (box 2) to $\boldsymbol{R}^{H1}(\boldsymbol{O}), \boldsymbol{R}^{H2}(\boldsymbol{O}), \ldots, \boldsymbol{R}^{HN}(\boldsymbol{O})$. These functions are estimates of the values of the task-input attributes and one can interpret them physiologically as the responses of the system. The system tries to make the error in each of these estimates small. This error is the discrepancy between $\boldsymbol{H}^i(\boldsymbol{I})$ and $\boldsymbol{R}^{Hi}(\boldsymbol{O})$. The system cannot know exactly what this discrepancy is, since it does not have access to the input, only to its estimates. However, the system can estimate the expected amount of error (box 3). Grzywacz and Balboa (2002) showed that the mean Bayesian expected loss over all possible responses is

$$E_i(\boldsymbol{A}, \boldsymbol{\Lambda}) = \int_{\boldsymbol{I}} \int_{\boldsymbol{R}^{Hi}} P(\boldsymbol{R}^{Hi}|\boldsymbol{I}, \boldsymbol{A}) P(\boldsymbol{I}|\boldsymbol{\Lambda}) L(\boldsymbol{I}, \boldsymbol{R}^{Hi} : \boldsymbol{\Lambda}, \boldsymbol{A}), \tag{1}$$

where $\boldsymbol{A}$ are the adaptation parameters of the adaptive stage, $\boldsymbol{\Lambda}$ are the hyperparameters of the prior distribution that the system selected and L is the loss function[3], which measures the cost of deciding that the ith attribute is $\boldsymbol{R}^{Hi}$ given that the input is $\boldsymbol{I}$ (for examples, see section 3 and Grzywacz and Balboa 2002). One can give intuitive and practical interpretations of the probability terms in this equation. The first term, the likelihood function $P(\boldsymbol{R}^{Hi}|\boldsymbol{I}, \boldsymbol{A})$, embodies the knowledge about the sensory mechanisms. In other words, $P(\boldsymbol{R}^{Hi}|\boldsymbol{I}, \boldsymbol{A})$ tells us how the system responds when the stimulus is $\boldsymbol{I}$ and the adaptation parameters are $\boldsymbol{A}$. Such probabilistic dependence on $\boldsymbol{I}$ and $\boldsymbol{A}$ is already present in the $\boldsymbol{O}$ variables, which then pass it to $\boldsymbol{R}^{Hi}$. The second term, $P(\boldsymbol{I}|\boldsymbol{\Lambda})$, is the prior distribution of the input when the environment is $\boldsymbol{\Lambda}$.

Equation (1) has two differences from those typically based in standard Bayesian decision theory (Berger 1985; see the development of the equation in Grzywacz and Balboa (2002)). First, in a typical theory, one considers a fixed environment (no hyperparameters) and a fixed system (no adaptation parameters). Second, one does not average over responses (second integral). Rather, given an individual response ($\boldsymbol{R}^{Hi}$), one chooses an action (in our case, the estimation of $\boldsymbol{H}^i(\boldsymbol{I})$) and for it, calculates the mean error (the Bayesian expected loss) over all possible inputs ($\boldsymbol{I}$). One then finds the action that minimizes the error. In equation (1),

[3] The loss function is parametric on both $\boldsymbol{\Lambda}$ and $\boldsymbol{A}$. The dependence on the latter is due to the dependence of the input–output relationship on the adaptation state of the system. In turn, the dependence on $\boldsymbol{\Lambda}$ indicates that the error of certain responses may depend on the state of nature. For instance, at night, certain errors may be more permissible than at day.

the action is $\boldsymbol{R}^{Hi}$ itself, as this is the estimation of $\boldsymbol{H}^i(\boldsymbol{I})$. Moreover, the purpose of our theory is to find the best adaptation parameters. In other words, we must find those parameters that lead to the best decision over all possible responses, not just an individual response. Consequently, one must average the errors over the responses, which is the purpose of the second integral. Because of it, the error is not a function of the action $\boldsymbol{R}^{Hi}$ but of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$, the parameters of the model. The mean Bayesian expected loss expressed in equation (1) can thus be interpreted as the mean error the system makes if its adaptation parameters are $\boldsymbol{A}$ and the environment is in state $\boldsymbol{\Lambda}$. Because the goal is to estimate N attributes, the total error is

$$E(\boldsymbol{A}, \boldsymbol{\Lambda}) = \sum_{i=1}^{N} E_i(\boldsymbol{A}, \boldsymbol{\Lambda}). \tag{2}$$

A way to find $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ would be to find their values that minimize $E(\boldsymbol{A}, \boldsymbol{\Lambda})$. However, this way would fail to recognize the knowledge that most of the time the environment does not change or it tends to change slowly (section 1). In other words, if we know what $\boldsymbol{\Lambda}$ was a short while ago, then it is highly likely that $\boldsymbol{\Lambda}$ did not change. Hence, we propose that the system estimates $\boldsymbol{\Lambda}$ from past responses (box 5 in figure 1). With $\boldsymbol{\Lambda}$ given, then we can estimate $\boldsymbol{A}$ (box 4) as

$$\boldsymbol{A}(\boldsymbol{\Lambda}) = \arg\min_{A^*} E(\boldsymbol{A}^*, \boldsymbol{\Lambda}). \tag{3}$$

In other words, $\boldsymbol{A}$ is the adaptation argument that minimizes the error for a fixed value of $\boldsymbol{\Lambda}$.

The goal is thus to estimate $\boldsymbol{\Lambda}$ from past responses. Moreover, responses are related to the environment through the adaptation state of the system. Therefore, one must in general take into account the past adaptation states to estimate $\boldsymbol{\Lambda}$. The most general way to relate $\boldsymbol{\Lambda}$ to past responses and adaptation states is to calculate $P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}_k^{Hi}, \breve{\boldsymbol{A}}_k)$. Using Bayes' rule and a few algebraic manipulations (Ho and Lee 1964, Burgi *et al* 2000), we get

$$P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}_k^{Hi}, \breve{\boldsymbol{A}}_k) = \frac{P(\boldsymbol{R}_k^{Hi}, \boldsymbol{A}_k|\boldsymbol{\Lambda}_k, \breve{\boldsymbol{R}}_{k-1}^{Hi}, \breve{\boldsymbol{A}}_{k-1})}{P(\boldsymbol{R}_k^{Hi}, \boldsymbol{A}_k|\breve{\boldsymbol{R}}_{k-1}^{Hi}, \breve{\boldsymbol{A}}_{k-1})} P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}_{k-1}^{Hi}, \breve{\boldsymbol{A}}_{k-1}). \tag{4}$$

One way to get $\boldsymbol{\Lambda}_k$ from this equation is to set it from maximum *a posteriori* estimation:

$$\boldsymbol{\Lambda}_k(\breve{\boldsymbol{R}}_k^{Hi}, \breve{\boldsymbol{A}}_k) = \arg\max_{\boldsymbol{\Lambda}_k^*} P(\boldsymbol{\Lambda}_k^*|\breve{\boldsymbol{R}}_k^{Hi}, \breve{\boldsymbol{A}}_k). \tag{5}$$

However, other estimations using different loss functions are possible (Berger 1985).

Equation (4) can be understood as a form of Kalman filtering. The rightmost term represents the *prediction stage*, in which we find the likely values of the current environment as predicted by the past measurements (box 6 in figure 1). In turn, the numerator in the division represents the *measurement stage*. The new measurements are combined using Bayes' theorem (arrow from box 2 to 5) and, if consistent, reinforce the prediction (arrow from box 6 to 5) and decrease the uncertainty about the current environment; inconsistent measurements may increase the uncertainty further. That $\boldsymbol{A}$ appears as an argument of the probability function of the measurement stage is a mathematical consequence of Bayes' theorem. However, we will argue in section 2.3 that $\boldsymbol{A}$ is not a random variable, simplifying equation (4). (Finally, the denominator is just a normalizing term.)

The measurement stage of equation (1) relates the environment to the responses[4]. Consequently, for a sensory system, this stage depends on the underlying biological mechanisms. Later in the paper, we will discuss how to go from these mechanisms to the numerator of equation (1). In turn, the prediction stage of equation (1) must somehow take

[4] The probability function describing the measurement stage is thus a second likelihood function, the first having appeared in equation (1). However, as we will see in section 2.3, these two likelihood functions are tightly related.

into account that environments are often stable or do not change rapidly. The predictive distribution function can be expressed as (Ho and Lee 1964, Burgi *et al* 2000)

$$P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}^{Hi}_{k-1}, \breve{\boldsymbol{A}}_{k-1}) = \int_{\boldsymbol{\Lambda}_{k-1}} P(\boldsymbol{\Lambda}_k|\boldsymbol{\Lambda}_{k-1}) P(\boldsymbol{\Lambda}_{k-1}|\breve{\boldsymbol{R}}^{Hi}_{k-1}, \breve{\boldsymbol{A}}_{k-1}). \tag{6}$$

The importance of this equation is that its rightmost term is identical[5] to the leftmost term in equation (4), except for being one step back in time. Hence, if one knows how to model the measurement stage of equation (4) and one knows the first term of the integral of equation (6), one can solve both equations recursively. The first term of the integral of equation (6) is the knowledge of how slowly the environment changes. It is, therefore, a second prior term necessary in adapting systems. The first prior term, $P(\boldsymbol{I}|\boldsymbol{\Lambda})$, appeared in equation (1). Incidentally, as we now treat the environment as a function of time, it is necessary to rewrite equations (1), (2) and (3), respectively, as

$$E_i(\boldsymbol{A}_k, \boldsymbol{\Lambda}_k) = \int_{\boldsymbol{I}} \int_{\boldsymbol{R}^{Hi}} P(\boldsymbol{R}^{Hi}|\boldsymbol{I}, \boldsymbol{A}_k) P(\boldsymbol{I}|\boldsymbol{\Lambda}_k) L(\boldsymbol{I}, \boldsymbol{R}^{Hi} : \boldsymbol{\Lambda}_k, \boldsymbol{A}_k), \tag{7}$$

$$E(\boldsymbol{A}_k, \boldsymbol{\Lambda}_k) = \sum_{i=1}^{N} E_i(\boldsymbol{A}_k, \boldsymbol{\Lambda}_k), \tag{8}$$

and

$$\boldsymbol{A}_k(\boldsymbol{\Lambda}_k) = \arg\min_{\boldsymbol{A}^*_k} E(\boldsymbol{A}^*_k, \boldsymbol{\Lambda}_k). \tag{9}$$

2.3. Simplification of the measurement stage

The measurement stage is generally complex, as indicated by its probability function in equation (4), namely $P(\boldsymbol{R}^{Hi}_k, \boldsymbol{A}_k|\boldsymbol{\Lambda}_k, \breve{\boldsymbol{R}}^{Hi}_{k-1}, \breve{\boldsymbol{A}}_{k-1})$. This function indicates that the present responses not only depend on the environment, but also on past responses and past adaptation states. The dependence on past responses can be understood by noting that responses do not decay instantaneously, but have decay time constants. Similarly, particular adaptation states may have slow decays and thus affect current responses. For instance, let us assume that a neuron was in its refractory period 0.5 ms ago. The probability that the refractory period remains is high, affecting the probability of particular neural responses.

Fortunately, one can make biologically plausible simplifications of the measurement stage for sensory adaptation. The first simplification arises from the slowness of adaptation compared to the responses. Because of this slowness, it is reasonable to assume that the probability that $\boldsymbol{A}_k = \boldsymbol{A}_{k-1}$ is close to one. Moreover, in all known biological sensory systems, the response depends on the current adaptation states and not on past ones. Hence, one can write to a good approximation that

$$P(\boldsymbol{R}^{Hi}_k, \boldsymbol{A}_k|\boldsymbol{\Lambda}_k, \breve{\boldsymbol{R}}^{Hi}_{k-1}, \breve{\boldsymbol{A}}_{k-1}) \approx P(\boldsymbol{R}^{Hi}_k|\boldsymbol{\Lambda}_k, \breve{\boldsymbol{R}}^{Hi}_{k-1}, \boldsymbol{A}_k).$$

One can simplify this equation further by assuming that the responses from the past have decayed already at the moment of measurement of the present response. One can make this assumption, as the goal of equation (4) is to compute the probability of the environment, which changes much more slowly compared to the responses. This slowness yields

$$P(\boldsymbol{R}^{Hi}_k, \boldsymbol{A}_k|\boldsymbol{\Lambda}_k, \breve{\boldsymbol{R}}^{Hi}_{k-1}, \breve{\boldsymbol{A}}_{k-1}) \approx P(\boldsymbol{R}^{Hi}_k|\boldsymbol{\Lambda}_k, \boldsymbol{A}_k). \tag{10}$$

[5] This statement assumes that $k \gg 1$, as, strictly speaking, $\breve{\boldsymbol{R}}^{Hi}_{k-1}$ and $\breve{\boldsymbol{A}}_{k-1}$ are each one element shorter than $\breve{\boldsymbol{R}}^{Hi}_k$ and $\breve{\boldsymbol{A}}_k$. If $k \gg 1$, then $P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}^{Hi}_k, \breve{\boldsymbol{A}}_k)$ and $P(\boldsymbol{\Lambda}_k|\breve{\boldsymbol{R}}^{Hi}_k, \breve{\boldsymbol{A}}_k)$ will not depend on the first elements of the time series. Consequently, the fundamental difference between them will be irrelevant.

With these simplifications, one can write a convenient expression for the measurement stage by using the likelihood and prior functions of equation (7) (see discussion after equation (1)), because

$$P(\boldsymbol{R}_k^{Hi}|\boldsymbol{\Lambda}_k, \boldsymbol{A}_k) = \int_{\boldsymbol{I}} P(\boldsymbol{R}_k^{Hi}|\boldsymbol{I}, \boldsymbol{A}_k) P(\boldsymbol{I}|\boldsymbol{\Lambda}_k). \tag{11}$$

Therefore, the likelihood function expressed by the measurement stage in equation (4) is tightly related to the likelihood function expressed in equation (7). By substituting equations (10) and (11) for the measurement stage in equation (4), one gets, to a good approximation,

$$P(\boldsymbol{\Lambda}_k|\check{\boldsymbol{R}}_k^{Hi}, \check{\boldsymbol{A}}_k) = K\, P(\boldsymbol{\Lambda}_k|\check{\boldsymbol{R}}_{k-1}^{Hi}, \check{\boldsymbol{A}}_{k-1}) \int_{\boldsymbol{I}} P(\boldsymbol{R}_k^{Hi}|\boldsymbol{I}, \boldsymbol{A}_k) P(\boldsymbol{I}|\boldsymbol{\Lambda}_k), \tag{12}$$

where K is a normalization constant replacing the denominator in equation (4).

3. Retinal contrast adaptation

We use the phenomenon of retinal contrast adaptation to illustrate how the theory works. It was felt that this phenomenon was a good one for this illustration, since there are good data on the distribution of contrasts in natural images. Furthermore, there is much literature about retinal responses to contrast. In what follows, we describe the basic phenomenology (section 3.1), present the model (section 3.2) and provide simulation methods (section 3.3) and results (section 3.4).

3.1. Experimental data

Retinal ganglion cells adapt to image contrast, that is, the range of light intensities in the image (Smirnakis *et al* 1997). These cells also adapt to spatial correlations within the scene (Smirnakis *et al* 1997). Increasing the contrast of the environment speeds the kinetics and lowers the sensitivity of ganglion cells (Benardete *et al* 1992, Chander and Chichilnisky 2001, Kim and Rieke 2001, Rieke 2001). The time course of these changes is typically of the order of seconds. For instance in salamander, the decline in sensitivity after a contrast increase has two distinct kinetic components, with fast ($<$2 s) and slow ($>$10 s) time constants (Kim and Rieke 2001). However, the rise in sensitivity with falling contrasts has a single, intermediate time constant (4–18 s—Kim and Rieke 2001). These time constants are space-dependent, appearing to rise with the size of the stimulus (Brown and Masland 2001). Such a complex behaviour is possibly due to the existence of multiple mechanisms of contrast adaptation (Smirnakis *et al* 1997). These mechanisms include bipolar-cell processes (Brown and Masland 2001, Rieke 2001), ganglion-cell processes (Kim and Rieke 2001) and different processes for the on and off pathways (Chander and Chichilnisky 2001, Kim and Rieke 2001, Rieke 2001). Finally, it is important to point out that these adaptation mechanisms appear to be present in the human retina (Heinrich and Bach 2001) and to have perceptual correlates (Anstis 1996, DeMarco *et al* 1997, Freeman and Badcock 1999).

3.2. Model

To illustrate the new Kalman theory of sensory adaptation, we propose a simple model for the changes in sensitivity due to contrast adaptation (leaving kinetic changes for a later publication). Before developing the equations, we provide a physical description of the model, using figure 2 for this purpose. We propose that one of the goals of the retina is to transmit maximal amount of information about contrast. (This is related to, but not the same as, the principle of optimal

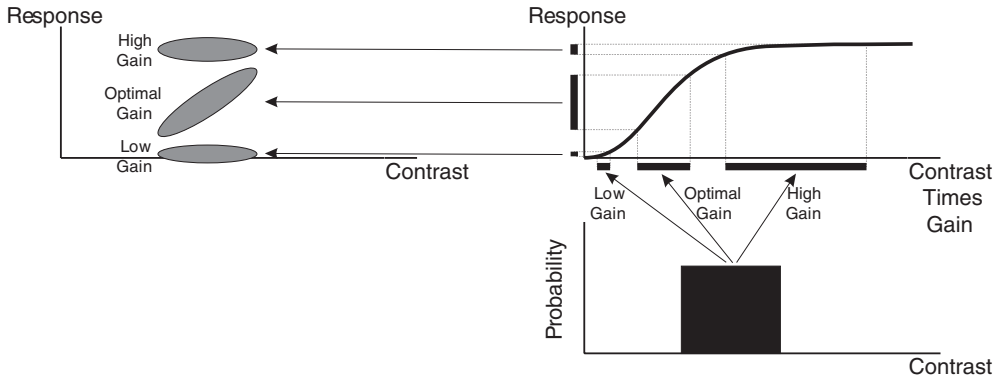

Figure 2. Maximization of contrast information through contrast gain selection. The bottom plot illustrates a hypothetical contrast distribution. This distribution is transformed into different ranges of inputs (horizontal bars) by different contrast gains. In turn, these inputs give rise to different response ranges (vertical bars) after transformation by the input-response curve (sigmoidal line; top right plot). If the gain and thus the inputs are too small or too large, then so is the response range. Only at intermediate values of gain are the response ranges large. After application of noise, small response ranges give rise to relatively uncorrelated contrast-response distributions (horizontal ellipses in left plot). However, a sufficiently large range produces correlations between contrast and response, making their joint distribution informative (tilted ellipses in left plot).

variance estimation used by DeWeese and Zador (1998).) Let the black rectangle in figure 2 represent the range of contrasts in some environment. The model controls the gain from contrast to ganglion-cell response (diagonal arrows in the figure). If the gain is too small, then most responses are essentially zero (plus noise—small lower vertical line). Consequently, the responses do not provide information about contrast to the rest of the brain (lower horizontal ellipse). Similarly, if the gain is too high, then the responses are essentially saturated (plus noise—small upper vertical line). Again, the brain cannot gather information about contrast (top horizontal ellipse). The only hope is that the gain is set correctly such that the distribution of responses due to the distribution of contrasts in the environment has most of its values away from zero and from saturation (large middle vertical line). In this case, the responses are informative about contrast (tilted ellipse). To set the gain correctly, we do two things. First, we use the knowledge that the distribution of contrasts in the environment does not change rapidly in time. Second, we impose that the coding of contrast is maximally informative.

As in section 2.1, to facilitate the reading of the following theoretical text, we now link its equations to the physical description of the model above. The gain to be controlled by the adaptation is defined in the measurement stage (equation (15)). In turn, the distribution of contrasts in the environment (prior distribution) appears in equation (13). Equation (14) imposes the knowledge that the distribution of contrasts in the environment does not change rapidly in time (prediction stage). Finally, the imposition that the coding of contrast is maximally informative (loss function) is in equations (16) and (17).

We now switch to the mathematical formulation of the model. To describe the predictions of the model, including the simplifications introduced in section 2.3, one must solve equations (5)–(9) and (12). These equations call for the specification of three probability functions and a loss function. These functions are $P(\boldsymbol{I}|\boldsymbol{\Lambda}_k)$, $P(\boldsymbol{\Lambda}_k|\boldsymbol{\Lambda}_{k-1})$, $P(\boldsymbol{R}_k^{Hi}|\boldsymbol{I}, \boldsymbol{A}_k)$ and $L(\boldsymbol{I}, \boldsymbol{R}^{Hi} : \boldsymbol{\Lambda}_k, \boldsymbol{A}_k)$.

To specify $P(\boldsymbol{I}|\boldsymbol{\Lambda}_k)$ for the model, we again begin with the approach of Grzywacz and Balboa (2002). For simplicity, they first assumed that, in any given environment, $P(\boldsymbol{I}|\boldsymbol{\Lambda}_k)$ is

constant over the range of naturally occurring images and zero outside it. Then, they used a sample of natural images to estimate the occurrence of the attributes in images. In our case, the attribute of interest is contrast (c). (Because H^i is contrast, we will denote from now on $H^i = H^c$ and $R^{Hi} = R^c$.) Hence, the procedure of Grzywacz and Balboa is similar to having the $\boldsymbol{I}$ being small patches of images from which we extract contrasts. In other words, $P(\boldsymbol{I}|\boldsymbol{\Lambda}_k)$ becomes the distribution of contrasts in the environment. Fortunately, there is much data on the natural distribution of contrasts (Ruderman and Bialek 1994, Zhu and Mumford 1997, Balboa and Grzywacz 2000c). In particular, Balboa and Grzywacz (2000c) pointed out that this distribution is approximately exponential. Therefore, we can approximate the first prior function as

$$P(H^c(\boldsymbol{I}(\boldsymbol{r})) = c|\boldsymbol{\Lambda}_k = c_k^*) = \frac{1}{c_k^*}\mathrm{e}^{c/c_k^*}, \tag{13}$$

where $\boldsymbol{r}$ are retinal positions, c is contrast and c_k^* is the estimated mean contrast in the environment at time t_k. The importance of equation (13) is that it parametrizes the environment by its mean contrast.

With this parametrization, the second prior function ($P(\boldsymbol{\Lambda}_k|\boldsymbol{\Lambda}_{k-1})$) specifies how much the mean contrast varies in time. No data are available on this yet and so we make the simple assumption that this probability function is Gaussian, namely

$$P(\boldsymbol{\Lambda}_k = c_k^*|\boldsymbol{\Lambda}_{k-1} = c_{k-1}^*) = \frac{1}{\sqrt{2\pi}((t_k - t_{k-1})/\tau)}\exp\left(-\frac{(c_k^* - c_{k-1}^*)^2}{2((t_k - t_{k-1})/\tau)^2}\right), \tag{14}$$

where τ is the time constant for how much the mean contrast typically varies during $t_k - t_{k-1}$. Large τ means that a change of mean contrast is unlikely, whereas small τ means the opposite.

Next, we must specify the likelihood function ($P(\boldsymbol{R}_k^c|\boldsymbol{I}, \boldsymbol{A}_k)$) of the measurement stage. The main image attribute of interest here is contrast and thus we look at how responses depend on it. The retina produces responses that tend to be proportional to contrast (Tranchina *et al* 1987). This applies strongly for some ganglion cells (Hochstein and Shapley 1976, de Monasterio 1978, Victor 1987, Lee 1996, Benardete and Kaplan 1997), but even nonlinear cells behave like that for sufficiently low contrasts above a small threshold (Merwine *et al* 1995). At high contrasts, responses invariably saturate in a manner consistent with a Michaelis–Menten (Naka–Rushton) behaviour (Merwine *et al* 1995, Dahari and Spitzer 1996, Tyler and Liu 1996). Accordingly, we assume that responses have this behaviour plus a Gaussian noise, namely

See endnote 1

$$P(\boldsymbol{R}_k^c(\boldsymbol{r}) = R_k^c|H^c(\boldsymbol{I}(\boldsymbol{r})) = c, \boldsymbol{A}_k = \gamma_k) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(R_k^c - (\gamma_k c/(1+\gamma_k c)))^2}{2\sigma^2}\right), \tag{15}$$

where σ is the standard deviation of the noise and γ_k is the contrast gain[6]. Hence, in this model, adaptation controls contrast sensitivity by adjusting the contrast gain of ganglion cells.

Finally, we must specify the loss function $L(\boldsymbol{I}, \boldsymbol{R}^c : \boldsymbol{\Lambda}_k, \boldsymbol{A}_k)$. To produce a maximally informative code for contrast, the loss function should penalize responses, providing little information about contrasts. To analyse what these responses are, we consider the mutual information (Dayan and Abbott 2001) between responses and contrast, that is

$$M(c, R^c : \boldsymbol{\Lambda}_k = c_k^*, \gamma_k) = \sum_c \sum_{R^c} P(c, R^c|c_k^*, \gamma_k)\log\left(\frac{P(c, R^c|c_k^*, \gamma_k)}{P(c|c_k^*, \gamma_k)P(R^c|c_k^*, \gamma_k)}\right). \tag{16}$$

[6] Without loss of generality, this equation sets the saturation response before noise to one. This is possible because the absolute value of the response is scaled in terms of σ. Another peculiarity of the response defined by equation (15) is that it can attain negative values due to noise. This is not a problem if one thinks that the zero is just a background activity.

It is possible to simplify this equation by noting that

$$P(c, R^c|c_k^*, \gamma_k) = P(R^c|c, c_k^*, \gamma_k)P(c|c_k^*, \gamma_k).$$

One can simplify this further in two ways. First, one can drop from it the conditioning on γ_k from the second probability of the right-hand side. This is because contrasts in natural images do not depend on the adaptation state of the system. Second, one can drop the conditioning on c_k^* in the first probability of the right-hand side, since knowing c completely specifies the probability of R^c. Therefore

$$P(c, R^c|c_k^*, \gamma_k) = P(R^c|c, \gamma_k)P(c|c_k^*).$$

By substituting this for the appropriate terms in equation (16), some of the $P(c|c_k^*)$ cancel out to yield

$$M(c, R^c : c_k^*, \gamma_k) = \sum_c \sum_{R^c} P(R^c|c, \gamma_k)P(c|c_k^*) \log\left(\frac{P(R^c|c, \gamma_k)}{P(R^c|c_k^*, \gamma_k)}\right).$$

This equation has the same functional form as equation (7) (using the notation in equations (13) and (15)). Consequently, to create a maximally informative code for contrast all one has to do is to define the loss function $L(\boldsymbol{I}, \boldsymbol{R}^{Hi} : \boldsymbol{\Lambda}_k, \boldsymbol{A}_k)$ as

$$L(H^c(\boldsymbol{I}(\boldsymbol{r})) = c, \boldsymbol{R}^c(\boldsymbol{r}) = R^c : \boldsymbol{\Lambda}_k = c_k^*, \boldsymbol{A}_k = \gamma_k) = -\log\left(\frac{P(R^c|c, \gamma_k)}{P(R^c|c_k^*, \gamma_k)}\right), \tag{17}$$

where the minus sign is to ensure that the loss function is smallest for maximally informative responses. This loss function is the negative of a log-likelihood ratio test (Zar 1984). Our test compares the likelihood of the response under the hypothesis that the contrast is c against the likelihood of the response under the null hypothesis that c_k^* is correct. A loss function based on this test encourages systems whose contrasts different from the mean produce responses significantly different from the responses to the mean. It should not be surprising that the log-likelihood ratio appears in maximally informative systems. Mutual information is related to the log-likelihood ratio test for multinomial distribution (Dayan and Abbott 2001).

3.3. Simulation methods

We simulated the equations of the model with MATLAB (MathWorks, Natick, MA). For all simulations, the possible contrasts, responses and environmental mean contrasts were fixed. The contrasts were from 0.1 to 1 in steps of 0.1, the responses were from -1 to 2 in steps of 0.1 and the mean contrasts were from 0.04 to 0.5 in steps of 0.005. Each simulation consisted of modulation of the mean contrast over 1000 time steps (in intervals of 100 ms). The next section describes the mean contrasts specified in the simulations. The initial conditions of each simulation were specified by preparatory simulations. Those simulations used constant mean contrast in the environment, and assumed homogeneous past probability of mean contrast (last probability function of equation (6)) and that the adaptation gain = 5 (equation (15)). We then used the final results of those simulations to set the probability distribution and the adaptation gain for the next simulation, which always started with the same mean contrast as the preparatory simulation. The parameters τ (equation (14)) and σ (equation (15)) were the only simulation parameters. We varied the former across simulations (see the next section), but kept $\sigma = 0.1$.

Because the possible contrasts and environmental mean contrasts were fixed, we pre-calculated the prior functions in equations (13) and (14) before the simulations. In turn, we calculated the likelihood and loss functions (equations (15) and (17), respectively) on every time step. This is because they depended on current estimates of the mean contrast of the

environment and of the adaptation gain. We estimated first the former through equations (5), (6) and (12), and then estimated the latter through equations (7) and (9). The minimization prescribed by equation (9) was performed with a Nelder–Mead simplex (direct search) method (Nelder and Mead 1965). In turn, the full probability distribution of the left-hand side of equation (12) was necessary for the iterative step in equation (6). Therefore, the maximization procedure in equation (5) simply looked for the mean contrast yielding the maximum in this distribution (without any interpolation). The calculation of the integral in equation (12) required the specification of the current response. Strictly speaking, one should draw this response from the likelihood distribution in a Monte Carlo manner. However, to speed up the convergence of the system, we used the most likely response at every step.

3.4. Results

Exploration of the model revealed much richness in detail, as we will see below. Analysis of these details is beyond the scope of this paper, since its focus is the Kalman theory of adaptation. We will engage in such an analysis elsewhere. Here, we will just analyse one prediction of the model that illustrates the inner workings of the theory. Moreover, this prediction is surprising and consistent with available data, making the model a good candidate to explain retinal contrast adaptation.

If the model is applicable to the retina, then as a minimum, it must account for the decrease in contrast gain as the mean contrast in the environment rises. Conversely, the model must also account for the increase in gain as environmental contrasts fall. Many models can probably account for these data. The sign of the power of a model of contrast adaptation would be in explaining the surprising difference between the effects of contrast rise and fall. The gain reduction with sudden contrast augmentation occurs in two phases, one with fast and another with slow time constants. On the other hand, the gain increase with sudden contrast reduction occurs in one phase with an intermediate time constant. Figure 3 shows that this is exactly what our model predicts.

In this example, the mean contrast of the environment was kept at 0.14 for the first 20 s of the simulation. We then switched the mean contrast up to 0.42 until 60 s, and then back down to 0.14 until 100 s. The simulation of the effects of these contrast modulations used $\tau = 10$ s. Panel A of figure 3 shows that the adaptation gain decreases and increases when the mean contrast rises and falls respectively. An expansion of the timescale of the decay in adaptation gain (panel B) shows that the fall is rapid in the first couple of seconds (see inset). However, this decay slows down for the next 15 s. On the other hand, the increase in contrast gain appears to occur in a single phase with an intermediate time constant (panel C). These surprising results coincide with those of Kim and Rieke (2001).

In the rest of this section, we will focus on the two-phase decay of adaptation with a sudden rise of mean contrast. This two-phase decay of adaptation illustrates the inner workings of the Kalman theory. Three factors interact to control this decay. They are the gain itself, the estimate of the mean contrast in the environment and the actual mean contrast in the environment. If the actual mean contrast increases suddenly, then we get higher responses. As a result, they become more discrepant with the prediction stage and the estimated mean contrast starts to rise, pushing the gain down. The relationship between the estimated mean contrast and the gain is complex as revealed by equations (7), (13), (15) and (17). Changes in gain depend nonlinearly on the estimated mean contrast and on the gain itself. The second slow phase of decay reveals that this nonlinearity causes the gain to depend more weakly on the estimated mean contrast for certain values of these two variables. Although the estimated mean contrast may continue to change relatively rapidly due to the discrepancy of the responses, the gain may change more

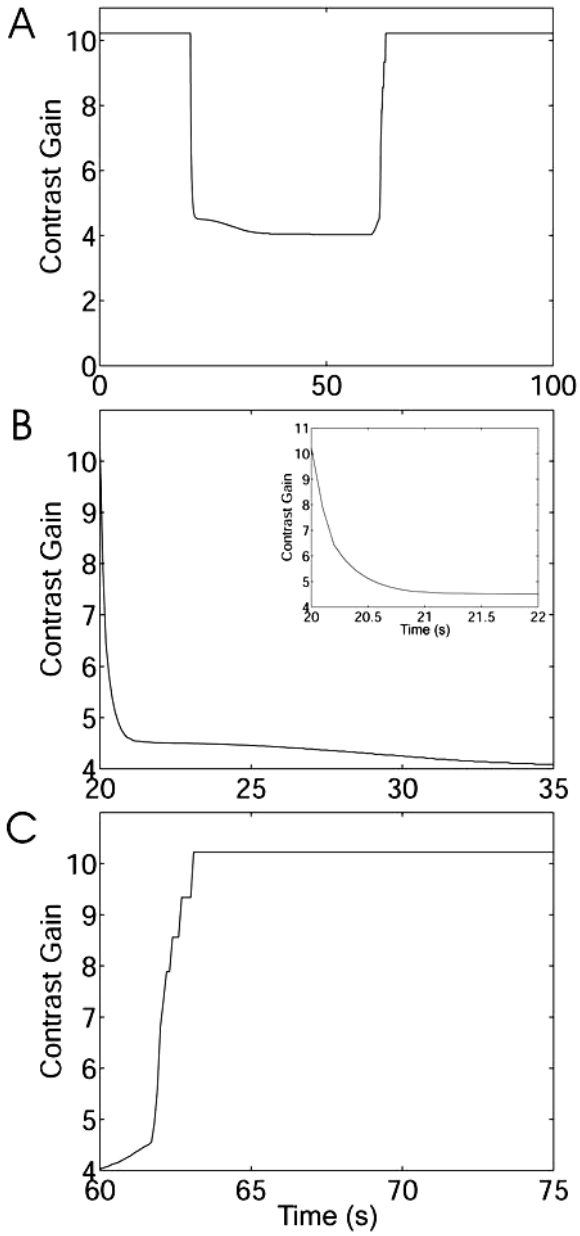

Figure 3. Contrast gain modulation in the model for contrast adaptation. A: this plot shows the simulated contrast gain when the mean contrast in the environment is modulated. In this example, the mean contrast goes from 0.14 ($0 \leqslant t < 20$ s) to 0.42 ($20 \leqslant t < 60$ s) and back ($60 \leqslant t < 100$ s). B: expanding the temporal axis after $t = 20$ s shows the detail of the decay of contrast gain after the sudden rise in mean contrast. One observes an initial fast decay (further expanded in the inset) followed by a slow decay. C: expanding the temporal axis just after $t = 60$ s shows the detail of the rise of contrast gain after the sudden fall in mean contrast. This rise occurs in one phase with a single, intermediate time constant.

slowly. In this case, the changes of gain may reflect more the time constants controlling the change of the estimated mean contrast. The second, slow phase of gain decay appears to have a

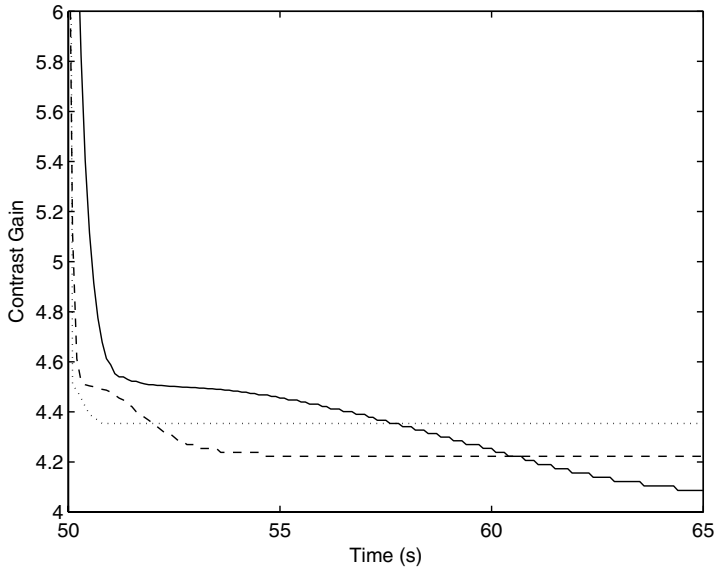

Figure 4. Dependence of the decay of contrast gain on the time constant of the prediction stage. The simulations were similar to those in figure 3, except that there was a single mean contrast increase from 0.14 to 0.42 at 50 s. We used three time constants of the prediction stage (equation (14)) in these simulations. These time constants were $\tau = 10$ s (full curve), $\tau = 3$ s (broken curve) and $\tau = 1$ s (dotted curve). It is apparent that the duration of the slow phase of the contrast-gain decay is roughly proportional to τ.

time constant similar to τ. Because τ is the time constant of the prediction stage (equations (6) and (14)), this suggests that this stage dominates the second phase. To test whether the second phase is dominated by τ, we repeated the simulations of figure 3 but with $\tau = 1$ and 3 s. Figure 4 compares the time constants of the decays for these values and for $\tau = 10$ s.

The figure shows that the late, slow decay is approximately proportional to τ. In contrast, the initial, fast decay is only mildly affected by τ, if at all. Consequently, τ, and thus the prediction stage of the Kalman filtering, appears to control the dynamics of late gain reduction when the contrasts of the environment increase suddenly.

What controls the early, fast phase of decay? According to the argument before figure 4, the contrast gain may depend more steeply on the estimated mean contrast of the environment during the early phase. In other words, small changes of the latter cause relatively large changes in the former. A prediction of this argument is that conditions that cause faster changes of estimated mean contrast should cause faster early decays, regardless of τ. The simplest way to cause such faster changes is to use deeper modulations of sudden contrast changes. Figure 5 illustrates what happens as the depth of modulation is varies.

This figure shows a portion of the initial decays for contrast modulations from 0.14 to 0.28 (full curve) and from 0.14 to 0.42 (dotted curve). Although the decays are similar in the beginning, the one for the deeper contrast modulation becomes faster after about 200 ms. Hence, its contrast gain reaches a lower level sooner, becoming less steeply dependent on changes of the estimated mean contrast at about 1 s. In turn, the contrast gain for the shallower contrast modulation takes longer to reach that stage ($\approx$1.5 s). After reaching that stage, the estimated mean contrast is already near its final value. The result is that the gain does not change much after that. Consequently, the decay for a shallow contrast modulation is predicted to

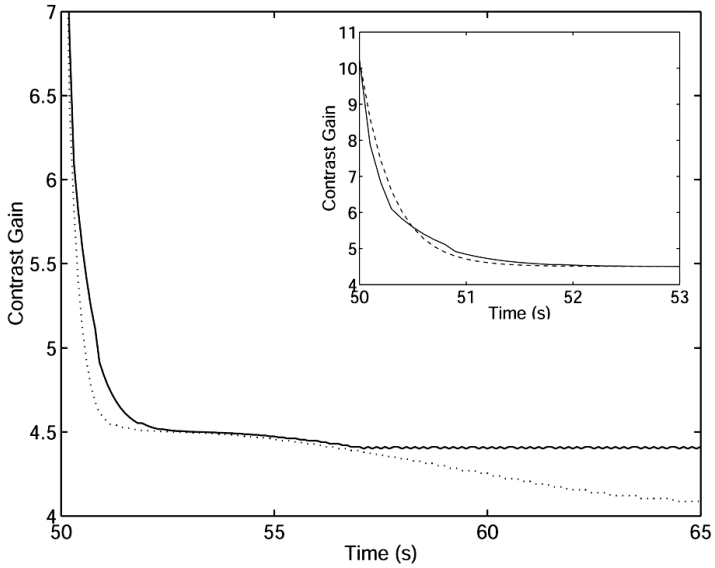

Figure 5. Dependence of the decay of contrast gain on the depth of the mean contrast modulation. This simulation was identical to that in figure 4, except that we used a single time constant ($\tau = 10$ s) and two depths of mean-contrast modulation. These depths were from 0.14 to 0.42 (dotted curve) and from 0.14 to 0.28 (full curve). The deeper mean-contrast modulation produced a faster initial decay than the shallower one. The initial decay for the latter was approximately exponential (broken curve in the inset). In turn, although the deep mean-contrast modulation produced a late, slow decay, the shallower modulation led to essentially no late decay.

have a single, relatively fast phase. Interestingly, as the inset shows, this fast decay appears to be approximately exponential (broken curve).

4. Discussion

We showed that, if one extends a Bayesian theory of sensory adaptation (Grzywacz and Balboa 2002) to explain how the system knows the parameters of the environment, the resulting framework becomes similar to Kalman filtering. As that form of filtering, the new framework has a prediction stage and a measurement stage (figure 1). The prediction stage tries to guess what the new parameters of the environment will be based on what they have been. In turn, the measurement stage gets new responses and computes their probabilities given the parameters predicted by the prediction stage. If these probabilities are high, then the system assumes these parameters as correct. Otherwise, the system selects parameters that are between those predicted by the prediction stage and those compatible with the obtained responses. With the new estimated parameters in hand, the new framework then sets the parameters of sensory adaptation. These parameters are chosen to minimize the errors in the performance of particular tasks. The errors are defined using Bayesian decision theory.

If sensory adaptation is a form of Kalman filtering, then there are some important consequences for the biology. The sensory system must store three types of knowledge to perform the tasks optimally. One type is the knowledge about the relevant statistics of the environment. In Bayesian theories, this type of knowledge is called prior probability function. Another type is the knowledge about the limitations of the biological processes and is called the likelihood function. The necessity of the likelihood function implies that adaptation is a

constrained optimization, with the constraint coming from the available biological hardware. Finally, the system must store knowledge about how the environment changes. This is a second type of prior probability function, but instead of saying what the statistics of the environment are, it says how rapidly they change in time.

Another consequence of adaptation being possibly a Kalman filtering is for the transient responses observed with sudden changes in stimulation. We would interpret these responses as the biological system being conservative. In other words, the transients would be due to the system not wanting to change until it is sure that the environment itself changed. Therefore, if the environment indeed changes, then the responses of the system would be produced transiently with the old adaptation parameters. This view of transient responses differs from the view that their role is to enhance the detection of changes in the environment (Stelmach *et al* 1984, Werblin *et al* 1988, Levanen and Sams 1997). These views are compatible, but they emphasize different aspects of the sensory signal.

A final consequence mentioned here is the prediction that sensory systems have optimal adaptation time courses. Recent work by Fairhall *et al* (2001) provided evidence that visual systems may have optimal or near-optimal adaptation dynamics. These investigators examined the dynamics of motion-sensitive cells in the fly's visual system. The experiments were such that the statistical properties of the visual environment were evolving dynamically. The results showed that the speed with which information processing was optimized by adaptation approached the physical limit imposed by statistical sampling and noise.

4.1. Retinal contrast adaptation

We illustrated the Kalman theory of adaptation with an example based on retinal contrast adaptation. For this purpose, we used knowledge about the statistics of contrasts in the environment (equation (13)) and about contrast processing in the retina (equation (15)). In addition, we made assumptions about how slowly the distribution of contrasts in the environment changes (equation (14)). And we assumed that one of the roles of retinal responses is to be maximally informative about contrasts (figure 2 and equations (16) and (17)). With these pieces of knowledge and assumptions, we could account for some complex features of the retinal contrast-adaptation data (figure 3). In particular, we could account for the two phases of contrast gain decay with sudden rise in the mean contrast of the environment. Furthermore, we accounted for the one-phase gain increase when the mean contrast falls (Chander and Chichilnisky 2001, Kim and Rieke 2001, Rieke 2001). (Interestingly, this asymmetry in the effects of contrast increase and decrease is similar to the results of DeWeese and Zador (1998). This is important because they did not assume maximal contrast information but the related optimal-variance estimation.) Further analysis of the model showed that the slow phase is dominated by the prediction stage of the Kalman filtering (figure 4), whereas the fast phase is dominated by the measurement stage (figure 5). In the retina, these two phases may also have separate mechanisms. Bipolar processes may underlie the fast phase (Brown and Masland 2001, Rieke 2001), while ganglion-cell processes may mediate the slow phase (Kim and Rieke 2001). One is thus tempted to speculate that outer plexiform layer mechanisms and bipolar-cell processes are involved in the measurement stage. Similarly, inner plexiform layer mechanisms and ganglion-cell processes may be involved in the prediction stage.

4.2. Other examples of retinal adaptation

The example of retinal contrast adaptation suggests that one can identify mechanisms implementing the various parts of the new theory (see figure 1). We already discussed for contrast adaptation the possible loci of the measurement and prediction stages. It is also

possible to postulate that the error measurement stage is in the inner plexiform layer. This is because the measurement of error must wait for the results of the interaction between the measurement and prediction stages. On the other hand, the adaptation feedback probably has a site of action on the outer plexiform layer. This is so, as it must affect the measurement stage.

This example was a case where adaptation is performed by the interaction of a network of cells contained in the retina. However, the mechanisms of retinal adaptation do not have to be network based and do not have to be located in the retina itself. Photoreceptor adaptation is a good example of a non-network process. The entire process takes place in a single cell by feedback of Ca^{2+} onto an enzymatic cascade (Koch and Stryer 1988, Nakatani and Yau 1988, Matthews *et al* 1990, Pugh and Lamb 1990, Gray-Keller and Detwiler 1994, Lagnado and Baylor 1994, Hackos and Korenbrot 1997). As the Ca^{2+} concentration is due mostly to processes in the cell's membrane, we postulate that the error signal is computed there. One can think of the adaptation error signal as the departure from equilibrium of Ca^{2+} concentration when Na^{+} channels open or close in the membrane.

An example of a non-retinal site of the computation of retinal adaptation is the modulation of spinules in teleost fishes. These are spine-like processes that extend from horizontal cells to the cone pedicle under light adaptation. The role of these processes may be to regulate the feedback from the horizontal cells to the cones (Raynauld *et al* 1979, Weiler and Wagner 1984, Djamgoz *et al* 1988). Spinule modulation is much slower than the contrast and photoreceptor adaptations mentioned above, lasting about 60 min. Hence, it is possible to postulate that the prior function of the prediction stage for this adaptation mechanism (see equations (6) and (14)) is related to a slow change of the environment. One possible idea is that it is related to adaptation to dawning (De Juan *et al* 1999). What is interesting for us here is that the error signal for spinule adaptation appears to be computed at an extra-retina location (De Juan *et al* 1996, De Juan and García 2001, Garcia *et al* 2002). Therefore, the prediction and error-measurement stages are in another part of the brain and their output is carried back to the retina through afferent fibres (Zucker and Dowling 1987).

4.3. Conclusion

The Kalman filtering theory of sensory adaptation hypothesizes that sensory systems use dynamic knowledge about the environment to optimize their own performances. The retinal examples above suggest that one may find direct correlates of the abstract elements of the theory in biology. Theories like these may thus close the gap between our cellular and computational understandings of the brain. We emphasize the roles of the likelihood function, the measurement stage and the loss function in closing this gap. The latter theoretical element forces us to specify the tasks that the biological system is performing. In turn, the former two elements force us to be clear about the biological constraints of optimization.

Acknowledgments

We would like to thank Mr Joaquín Rapela for discussions on various theoretical issues related to this project during its early phases. We would also like to thank Ms Lia Diepstraten for help with the preparation of the manuscript. This work was supported by National Eye Institute grants EY08921 and EY11170 to NMG and by grant PB-96-0414 from DGICYT to JDJ.

References

Anstis S 1996 Adaptation to peripheral flicker *Vis. Res.* **36** 3479–85

Atick J J and Redlich A N 1992 What does the retina know about natural scenes? *Neural Comput.* **4** 196–210

Balboa R M and Grzywacz N M 2000a The role of early retinal lateral inhibition: more than maximizing luminance information *Vis. Neurosci.* **17** 77–89

Balboa R M and Grzywacz N M 2000b The minimal-local asperity hypothesis of early retinal lateral inhibition *Neural Comput.* **12** 1485–517
Balboa R M and Grzywacz N M 2000c Occlusions and their relationship with the distribution of contrasts in natural images *Vis. Res.* **40** 2661–9
Baylor D A, Lamb T D and Yau K W 1979 Responses of retinal rods to single photons *J. Physiol.* **288** 613–34
Benardete E A and Kaplan E 1997 The receptive field of the primate P retinal ganglion cell, I: linear dynamics *Vis. Neurosci.* **14** 169–85
Benardete E A, Kaplan E and Knight B W 1992 Contrast gain control in the primate retina: P cells are not X-like, some M cells are *Vis. Neurosci.* **8** 483–6
Berger J O 1985 *Statistical Decision Theory and Bayesian Analysis* (New York: Springer)
Brown S P and Masland R H 2001 Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells *Nat. Neurosci.* **4** 44–51
Burgi P Y, Yuille A L and Grzywacz N M 2000 Probabilistic motion estimation based on temporal coherence *Neural Comput.* **12** 1839–67
Chander D and Chichilnisky E J 2001 Adaptation to temporal contrast in primate and salamander retina *J. Neurosci.* **21** 9904–16
Dahari R and Spitzer H 1996 Spatiotemporal adaptation model for retinal ganglion cells *J. Opt. Soc. Am.* A **13** 419–35
Datum M S, Palmieri F and Moiseff A 1996 An artificial neural network for sound localization using binaural cues *J. Acoust. Soc. Am.* **100** 372–83
Dayan P and Abbott L F 2001 *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (Cambridge, MA: MIT Press)
De Juan J and García M 2001 Spinules and nematosomes in retinal horizontal cells: a 'thorny' issue *Prog. Brain Res.* **131** 519–36
De Juan J, García M and Cuenca N 1996 Formation and dissolution of spinules and changes in nematosome size require optic nerve integrity in black bass *(Micropterus salmoides)* retina *Brain Res.* **707** 213–20
De Juan J, Sáez F, García M, Rallegue R and Sánchez-Vázquez F J 1999 Ocular melatonin is not involved in cone elongation in natural conditions in sea bass *(Dicentrarchus labrax)* retina *Invest Opthalmol. Vis. Sci.* **40** S611
DeMarco P J Jr, Brigell M G and Gordon M 1997 The peripheral flicker effect: desensitization of the luminance pathway by static and modulated light *Vis. Res.* **37** 2419–25
de Monasterio F M 1978 Properties of concentrically organized X and Y ganglion cells of macaque retina *J. Neurophysiol.* **41** 1394–417
DeWeese M and Zador A 1998 Asymmetric dynamics in optimal variance adaptation *Neural Comput.* **10** 1179–202
Djamgoz M B A, Downing J E G, Kirsch M, Prince D J and Wagner H J 1988 Plasticity of cone horizontal cell function in cyprinid fish retina: effects of background illumination of moderate intensity *J. Neurocytol.* **17** 701–10
Fairhall A L, Lewen G D, Bialek W and de Ruyter Van Steveninck R R 2001 Efficiency and ambiguity in an adaptive neural code *Nature* **412** 787–92
Field D J 1994 What is the goal of sensory coding? *Neural Comput.* **6** 559–601
Freeman A W and Badcock D R 1999 Visual sensitivity in the presence of a patterned background *J. Opt. Soc. Am.* A **16** 979–86
Fuortes M G F and Yeandle S 1964 Probability of occurrence of discrete potential waves in the eye of the Limulus *J. Physiol.* **47** 443–63
Garcia M, Grzywacz N M and De Juan J 2002 Interocular effect of intravitreal injection of 6-hydroxydopamine and dopamine on spinule formation in teleost retina *Histol. Histopathol.* **17** 721–9
Gray-Keller M P and Detwiler P B 1994 The calcium feedback signal in the phototransduction cascade of vertebrate rods *Neuron* **13** 849–61
Grzywacz N M and Balboa R M 2002 A Bayesian framework for sensory adaptation *Neural Comput.* **14** 543–59
Hackos D H and Korenbrot J I 1997 Calcium modulation of ligand affinity in the cyclic GMP-gated ion channels of cone photoreceptors *J. General Physiol.* **110** 515–28
Heinrich T S and Bach M 2001 Contrast adaptation in human retina and cortex *Invest Ophthalmol. Vis. Sci.* **42** 2721–7
Ho Y C and Lee R C K 1964 A Bayesian approach to problems in stochastic estimation and control *IEEE Trans. Autom. Control* **9** 333–9
Hochstein S and Shapley R M 1976 Quantitative analysis of retinal ganglion cell classifications *J. Physiol.* **262** 237–64
Isard M and Blake A 1996 Contour tracking by stochastic propagation of conditional density *Proc. Eur. Conf. on Comput. Vision (Cambridge, UK)* pp 343–56
Kalman R E 1960 A new approach to linear altering and prediction problems *Basic Eng.* **82** 35–45
Kim K J and Rieke F J 2001 Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells *Neurosci.* **21** 287–99

See endnote 2
See endnote 3

Koch K W and Stryer L 1988 Highly cooperative feedback control of retinal rod guanylate cyclase by calcium ions *Nature* **344** 64–6

Lagnado L and Baylor D A 1994 Calcium controls light-triggered formation of catalytically active rhodopsin *Nature* **367** 273–7

Laughlin S B 1989 The role of sensory adaptation in the retina *J. Exp. Biol.* **146** 39–62

Lee B B 1996 Receptive field structure in the primate retina *Vis. Res.* **36** 631–44

Levanen S and Sams M 1997 Disrupting human auditory change detection: Chopin is superior to white noise *Psychophysiology* **34** 258–65

Matthews H R, Fain G L, Murphy R L W and Lamb T D 1990 Light adaptation in cone photoreceptors of the salamander: a role for cytoplasmic calcium *J. Physiol.* **420** 447–79

Merwine D K, Amthor F R and Grzywacz N M 1995 Interaction between center and surround in rabbit retinal ganglion cells *J. Neurophysiol.* **73** 1547–67

Nakatani K and Yau K W 1988 Calcium and light adaptation in retinal rods and cones *Nature* **334** 69–71

Nelder J A and Mead R 1965 A simplex method for function minimization *Comput. J.* **7** 308–13

Pugh E N Jr and Lamb T D 1990 Cyclic GMP and calcium: the internal messengers of excitation and adaptation in vertebrate photoreceptors *Vis. Res.* **30** 1923–48

Rao R P 1999 An optimal estimation approach to visual perception and learning *Vis. Res.* **39** 1963–89

Rao R P and Ballard D H 1997 Dynamic model of visual recognition predicts neural response properties in the visual cortex *Neural Comput.* **9** 721–63

Raynauld J P, Laviolette J R and Wagner H J 1979 Goldfish retina: a correlate between cone activity and morphology of the horizontal cell in cone pedicles *Science* **204** 1436–8

Rieke F 2001 Temporal contrast adaptation in salamander bipolar cells *J. Neurosci.* **21** 9445–54

Ruderman D L and Bialek W 1994 Statistics of natural images: scaling in the woods *Phys. Rev. Lett.* **73** 814–17

Rushton W A H 1961 The intensity factor in vision *Light and Life* ed W D McElroy and H B Glass (Baltimore, MD: Johns Hopkins University Press) pp 706–22

Smirnakis S M, Berry M J, Warland D K, Bialek W and Meister M 1997 Adaptation of retinal processing to image contrast and spatial scale *Nature* **386** 69–73

Srinivasan M V, Laughlin S B and Dubs A 1982 Predictive coding: a fresh view of inhibition in the retina *Proc. R. Soc.* B **216** 427–59

Stelmach L B, Bourassa C M and Di Lollo V 1984 Detection of stimulus change: the hypothetical roles of visual transient responses *Percept. Psychophys.* **35** 245–55

Sum J, Leung C S, Young G H, Chan L W and Kan W K 1999 An adaptive Bayesian pruning for neural networks in a non-stationary environment *Neural Comput.* **11** 965–76

Thorson J and Biederman-Thorson M 1974 Distributed relaxation processes in sensory adaptation *Science* **183** 161–72

Tranchina D, Gordon J, Shapley R and Toyoda J 1987 Linear information processing in the retina: a study of horizontal cell responses *Proc. Natl Acad. Sci. USA* **78** 6540–2

Tyler C W and Liu L 1996 Saturation revealed by clamping the gain of the retinal light response *Vis. Res.* **36** 2553–62

Victor J D 1987 The dynamics of the cat retinal X cell centre *J. Physiol.* **386** 219–46

Weiler R and Wagner H J 1984 Light-dependent change of cone horizontal cell interactions in carp retina *Brain Res.* **298** 1–9

Werblin F, Maguire G, Lukasiewicz P, Eliasof S and Wu S M 1988 Neural interactions mediating the detection of motion in the retina of the tiger salamander *Vis. Neurosci.* **1** 317–29

Zar J H 1984 *Biostatistical Analysis* (Englewood Cliffs, NJ: Prentice-Hall)

Zhu S C and Mumford D 1997 Prior learning and Gibbs reaction-diffusion *IEEE Trans. Pattern Anal. Mach. Intell.* **19** 1236–50

Zucker C L and Dowling J E 1987 Centrifugal fibres synapse on dopaminergic interplexiform cells in the teleost reti *Nature* **330** 166–8

Queries for IOP paper 154125

Journal: **Network**
Author: **N M Grzywacz and J de Juan**
Short title: **Sensory adaptation as Kalman filtering: illustration with contrast adaptation**

Page 9

Query 1:
Author: Tranchina *et al* (1981) changed to Tranchina *et al* (1987) as in reference list. OK?

Page 17

Query 2:
Author: Datum *et al* (1996): 1996 OK?

Query 3:
Author: Isard and Blake (1996): Publisher?