

# Natural signal statistics and sensory gain control

Odelia Schwartz<sup>1</sup> and Eero P. Simoncelli<sup>2</sup>

---

<sup>1</sup>*Center for Neural Science, New York University, New York, NY 10003, USA*

<sup>2</sup>*Howard Hughes Medical Institute, Center for Neural Science, and  
Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA*

*Correspondence should be addressed to E.S. (eero.simoncelli@nyu.edu)*

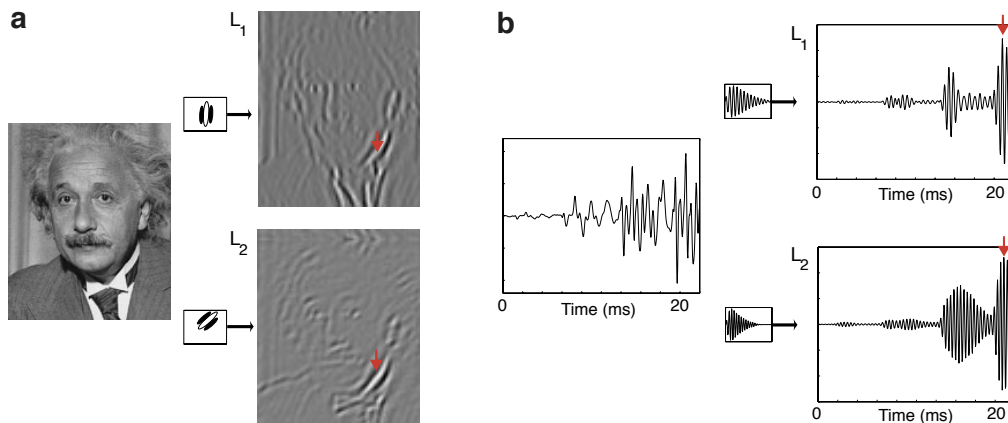
**We describe a form of nonlinear decomposition that is well-suited for efficient encoding of natural signals. Signals are initially decomposed using a bank of linear filters. Each filter response is then rectified and divided by a weighted sum of rectified responses of neighboring filters. We show that this decomposition, with parameters optimized for the statistics of a generic ensemble of natural images or sounds, provides a surprisingly good characterization of the nonlinear response properties of typical neurons in primary visual cortex or auditory nerve, respectively. These results suggest that nonlinear response properties of sensory neurons are not an accident of biological implementation, but serve an important functional role.**

Signals arising in the natural world are highly structured. To an observer with knowledge of these structures the signals are redundant, because one spatial or temporal portion of a given signal may be predicted from others. Indeed, this is why modern communication technologies are able to efficiently compress and transmit signals. It is widely assumed that neurons in sensory areas of the brain are adapted, through processes of evolution and development, to the signals to which they are exposed. This concept was made more precise by Atneave [1] who hypothesized that a guiding principle for the design of sensory systems is the statistically efficient representation of environmental information. Barlow [2] formalized a neural statement of this efficient coding hypothesis, in which he argued that the role of early sensory processing is to remove statistical redundancy or increase independence between neuronal responses to natural stimuli.

This hypothesis has led many researchers to “derive” models of sensory processing directly from statistical properties of natural signals (see [3] for review). In many such approaches, a linear basis is optimized such that the responses to natural signals are as statistically independent as possible. The basis functions resulting from such decomposition of natural images have been shown to have properties similar to receptive fields in primary visual cortex [e.g., 4, 5, 6]. Those resulting from decomposition of natural sounds are temporally localized bandpass filters, similar to those found in the peripheral auditory system [e.g., 7].

These linear models provide intriguing evidence for a link between statistics of natural signals and neuronal processing. But the statistical properties of natural signals are too complex to expect a linear transformation to produce an independent set of components. For example, visual images are formed from light reflected from different surfaces in the environment, but the light from these surfaces is typically combined nonlinearly, according to the rules of occlusion. Thus, it would seem unlikely that one could achieve a description as a linear superposition of independent patterns. In addition, it is well known that sensory neurons are highly nonlinear, even at relatively early stages of processing. Numerous nonlinear behaviors have been documented in physiological experiments in vision and audition over the past decades. These range from rectification and saturation of responses [e.g. 8, 9, 10], to suppression by non-optimal masks [e.g. 11, 12, 13, 14, 10, 15, 8], to changes in tuning curve shape as signal strength is varied [e.g. 16, 17, 18].

Here we offer a functional explanation for these sensory nonlinearities in terms of the efficient coding of natural signals. We show empirically that responses of a linear neural model to natural sounds and images exhibit striking statistical dependencies that cannot be eliminated with linear operations. We then describe a nonlinear model of neural processing which significantly reduces these statistical dependencies. We demonstrate through simulation that this model accounts for a number of nonlinear physiological behaviors that have been



**Fig. 1.** Linear filter responses to example image and sound stimuli. **(a)** A natural image convolved with two filters selective for the same spatial frequency, but different orientation and spatial position; the lower filter is oriented 45 degrees away, and shifted up by 4 pixels. At a given location, when the first filter responds weakly (gray areas) the second filter will also tend to respond weakly. But when the first filter responds strongly (black or white), the second filter is more likely to respond strongly. Red arrows indicate a location corresponding to a high contrast edge, in which both filters are responding strongly (the first positive, and the second negative). **(b)** A natural sound convolved with two filters tuned for different temporal frequencies (2000 and 2840 Hz center frequencies). Red arrows indicate a time at which both filters are responding strongly. Note also that when the first filter responds weakly, the second also tends to respond weakly.

observed in both visual and auditory neurons.

## Statistics of natural signals

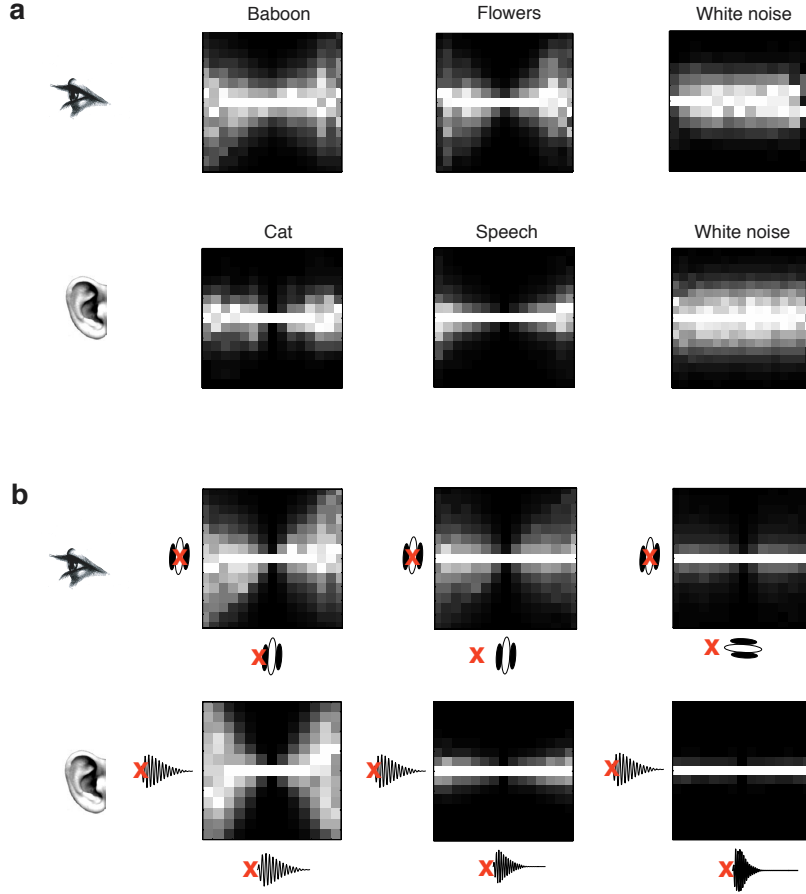
We begin by examining the statistical properties of natural signals processed with linear filters (see Methods). **Figure 1a** shows responses of a pair of non-overlapping oriented visual filters qualitatively similar to those used to characterize receptive fields in primary visual cortex (area V1). Note that both filters are likely to respond to prominent features in the image, such as around the eyes. These features are interspersed with regions in which both filters respond with very low activity. Similarly, **Fig. 1b** shows instantaneous responses of a pair of bandpass filters typical of the peripheral auditory system. As in the visual case, we observe the co-occurrence of large amplitude features separated by low amplitude intervals.

These examples suggest that responses of typical sensory filters to natural signals are not statistically independent. In order to examine this dependency more explicitly, consider a conditional histogram of the linear response of one visual filter ( $L_2$ ) given the response of another visual filter ( $L_1$ ) at a fixed relative orientation, size, and position (Fig. 2). Statistical independence means that knowledge of  $L_1$  should provide no information about  $L_2$ . Graphically, this is equivalent to saying that all vertical cross sections of the conditional his-

togram should be identical. In the histogram of **Fig. 2**, the mean of  $L_1$  (i.e. the mean value of each vertical cross section) is zero, independent of  $L_2$ , indicating that the two responses are uncorrelated. But the width of the distribution of  $L_1$  increases with the amplitude of  $L_2$ . That is, the *variance* of  $L_2$  depends on  $L_1$ .

This form of dependency appears to be ubiquitous - it is evident in a wide variety of natural images and sounds. It occurs even when the filters are chosen to be orthogonal, non-overlapping, or from a set that is optimized for statistical independence [e.g. 5]. The strength of the dependency does, however, depend on the specific pair of filters chosen, as illustrated in **Fig. 3**. Nevertheless, we emphasize that this dependency is a property of natural signals, and is *not* due purely to properties of this particular set of linear filters. For example, no such dependency is observed when the input consists of white noise.

We formalize the conditional relationship between a given pair of linear filter responses  $\{L_1, L_2\}$  with a model in which the variance of  $L_2$  is proportional to the squared value of  $L_1$  plus a constant (Methods, equation (1)). For a pair of filters with strongly dependent responses, this proportion is larger; for a pair whose responses are independent, this proportion is zero. Since  $L_2$  also depends on the responses of a number of other filters within a local neighborhood, we form a generalization of this conditional variance model in which



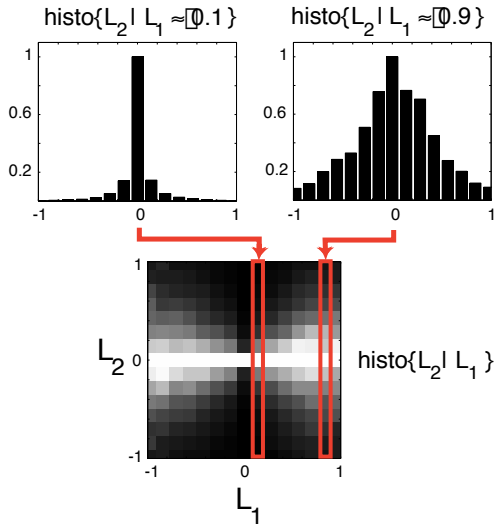
**Fig. 3.** Examples of variance dependency in natural signals. **(a)** Responses of two filters to several different signals (filters as in Fig. 1). Dependency is strong for natural signals, but is negligible for white noise. **(b)** Responses of different pairs of filters to a fixed natural signal. The strength of the variance dependency depends on the filter pair. For the image, the red X represents a fixed spatial location on the retina. The ordinate response is always computed with a vertical filter (shifted 4 pixels), vertical filter (shifted 12 pixels) and horizontal filter (shifted 12 pixels). For the sound, the red X represents a fixed time. Temporal frequency of ordinate filter is 2000 Hz. Temporal frequency of abscissa filter are 2000 Hz (shifted 9 msec in time), 2840 Hz (shifted 9 msec) and 4019 Hz (shifted 9 msec).

$L_2$  is proportional to a weighted sum of the squared responses over the neighborhood and an additive constant. We compute a set of optimal weights and additive constant by maximizing the likelihood of the conditional distribution over an ensemble of images or sounds (Methods, equation (6)). Intuitively, these weights are larger for pairs of filters that have stronger dependency. The constant represents the residual variance that cannot be predicted from neighboring filters.

If this model fully describes the dependency between filter responses, how can these responses be made independent? Given that the dependency governs only the variance, the natural solution is to *divide* the squared response of each filter by its variance, as predicted from

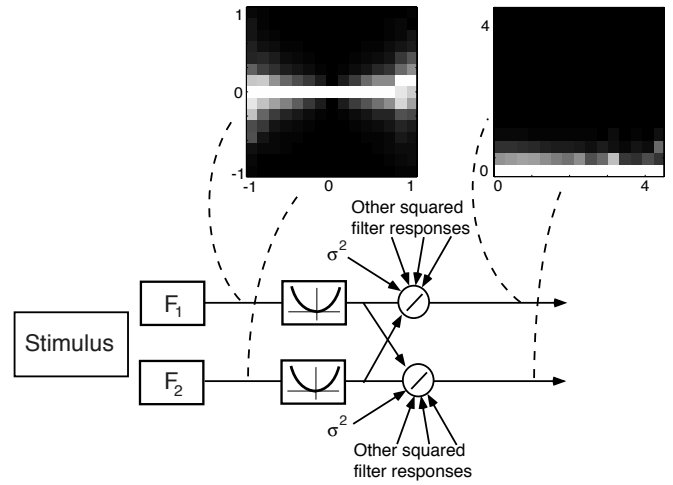
a linear combination of its neighbors (see Methods). **Figure 4** summarizes this model. A natural signal is passed through a bank of linear filters (only two are depicted, for readability). In the gain control stage, the squared response of each filter is divided by a weighted combination of squared responses of other filters in the population plus an additive constant (Methods, equation (4)). The resulting responses are significantly more independent. Related work examines models for variance dependency, as well as the conditions under which division is optimal [19].

The model illustrated in **Fig. 4** incorporates a form of automatic gain control known as “divisive normalization” that has been used to account for many nonlinear



**Fig. 2.** Joint statistics of a typical natural image as seen through two linear filters. Top: one-dimensional histograms of the linear response of a vertical filter ( $L_2$ ), conditioned on two different values of the response of a diagonal spatially shifted filter ( $L_1$ ). Pairs of responses are gathered over all image positions, and a joint histogram is constructed by counting the frequency of occurrence of each pair of responses. The two one-dimensional histograms are vertical slices of this joint histogram. Differing widths of these histograms clearly indicate that the filter responses are not statistically independent. Bottom: A grayscale image depicting the full two-dimensional conditional histogram. Pixel intensity is proportional to the bin counts, except that each column is independently re-scaled to fill the range of intensities. Responses of  $L_1$  and  $L_2$  are roughly decorrelated (expected value of  $L_2$  is approximately 0, independent of  $L_1$ ) but not statistically independent. Specifically, the variance of distribution of  $L_2$  increases with increasing value (both positive and negative) of  $L_1$ .

steady-state behaviors of neurons in primary visual cortex [e.g. 20, 21, 10]. Normalization models have been motivated by several basic properties. First, gain control allows a system with limited response range to handle a wider dynamic range of input. Divisive normalization achieves this goal, producing sigmoidal contrast-response functions similar to those seen in neurons. In addition, it seems advantageous for tuning curves in stimulus parameters such as orientation to retain their shape at different contrasts, even in the presence of response saturation [20]. Previous models have accomplished this by computing a normalization signal that is independent of parameters such as orientation (achieved with a uniformly weighted sum over the entire neural population). A consequence of this design is that the models can account for the response suppression that



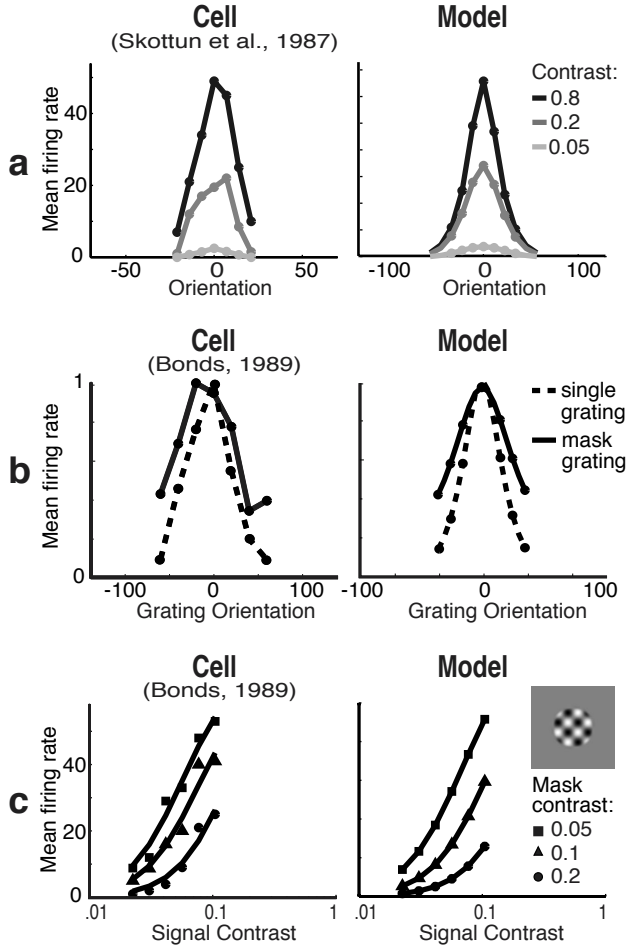
**Fig. 4.** Generic normalization model for vision and audition. Each filter response is divided by the weighted sum of squared responses of neighboring filters and an additive constant. Parameters are determined using Maximum Likelihood on a generic ensemble of signals (see Methods). The conditional histogram of normalized filter responses demonstrates that the variance of  $N_2$  is roughly constant, independent of  $N_1$ . The diagram is a schematic representation of the computation and is not meant to specify a particular mechanism or implementation (see Discussion).

occurs, for example, when a grating of non-optimal orientation is superimposed on a stimulus.

## Model simulations vs. physiology

We now compare our model with electrophysiological measurements from single neurons. In order to simulate an experiment, we choose a primary filter and a set of neighboring filters that will interact with this primary filter. We pre-compute the optimal normalization weights for an ensemble of natural signals (see Methods). We then simulate each experiment, holding all parameters of the model fixed, by computing the normalized responses of the primary filter to the experimental stimuli. We compare these responses to the physiologically measured average firing rates of neurons. Our extended normalization model, with all parameters chosen to optimize statistical independence of responses, can account for those nonlinear behaviors in V1 neurons previously modeled with divisive normalization (see above). **Figure 5** shows data and model simulations demonstrating preservation of orientation tuning curves and cross-orientation inhibition.

Our model also accounts for nonlinear behaviors



**Fig. 5.** Classical nonlinear behaviors of V1 neurons. (a) Contrast independence of orientation tuning [22]. (b) Orientation masking [23]. Dashed line indicates response to a single grating, as a function of orientation. Solid line indicates response to an optimal grating additively superimposed on a mask grating of variable orientation. All curves are normalized to have a maximum value of one. (c) Cross-orientation suppression [23]. Responses to optimal stimulus are suppressed by an orthogonal masking stimulus within the receptive field. This results in a rightward shift of the contrast response curve (on a log axis). Curves on cell data plot are fitted with a Naka-Rushton function,  $r(c) = c^2 / (ac^2 + b^2)$ .

not previously modeled using normalization. **Figure 6a** shows data from an experiment in which an optimal sinusoidal grating stimulus is placed inside the classical receptive field of a neuron in primary visual cortex of a Macaque monkey [24]. A mask grating is placed in an annular region surrounding the classical receptive field. Each curve in the figure indicates the response as a function of the center contrast for a particular surround con-

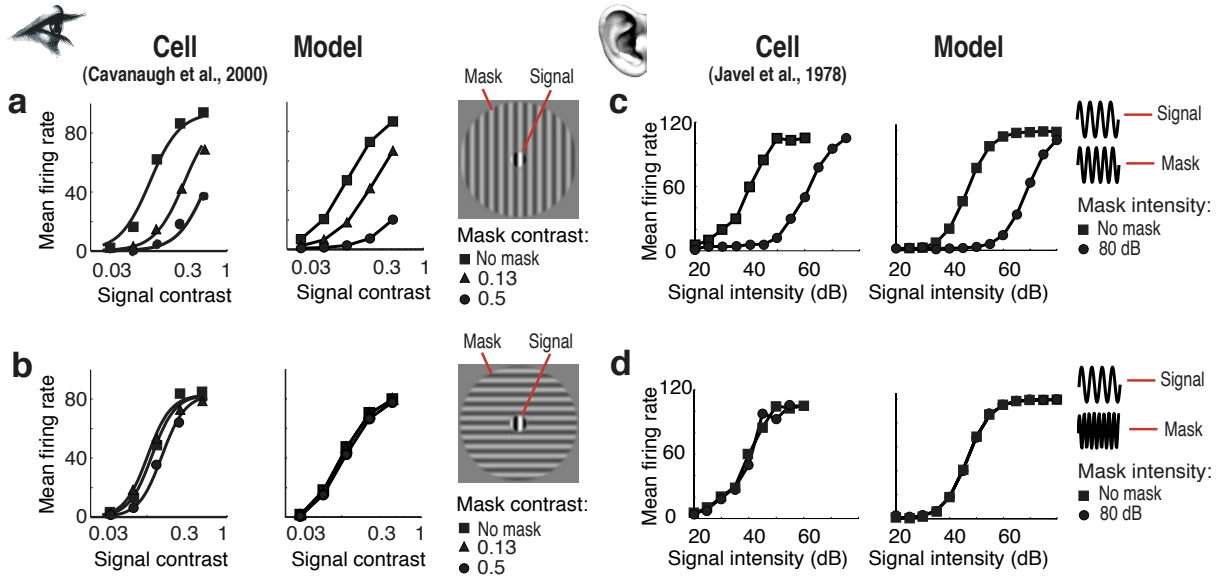
trast. The sigmoidal shape of the curves results from the squaring nonlinearity and the normalization. Presentation of the mask grating alone does not elicit a response from the neuron, but its presence suppresses the responses to the center grating. Specifically, the contrast response curves are shifted to the right (on a log axis), indicative of a divisive gain change. When the mask orientation is parallel to the center, this shift is much larger than when the mask orientation is orthogonal to the center (**Fig. 6b**).

Our model exhibits similar behavior (**Figs. 6a,b**), which is due to a suppressive weighting of neighboring model neurons with the same orientation preference that is stronger than that of neurons with perpendicular orientation preference (see also [25]). This weighting is determined by the statistics of our image ensemble, and is due to the fact that adjacent regions in natural images are more likely to have similar orientations than orthogonal orientations. For example, oriented structures in images (such as edges of objects) tend to extend along smooth contours, yielding strong responses in linear filters that are separated from each other spatially, but lying along the same contour (see also [26, 27]). Note that this behavior would not be observed in previous normalization models, because the parallel and orthogonal surround stimuli would produce the same normalization signal.

An analogous effect is seen in the auditory system. **Figure 6** shows example data recorded from a cat auditory nerve fiber [11], in which an optimal sinusoidal tone stimulus is combined with a masking tone. As in the visual data, the rate-level curves of the auditory nerve fiber shift to the right (on a log scale) in the presence of the masking tone (**Fig. 6c,d**). This shift is larger when the mask frequency is closer to the optimal frequency for the cell. Again, the model behavior is due to variations in suppressive weighting across neurons tuned for adjacent frequencies, which in turn arises from the statistical properties illustrated in **Fig. 3b**.

As mentioned above, a motivating characteristic of normalization models has been the preservation of tuning curve shape under changes in input level. However, the shapes of physiologically measured tuning curves for some parameters exhibit substantial dependence on input level in both audition [16] and vision [17, 18]. **Figure 7a** shows an example of this behavior in a neuron from primary visual cortex of a macaque monkey [24]. The graph shows the response of the cell as a function of the radius of a circular patch of sinusoidal grating, at two different contrast levels. The high-contrast responses are generally larger than the low-contrast responses, but in addition, the shape of the curve changes.





**Fig. 6.** Suppression of responses to optimal stimuli by masking stimuli. (a) Vision experiment [24]. Mean response rate of a V1 neuron as a function of contrast of an optimally oriented grating presented in the classical receptive field, in the presence of a surrounding parallel masking stimulus. Curves on cell data plots are fits of a Naka-Rushton equation with two free parameters [24]. (b) Mean response rate vs. center contrast, in the presence of an orthogonal surround mask. (c) Auditory experiment [11]. Mean response rate of an auditory nerve fiber a function of sound pressure level, in the presence of a non-optimal mask at 1.25 times the optimal frequency. (d) Mean response rate vs. sound pressure level, in the presence of a non-optimal mask at 2.08 times the optimal frequency. For all plots, maximum model response has been rescaled to match that of the cell.

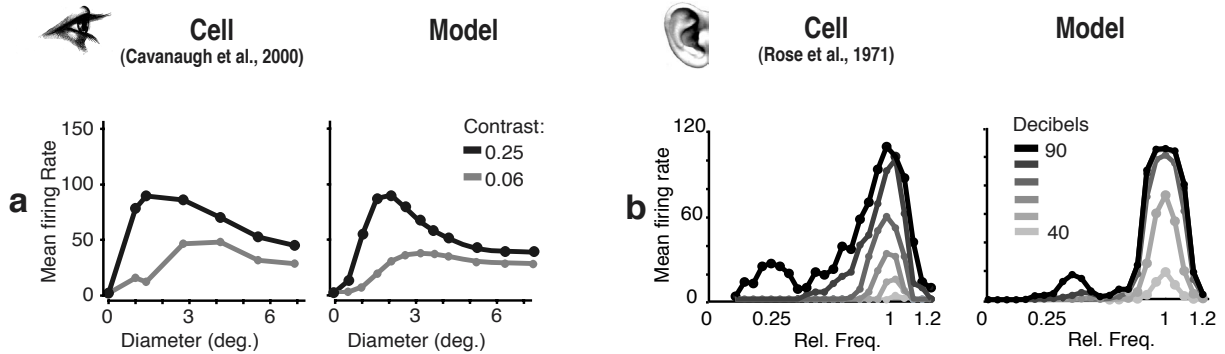
Specifically, for higher contrast, the peak response occurs at a smaller radius. The same behavior is seen in our model neuron.

**Figure 7b** shows an analogous result for a typical cell in the auditory nerve fiber of a squirrel monkey [16]. Responses are plotted as a function of frequency, for a number of different sound pressure levels. As the sound pressure level increases, the frequency tuning becomes broader, developing a “shoulder” and a secondary mode (**Fig. 7b**). Both cell and model show similar behavior, despite the fact that we have not adjusted the parameters to fit these data: all weights in the model are chosen by optimizing the independence of the responses to the ensemble of natural sounds. The model behavior arises because the weighted normalization signal is frequency-dependent. At low input levels, this frequency dependence is inconsequential because the additive constant dominates the signal. But at high input levels, this frequency dependence modulates the shape of the frequency tuning curve that is primarily established by the numerator kernel of the model. In **Fig. 7b**, the high contrast secondary mode corresponds to frequency bands with minimal normalization weighting.

## Discussion

We have described a generic nonlinear model for early sensory processing, in which linear responses are squared and then divided by a gain control signal computed as a weighted sum of the squared linear responses of neighboring neurons and a constant. The form of this model is chosen to eliminate the type of dependencies that we have observed between responses of pairs of linear receptive fields to natural signals (**Fig. 2**). The parameters of the model (in particular, the weights used to compute the gain control signal) are chosen to maximize the independence of responses to a particular set of signals. We have demonstrated that the resulting model accounts for a range of sensory nonlinearities in “typical” cells. Although there are quantitative differences amongst individual cells, the qualitative behaviors we have modeled have been observed in a number of laboratories. A remarkable aspect of our model is its ability to account for physiologically observed nonlinearities in two different modalities. This suggests a canonical neural mechanism for eliminating the statistical dependencies prevalent in typical natural signals.

The concept of gain control has been used previ-



**Fig. 7.** Nonlinear changes in tuning curves at different input levels. (a) Mean response rate of a V1 neuron as a function of stimulus radius for two different contrasts [24]. The peak response radius for both cell and model is smaller for the higher contrast. (b) Mean response rate of an auditory nerve fiber as a function of stimulus frequency for a range of sound pressure levels [16]. Tuning curve broadens and saturates at high levels. For all plots, maximum model response has been rescaled to match that of the cell.

ously to explain nonlinear behaviors of neurons. For example, a number of auditory models have incorporated explicit gain control mechanisms [28, 29, 8]. Visual models based on divisive normalization have been developed to explain nonlinear effects in cortical area V1 within the classical receptive field [20, 10]. The standard model assumes the response of each neuron is divided by an equally weighted sum of all other neurons and an additive constant. Our model utilizes a weighted sum for the normalization signal, and is thus able to account for a wider range of nonlinear behaviors. In addition, our model provides an ecological justification, through the efficient coding hypothesis [2], for such gain control models.

The ability of our model to account for nonlinear changes in tuning curve shape at different levels of input is particularly interesting. Such behaviors have been generally interpreted to mean that the fundamental tuning properties of cells depend on the strength of the input signal. But in our model, the fundamental tuning properties are determined by a fixed linear receptive field, and are modulated by a gain control signal with its own tuning properties. Although such behaviors may appear to be artifacts, our model suggests that they occur naturally in a system that is optimized for statistical independence over natural signals.

Our current model provides a functional description, and does not specify the circuitry or biophysics by which these functions are implemented. Our normalization computation is performed instantaneously and we have only modeled mean firing rates. Normalization behavior could potentially arise through a number of mechanisms. For example, feedforward synaptic de-

pression mechanisms have been documented and have been shown to exhibit gain control properties [e.g. 30]. Although such mechanisms may account for suppressive behaviors within the classical receptive field, they seem unlikely to account for such behaviors like those shown in Fig. 6. It has also been proposed that normalization could result from shunting inhibition driven by other neurons [e.g. 31, 32, 33]. This type of implementation necessarily involves recursive lateral or feedback connections and thus introduce temporal dynamics. Some researchers have described recurrent models that can produce steady-state responses consistent with divisive normalization in primary visual cortex [20, 10].

Some of the gain control behaviors we describe may be attributed to earlier stages of neural processing. Gain control is known to occur at the level of the retina [e.g. 9, 34], although selectivity for orientation does not arise before cortical area V1. In fact, Ruderman and Bialek [35] have suggested division by local contrast as a means of maximizing marginal entropy, thus providing a functional explanation for gain control in the retina. Our work differs conceptually in the choice of statistical criteria (independence between filters, as opposed to marginal statistics of one filter). In audition, outer hair cells have been implicated in providing gain control [8, 36], and some of the behaviors we describe at the level of the auditory nerve have also been documented in recordings from basilar membrane.

Our model is based on a mechanism that is fundamentally suppressive, but a number of authors have reported facilitative influences in both vision and audition [e.g. 37, 38, 39, 14]. Some of these facilitative effects might be explained by the use of masking stimuli

that inadvertently excite the receptive field of the neuron [e.g., 13, 40] thus causing suppression to overcome facilitation only at high contrasts or sound pressure levels of the mask. Facilitative effects might also be explained by dis-inhibition, in which a third cell inhibits a second cell, thus releasing its inhibition of the recorded cell. As mentioned above, our current model does not utilize a recurrent implementation and thus cannot predict such effects.

The relationship between the model and perception should also be explored. For example, psychophysical experiments suggest that visual detectability is enhanced along contours [41]. At first glance, this might appear to be inconsistent with our model, in which neurons that lie along contours will suppress each other. But the apparent contradiction is based on the unsubstantiated intuition that a reduction in the neural responses implies reduced detectability. Presumably, *any* difference in relative activity of neurons along the contour, as compared with the activity of neurons in other regions, could be used for contour detection. More generally, examination of the implications of our model for perception requires a method of extracting a percept from a population of neural responses. While this has not been done for contour detection, we find it encouraging that other basic percepts have been explained in the context of a population of neurons performing gain control (e.g. detectability of a grating in the presence of a mask [42]; perceptual segregation of visual textures [43]).

There are a number of important directions for further refinement of the connection between natural signal statistics and neuronal processing. We have optimized our model for a generic signal ensemble; it could be the case that neurons are specialized for particular subclasses of signals [e.g. 44]. Moreover, it would be interesting to model the mechanisms and associated timescales (i.e., evolution, development, learning, and adaptation) by which the optimization occurs. For example, some visual adaptation effects have been explained by adjusting model parameters according to the statistical properties of recent visual input [45, 46]. A more complete theory also requires an understanding of which groups of neurons are optimized for independence. A sensible assumption might be that each stage of processing in the system takes the responses of the previous stage and attempts to eliminate as much statistical redundancy as possible, within the limits of its computational capabilities. It remains to be seen how much of sensory processing can be explained using such a bottom-up criterion.

Future work should also be directed towards testing the efficient coding hypothesis experimentally. Some

support for the hypothesis has been obtained through recordings from groups of neurons [47, 48] under naturalistic stimulation conditions. We believe that improvements in both experimental techniques and statistical models of natural signals will continue to provide new opportunities to test and extend the efficient coding hypothesis proposed by Barlow forty years ago.

## Methods

For the auditory simulations we use a set of Gamma-tone filters as the linear front end [49]. We choose a primary filter with center frequency 2000 Hz, and a neighborhood of filters for the normalization signal: 16 filters with center frequencies 205 to 4768 Hz, and replicas of all filters temporally shifted by 100, 200, and 300 samples. For the visual simulations, linear receptive fields are derived using a multi-scale oriented decomposition known as the steerable pyramid [50]. The primary filter is vertically oriented with peak spatial frequency of  $1/8$  cycles/pixel. The filter neighborhood includes all combinations of two spatial frequencies, four orientations, two phases, and a spatial extent of three times the diameter of the primary filter. Responses are horizontally and vertically subsampled at 4-pixel intervals. To reduce the dimensionality of the weight vector that needs to be optimized, we assume that weights for two filters with differing phase are the same, thus guaranteeing a phase-invariant normalization signal. We also assume vertical and horizontal symmetry. We have verified that these simplification do not substantially alter the simulation results.

Our ensemble of natural sounds consists of 9 animal and speech sounds, each approximately 6 seconds long. The sounds are obtained from commercial compact disks and converted to sampling frequency of 22050 Hz. The natural image ensemble consists of 10 images obtained from a database of standard images used in image compression benchmarks (known as: boats, gold-hill, Einstein, Feynman, baboon, etc). We have obtained similar results using an intensity calibrated image set [6].

For a pair of filters, we model the variance of response of the first filter given the response of the second filter to a visual/auditory stimulus as follows:

$$\text{var}(L_1|L_2) = wL_2^2 + \sigma^2 \quad (1)$$

where  $L_1$  and  $L_2$  are the linear responses of the two filters. This conditional variance dependency is elimi-



nated by dividing:

$$R_1 = \frac{L_1^2}{wL_2^2 + \sigma^2} \quad (2)$$

We assume a generalization of this dependency to a population of filters. We model the variance dependency of the response of filter  $L_i$  given the responses of a population of filters  $L_j$  in a neighborhood  $N_i$ :

$$\text{var}(L_i | \{L_j, j \in N_i\}) = \sum_j w_{ji} L_j^2 + \sigma^2 \quad (3)$$

Again, the conditional variance dependency is eliminated by dividing:

$$R_i = \frac{L_i^2}{\sum_j w_{ji} L_j^2 + \sigma^2} \quad (4)$$

We would like to choose the parameters of the model (the weights  $w_{ji}$ , and the constant  $\sigma$ ) to maximize the independence of the normalized response to an ensemble of natural images and sounds. Such an optimization is computationally prohibitive. In order to reduce the complexity of the problem, we assume a Gaussian form for the underlying conditional distribution:

$$\mathcal{P}(L_i | \{L_j, j \in N_i\}) = \frac{1}{\sqrt{2\pi (\sum_j w_{ji} L_j^2 + \sigma^2)}} \exp \left[ \frac{-L_i^2}{2 (\sum_j w_{ji} L_j^2 + \sigma^2)} \right] \quad (5)$$

We then maximize the product of such distributions over the image/sound data at all positions/times  $x$ .

$$\hat{w}_{ji}, \hat{\sigma} = \arg \max_{w_{ji}, \sigma} \prod_x \mathcal{P}(L_i(x) | \{L_j(x), j \in N_i\}) \quad (6)$$

We solve for the optimal parameters numerically, using conjugate gradient descent. The weights are constrained to be positive, to avoid instability in optimization convergence and in simulations. Note that the value of  $\hat{\sigma}$  depends on the somewhat arbitrary scaling of the input signal (i.e., doubling the input strength would lead to a doubling of  $\hat{\sigma}$ ). Neurons also exhibit a range of different sensitivities. As such, we have chosen different values of  $\sigma$  in our simulations to match the sensitivity of the cell being modeled.

## References

- [1] F Attneave. Some informational aspects of visual perception. *Psych. Rev.*, 61:183–193, 1954.
- [2] H B Barlow. Possible principles underlying the transformation of sensory messages. In W A Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA, 1961.
- [3] E Simoncelli and B Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, May 2001.
- [4] B A Olshausen and D J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [5] A J Bell and T J Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [6] J H van Hateren and A van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B*, 265:359–366, 1998.
- [7] A J Bell and T J Sejnowski. Learning the higher-order structure of a natural sound. *Network: Computation in Neural Systems*, 7:261–266, 1996.
- [8] D Geisler. *From Sound to Synapse: Physiology of the Mammalian Ear*. Oxford University Press, New York, 1998.
- [9] R Shapley and C Enroth-Cugell. Visual adaptation and retinal gain control. *Progress in Retinal Research*, 3:263–346, 1984.
- [10] M Carandini, D J Heeger, and J A Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17:8621–8644, 1997.
- [11] E Javel, D Geisler, and A Ravindran. Two-tone suppression in auditory nerve of the cat: Rate-intensity and temporal analyses. *J. Acoust. Soc. Am.*, 63(4):1093–1104, 1978.
- [12] J J Knierim and D C Van Essen. Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67:961–980, 1992.
- [13] G C DeAngelis, R D Freeman, and I Ohzawa. The organization of suppression in receptive fields of neurons in the cat’s visual cortex. *Journal of Neurophysiology*, 68:144–163, 1994.
- [14] A M Sillito, K L Grieve, H E Jones, J Cudeiro, and J Davis. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, 378:492–496, 1995.

- [15] J B Levitt and J S Lund. Contrast dependence of contextual effects in primate visual cortex. *Nature*, 387:73–76, 1997.
- [16] J E Rose, D J Anderson, and J F Brugge. Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *Journal Neurophys.*, 34:685–699, 1971.
- [17] M P Sceniak, D L Ringach, M J Hawken, and R Shapley. Contrast’s effect on spatial summation by macaque V1 neurons. *Nat Neurosci.*, 8:733–739, 1999.
- [18] M K Kapadia, G Westheimer, and C D Gilbert. Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc Natl Acad Sci*, 21:12073–12078, 1999.
- [19] M J Wainwright, E P Simoncelli, and A S Willsky. Random cascades on wavelet trees and their use in modeling and analyzing natural imagery. *Applied and Computational Harmonic Analysis*, 11(1), July? 2001. Special issue on wavelet applications. In press.
- [20] D J Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.
- [21] W S Geisler and D G Albrecht. Cortical neurons: Isolation of contrast gain control. *Vision Research*, 8:1409–1410, 1992.
- [22] B C Skottun, R D Freeman, G Sclar, I Ohzawa, and R D Freeman. The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J. Neurophysiol.*, 57:773–786, 1987.
- [23] A B Bonds. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, 2:41–55, 1989.
- [24] J R Cavanaugh. *Properties of the Receptive Field Surround in Macaque Primary Visual Cortex*. PhD thesis, Center for Neural Science, New York University, August 2000.
- [25] E P Simoncelli and O Schwartz. Image statistics and cortical normalization models. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors, *Adv. Neural Information Processing Systems*, volume 11, pages 153–159, Cambridge, MA, 1999. MIT Press.
- [26] W S Geisler, J S Perry, B J Super, and D P Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41(6):711–724, March 2001.
- [27] M Sigman, G A Cecchi, C D Gilbert, and M O Magnasco. On a common circle: Natural scenes and Gestalt rules. *Proc. National Academy of Sciences*, 98(4):1935–1940, 2001.
- [28] R F Lyon. Automatic gain control in cochlear mechanics. In P Dallos et al., editor, *The Mechanics and Biophysics of Hearing*, pages 395–420. Springer-Verlag, 1990.
- [29] K Wang and S Shamma. Self-normalization and noise-robustness in early auditory representations. In *IEEE Trans. Speech and Audio Proc.*, volume 2, pages 421–435, 1994.
- [30] L F Abbott, J A Varela, K Sen, and S B Nelson. Synaptic depression and cortical gain control. *Science*, 275:220–224, 1997.
- [31] L J Borg-Graham, C Monier, and Y Frénac. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393:369–373, 1998.
- [32] J A Hirsch, J-M Alonso, R C Reid, and L M Martinez. Synaptic integration in striate cortical simple cells. *J. Neurosci.*, 18(22):2517–9528, 1998.
- [33] J Anderson, M Carandini, and D Ferster. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.*, 84:909–926, 2000.
- [34] J D Victor. The dynamics of the cat retinal X cell centre. *J. Physiol.*, 386:219–246, 1987.
- [35] D L Ruderman and W Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Letters*, 73(6):814–817, 1994.
- [36] H B Zhao and J Santos-Sacchi. Auditory collusion and a coupled couple of outer hair cells. *Nature*, 399(6734):359–362, 1999.
- [37] L Maffei and A Fiorentini. The unresponsive regions of visual cortical receptive fields. *Vision Research*, 16:1131–1139, 1976.
- [38] J I Nelson and B J Frost. Intracortical facilitation among co-oriented, coaxially aligned simple cells in cat striate cortex. *Exp. Brain Res.*, 61:54–61, 1985.
- [39] C D Gilbert and T N Wiesel. The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, 30:1689–1701, 1990.

- [40] B Delgutte. Physiological models for basic auditory percepts. In H. Hawkins and T. McMullen, editors, *Auditory Computation*, pages 157–220. Springer-Verlag, 1996.
- [41] D J Field, A Hayes, and R F Hess. Contour integration by the human visual system: Evidence for a local “association field”. *Vision Research*, 33:173–93, 1993.
- [42] J M Foley. Human luminance pattern mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A*, 11(6):1710–1719, 1994.
- [43] N Graham and A Sutter. Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, 40:2737–61, 2000.
- [44] F Rieke, D A Bodnar, and W Bialek. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B*, 262:259–265, 1995.
- [45] Horace B Barlow and P Foldiak. Adaptation and decorrelation in the cortex. In R Durbin, C Miall, and G Mitchinson, editors, *The Computing Neuron*, chapter 4, pages 54–72. Addison-Wellesley, New York, 1989.
- [46] M J Wainwright, O Schwartz, and E P Simoncelli. Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In R Rao, B Olshausen, and M Lewicki, editors, *Probabilistic Models of the Brain: Perception and Neural Function*, chapter 10. MIT Press, 2001. In press.
- [47] W E Vinje and J L Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287:1273–1276, Feb 2000.
- [48] S Nirenberg, S M Carcieri, A L Jacobs, and P E Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411:698–701, June 2001.
- [49] M Slaney. An efficient implementation of the Patterson and Holdworth auditory filter bank. *Apple Technical Report 35*, 1993.
- [50] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger. Shiftable multi-scale transforms. *IEEE Trans Information Theory*, 38(2):587–607, March 1992. Special Issue on Wavelets.

**Acknowledgements:** We thank James Cavanaugh, Wyeth Bair, and J. Anthony Movshon for providing us with physiological data.