Neuron, Vol. 26, 695–702, June, 2000, Copyright ©2000 by Cell Press

Adaptive Rescaling Maximizes Information Transmission

Naama Brenner, William Bialek,* and Rob de Ruyter van Steveninck
NEC Research Institute
Princeton, New Jersey 08540

Summary

Adaptation is a widespread phenomenon in nervous systems, providing flexibility to function under varying external conditions. Here, we relate an adaptive property of a sensory system directly to its function as a carrier of information about input signals. We show that the input/output relation of a sensory system in a dynamic environment changes with the statistical properties of the environment. Specifically, when the dynamic range of inputs changes, the input/output relation rescales so as to match the dynamic range of responses to that of the inputs. We give direct evidence that the scaling of the input/output relation is set to maximize information transmission for each distribution of signals. This adaptive behavior should be particularly useful in dealing with the intermittent statistics of natural signals.

Introduction

One of the major problems in processing the complex, dynamic signals that occur in the natural environment is providing an efficient representation of these data. More than 40 years ago, Attneave (1954) and Barlow (1961) suggested that steps in the neural processing of information could be understood as solutions to this problem of efficient representation. This idea was later developed by many groups, especially in the context of the visual system. Efficient representation requires a matching of the coding strategy to the statistical structure of incoming signals. At early stages of the visual pathway, for example, lateral inhibition or center–surround organization of receptive fields can be seen as a strategy for reducing redundancy among signals carried by neighboring neurons. At higher levels of processing, a description of images in terms of objects is more efficient than is a representation of light intensity in the pixels of the photoreceptor array (Attneave, 1954).

Much of the work that followed Attneave's and Barlow's ideas has focused on the matching of neural performance to global statistical properties of natural signals. Such a matching is likely to have taken place on the long timescales of evolution or development. While this approach has been successful in some cases (Laughlin, 1981; Atick, 1992), it should be kept in mind that the statistical properties of natural signals are highly variable as we move in the world and as time passes. The mean light level, for example, changes by orders of magnitude as we leave a sunny region and enter a forest. Adaptation to mean light level ensures that our visual responses are matched to the average signal in real time, thus maintaining sensitivity to the fluctuations around this mean. But the fluctuations themselves are intermittent, such that periods of large fluctuations are interspersed with periods of relative "quiet." Recent observations indicate that the intermittency of natural signals is of a special form: statistics such as the variance and correlation function are stationary over some regions, with slowly modulated parameters over larger regions (Ruderman and Bialek, 1994; Nelken et al., 1999). The principle of efficient coding suggests that the nervous system should adapt its strategy to these local statistical properties of the stimulus. Evidence for such statistical adaptation in the early stages of vision has been found in the fly (van Hateren, 1997) and in the vertebrate retina (Smirnakis et al., 1997), where mechanisms of gain control are well known (Shapley and Victor, 1979a, 1979b, 1980). From a more functional point of view, one would like to observe directly the changes in processing strategy and to demonstrate that these changes in fact enhance the efficiency of coding in response to changes in the statistics of visual inputs.

Here, we investigate the coding of dynamic velocity signals in a motion-sensitive neuron of the fly's visual system. Previous work has shown that this system adapts to constant velocity signals (Maddess and Laughlin, 1985; de Ruyter van Steveninck et al., 1986), to the variance of spatial image contrast (de Ruyter van Steveninck et al., 1996), and to the timescales in the stimulus (Borst and Egelhaaf, 1987; de Ruyter van Steveninck and Bialek, 1996). In this work, we focus on adaptation to the distribution from which velocity signals are drawn. We find a dramatic adaptive rescaling of the system's input/output relation with the standard deviation of the signal distribution. Further, we find that the magnitude of the rescaling selected by the adaptation process optimizes information transmission. Identical adaptive effects occur in response to signals that differ by orders of magnitude in timescale, suggesting that the molecular and cellular mechanisms underlying these adaptive effects span a similar range of timescales.

Results

We use as an experimental test case the H1 neuron in the visual system of the blowfly, which is sensitive to horizontal motion across the visual field. H1 responds by generating action potentials to motion in its preferred direction and is inhibited by motion in the opposite direction. The spike trains of H1 carry information about the time-dependent horizontal velocity (Strong et al., 1998). We record from H1 extracellularly to obtain a sequence of spike times (see Procedures). The high degree of stimulus control and the stability of recordings from this neuron have made it a well-studied system for problems of real time coding of time-varying signals (Rieke et al., 1997).

Here, we focus on how the system deals with changes

*To whom correspondence should be addressed (e-mail: bialek@research.nj.nec.com).

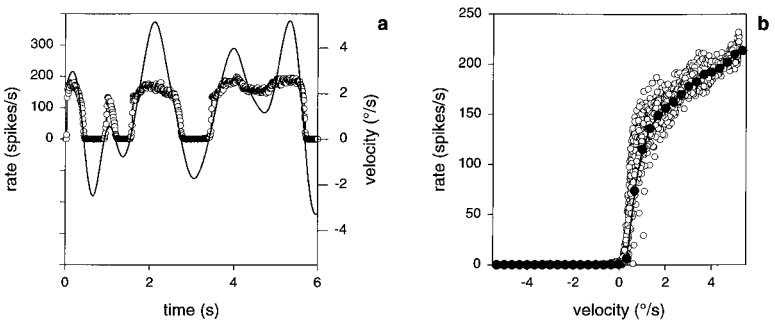

Figure 1. Response to Slowly Varying Inputs

The neural response function of H1 for a slowly varying stimulus ensemble.

(a) Open circles (left-hand axis), firing rate as a function of time, r(t), estimated by counting spikes in bins of 10 ms and averaging over 180 repeated presentations of the velocity stimulus s(t). Solid line (right-hand axis), stimulus velocity.

(b) The neural response r(s) is constructed by reading the firing rate at each time bin and plotting it as a function of the stimulus velocity 30 ms earlier. Open circles, scatter plot of rate versus velocity; closed circles, average response function, obtained by discretizing velocities in bins and averaging the corresponding rates.

in the dynamic range of motion signals, such as might occur in the transition from straight flight to chasing behavior. As a simplified version of this situation, we consider the encoding of horizontal velocity signals that are drawn from a distribution with zero mean, using the variance as a control parameter. We first describe the phenomenology of adaptive rescaling of input/output relations and the accompanying invariance of statistics of spike trains. Then, we relate this property to functionality, showing that adaptive rescaling enables the system to maximize information transmission.

Phenomenology: Adaptive Rescaling

In this section, we describe the phenomenology of adaptive rescaling of input/output relations. The experiment is designed as follows. Several stimulus ensembles are presented to the fly that differ only in their velocity variance; all other stimulus properties (for example, image contrast) are held fixed. Once the adaptation processes have reached a steady state, we construct the input/output relations and compare these relations as they are found with different ensembles. We focus on two types of stimulus dynamics, one with slowly, the other with rapidly, varying stimuli. These cases show very similar forms of adaptive rescaling, although the technicalities involved in constructing the input/output relation are different.

Consider first an ensemble of horizontal velocity stimuli, s(t), with a time variation much slower than the typical time between spikes. The value of the stimulus is then encoded by the local, slowly varying firing rate; this is a quasistatic extension of the classic rate code suggested by Adrian (1928). Figure 1a shows the firing rate as a function of time together with the slowly varying stimulus s(t). The velocity stimulus in this experiment had a correlation time of $\tau \approx 1$ s and a Gaussian distribution with standard deviation of 2.3°/s. A 20 s sample from this ensemble was presented repeatedly to the fly (180 repeats), and the firing rate r(t) was obtained by averaging over the presentations. The firing rate follows the velocity stimulus s(t), with rectification due to the direction selectivity of the neuron, and we may define an instantaneous relation r(s) between stimulus and firing rate. Figure 1b shows this function, which represents the input/output relation of the neuron when stimuli are drawn from the "context" provided by the distribution of signals, P[s(t)], used in this experiment. We emphasize that the firing rate as a function of image velocity is measured not by suddenly moving an image that was previously stationary; rather, the image motion is an ongoing random process, sometimes slower and sometimes faster, such that stimuli always are in their statistical context.

As seen from the scatter plot in Figure 1b, the instantaneous relation between rate and velocity is not perfectly unique; the scatter results from dependencies on other signal properties, such as acceleration. For the slowly varying stimuli considered here, this dependence on acceleration is negligible, while in other cases, it must be separated out (see below).

The stimulus in this experiment is changing so slowly that to keep track of its statistical properties requires processes on the timescale of many seconds. Although the encoding of velocity by H1 is characterized by integration times that range from 20 to 300 ms, there are also adaptation processes with timescales of many seconds, so it is reasonable to ask if "adaptation to statistics" exists. To see whether this is the case, we compare the neural response in Figure 1 with that measured in a similar stimulus ensemble, but with different standard deviation. In practice, we use the same long stimulus segment and multiply its amplitude by a constant, thus keeping the timescales in the stimulus the same. Figure 2 shows the time-dependent firing rates in two experiments, where the standard deviations were $\sigma_1 = 2.3°/s$ and $\sigma_2 = 4.6°/s$. While the velocity stimulus was doubled between the two experiments, the response is very similar, as observed also for retinal neurons (Meister and Berry, 1999). This suggests that spikes are not generated by the same fixed rule in these two cases and that the system does keep track of the overall structure of the stimulus ensemble.

Figure 3a shows the input/output relations of the same neuron in two experiments, obtained as in Figure 1b. These curves are similar in shape but have different scales. Normalizing the stimulus by its standard deviation, and the rate by the average rate, the two curves overlap, as shown in Figure 3b. The input/output relation represents average responses; Figure 3c shows the trial-to-trial variability in the response, represented by the standard deviation of the spike rate in 10 ms bins. This variability is small and weakly dependent on the rate. Moreover, Figure 3d shows that the variability in the

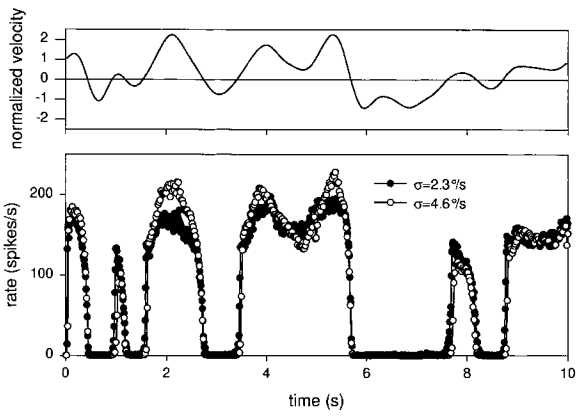

Figure 2. Rescaled Inputs Elicit the Same Responses

Time-dependent firing rates of H1 for two stimulus ensembles, with standard deviations of $\sigma_1 = 2.3°/s$ and $\sigma_2 = 4.6°/s$. The short segment of the normalized stimulus s(t), shown by a solid line, is identical in shape in the two experiments but has different scales. Although this scale was doubled, the response was almost identical, implying adaptation to the range of stimuli in the distribution from which they are drawn.

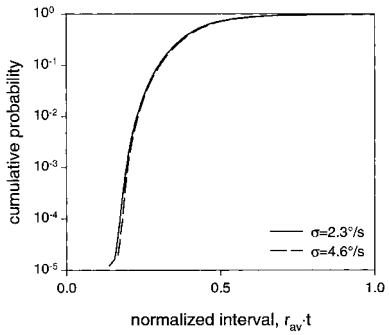

Figure 4. Interval Distributions with Slowly Varying Inputs

Cumulative distribution of intervals between successive spikes, measured in dimensionless units, $r_{av} \cdot t$, where r_{av} is the mean firing rate, and t is the interval between spikes. Data are presented from two experiments, where the stimulus distribution has standard deviations of $\sigma = 2.3°/s$ (solid line) and $\sigma = 4.6°/s$ (dashed line).

normalized rate (rate relative to the average over the experiment) also rescales among the two experiments.

The adaptive rescaling of the mean and variance of the neural response suggests that the system encodes stimulus fluctuations in units of the stimulus ensemble standard deviation. If the system is characterized by a nonlinear response function, then adaptive rescaling means it has an additional degree of freedom, a "stretch factor": the response function can be stretched or compressed by a constant factor to allow for incoming stimulus distributions of different widths. The rescaling breaks down for very large variances, in which the velocities are too rapid for the cell to follow, and the response drops dramatically.

If the neuron encodes its inputs in a normalized fashion, effectively scaling away the standard deviation or dynamic range of the input ensemble, then the statistical properties of the spike train—not just the time-dependent rates—should be invariant under changes of this standard deviation. This is a prediction about the structure of spike trains, independent of any assumptions about the nature of the code and its elementary symbols. Figure 4 shows the interspike interval distributions measured in the two steady states, plotted together in dimensionless time units. While the average firing rate changes from 69 to 75 spikes/s upon doubling the input standard deviation, the relative fluctuations are the same to high accuracy over several decades of probability.

Next, we consider an ensemble of horizontal velocity stimuli, s(t), with a time variation that is fast relative to the typical neural integration times. In this case, one cannot expect to find a simple relation between the momentary stimulus and the firing rate, since the temporal width of neural filtering becomes noticeable. The neural response is then determined by the local value of the filtered signal, and in general there can be more than one such filter. One would like to be able to uncover

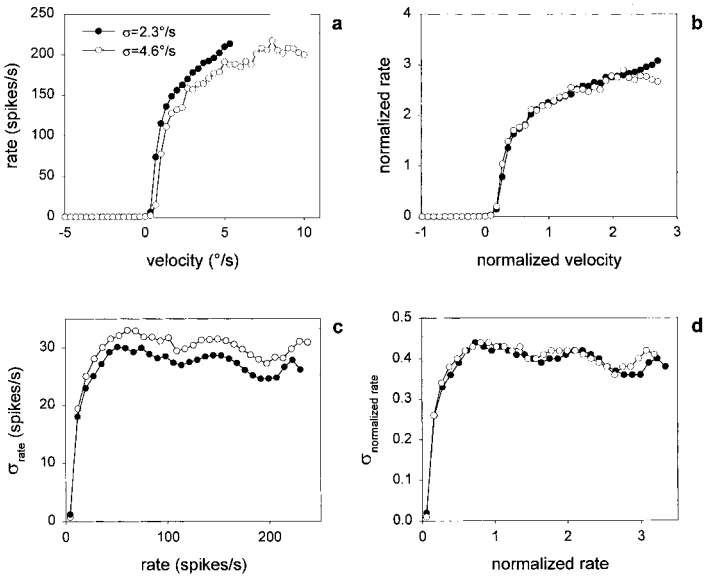

Figure 3. Signals and Noise

Average response functions and response variability in the two stimulus ensembles, $\sigma_1 = 2.3°/s$ (closed circles) and $\sigma_2 = 4.6°/s$ (open circles).

(a) Response in physical units, rate (spikes/s) as a function of stimulus velocity (degrees/s).

(b) Response in dimensionless units. The rate is normalized by the time-averaged firing rate and the velocity stimulus by the ensemble standard deviation.

(c) Variability in rate over trials as a function of average rate. The standard deviation in the firing rate was computed over the 180 trials for each 10 ms time bin and plotted as a function of the average rate in the bin.

(d) Variability in dimensionless units. The rate is normalized by the time-averaged rate.

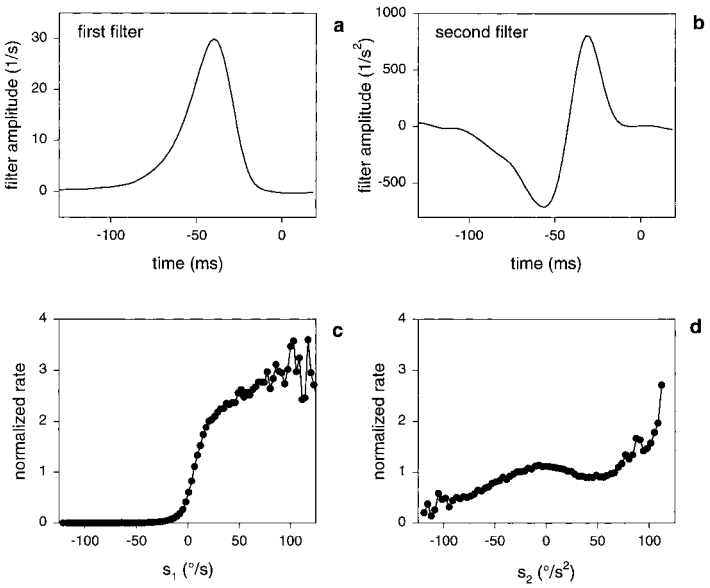

Figure 5. Velocity and Acceleration Sensitivity

The two dominant stimulus features that control the response of the H1 neuron (a and b) and the corresponding nonlinear neural response functions (c and d).

(a) The dominant filter is a smoothing filter, implying that H1 is sensitive to a smoothed version of the time-dependent velocity.

(b) The second filter is approximately the derivative of the first, implying that H1 is also sensitive to the smoothed acceleration. The two filters are normalized to units appropriate to their interpretation as velocity and acceleration. The firing rate is a nonlinear function of both stimulus dimensions, s_1 and s_2, and the one-dimensional projections of this function are shown in (c) and (d).

the relevant filters from the data, rather than to postulate them a priori, and this can indeed be done (see Procedures). By an extension of the reverse correlation method, we can show that the response of H1 is dominated by the time-dependent signal, as seen through two filters. Figures 5a and 5b depict the two filters, and they correspond to our intuition: the first filter smoothes the velocity signal over a window of about 50 ms, and the second filter is approximately the derivative of the first, corresponding to a smoothed acceleration.

Constructing the input/output relation (see Procedures), we describe the spike rate in H1 as a function of the two dominant stimulus dimensions, s_1 and s_2, corresponding to velocity and acceleration. For simplicity, we discuss here the two projections of this function separately; then, we have two input/output relations, $r(s_1)$ and $r(s_2)$, shown in Figures 5c and 5d. Note that although the filters defining s_1 and s_2 were found by linear analysis, the response functions are nonlinear. Consistent with the interpretation of s_1 as the smoothed velocity, the function in Figure 5c is qualitatively similar to that in the slowly varying limit (Figure 1b).

We performed experiments using rapidly varying stimuli with Gaussian statistics; the correlation time was 10 ms, and the standard deviation took four values, ranging from $\sigma = 18°/s$ to $\sigma = 180°/s$. Since the two filters are derived from the data in each case, there are generally some differences in the details of these filters for the different stimulus ensembles; however, they always have similar form and correspond to smoothed velocity and acceleration. Measuring the stimulus component s_1 in units of velocity, that is, degrees/s, we find that the

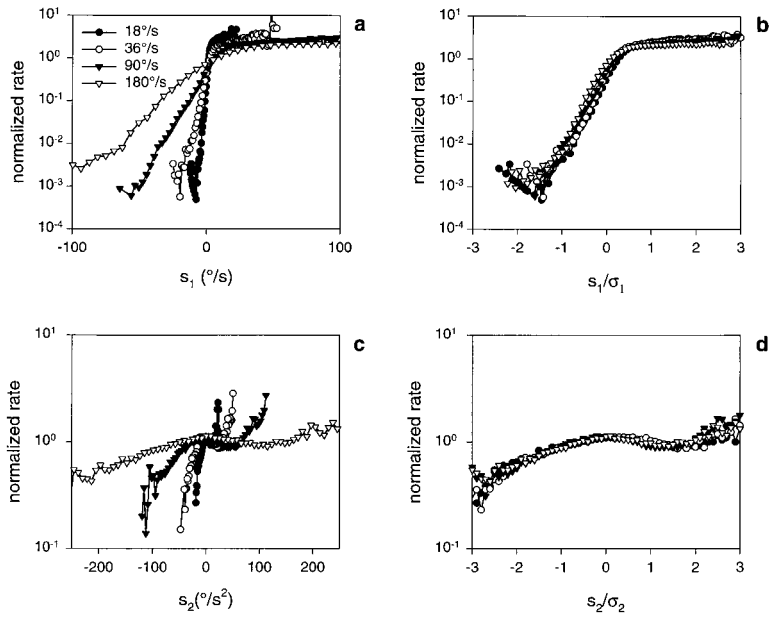

Figure 6. Rescaling of Responses to Dynamic Inputs

Adaptive rescaling of the input/output relations along the two leading dimensions.

(a and c) Response as a function of stimulus velocity as seen through the first (a) and second (c) filter (see Figure 5).

(b and d) Response as a function of stimulus projections, each normalized by its standard deviation.

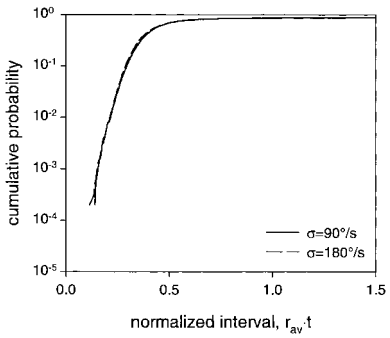

Figure 7. Interval Distributions with Rapidly Varying Inputs

Cumulative distribution of intervals between successive spikes, measured in dimensionless units, in two stimulus ensembles. Here, the stimulus varied rapidly—much faster than the typical interspike time. Data are presented from two experiments with stimulus standard deviations of $\sigma = 90°/s$ (solid line) and 180°/s.

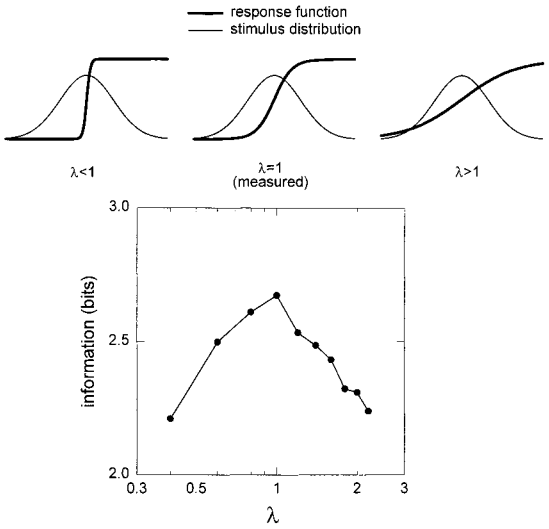

Figure 8. Optimizing Information Transmission

Information as a function of the stretch factor λ. The input/output relation measured in the experiment was artificially stretched or contracted by a factor λ, simulating the rescaling that occurs during adaptation. This is illustrated schematically in the three panels on top. For each value of λ, the stretched input/output relation and the distribution of stimuli used in the experiment determine a distribution of rates, which in turn determines the information with Equation 2. The point $\lambda = 1$ corresponds to the stretch factor measured in the experiment. The maximum at this point indicates that the process of adaptation selects a stretch factor that maximizes the information transmission.

input/output relations are very different, as shown in Figure 6a. Normalizing s_1 by its standard deviation, and the rate by its mean, these curves overlap (Figure 6b). The adaptive rescaling of the response function seen here is analogous to the effect described previously for the slowly varying stimulus. Because of better sampling, here we can observe the effect over ~3 decades in the response. A similar result is found for the input/output relation in the second dimension, $r(s_2)$, although the range of response variation along this dimension is smaller. Again, in physical units, the response functions look rather different (Figure 6c), but in normalized units, all of the curves from different ensembles overlap (Figure 6d).

Once again, adaptive rescaling predicts that higher order statistics of the spike trains are also invariant to changes in the dynamic range of the inputs. As a test of this, Figure 7 shows the interval distribution in two of the experiments. In dimensionless time units, the two interval distributions overlap across several decades of probability.

Functionality: Maximizing Information Transmission

Once we know that the neural input/output relation has the flexibility to rescale, a natural question is how the system "chooses" the stretch factor in response to the distribution of stimuli. Intuitively, stretching the input/output relation allows the system to match its limited dynamic range to the dynamic range of the inputs, as indicated schematically in Figure 8. This intuition is quantified by searching for coding strategies—in this case, stretching factors—that maximize the entropy of the distribution of outputs, as suggested by Laughlin (1981). More generally, one expects to find a matching that maximizes the information carried by the output about the input. This includes, in addition to output symbol entropy, the effect of the noise entropy or response variability (de Ruyter van Steveninck et al., 1997; Strong et al., 1998). While there has been much interest in optimization principles for neural coding and computation, there are few examples in which we can check directly that something is being optimized. Within the space of input/output relations parameterized by the stretch factor, however, we can test for optimization using only quantities that are measured in experiments.

The analysis of information requires identifying the elementary symbols of the code and their distribution in any given ensemble; this is straightforward for a rate code, and we present the results here for this case. Identifying the symbols is, however, more subtle for the case of rapidly varying stimuli (Brenner et al., 2000). A simplified analysis can be done in that case which leads to a similar conclusion. The elementary symbols of a quasistatic rate code are the momentary values of the firing rate, and we can find the entropy of these symbols from the distribution of signals in the experiment and the response function r(s), shown in Figure 3. We can also find what this entropy would be if the stretch factor were chosen differently, artificially stretching the response function by a factor of λ, and thus simulating the input/output relation observed when the system is adapted to a larger or smaller variance. A similar analysis can be done for the noise entropy, and in this way we can compare the transmitted information in the experiment ($\lambda = 1$) with the information that would be transmitted with a different choice of stretch factor ($\lambda \neq 1$) under the same stimulus ensemble (see Procedures, Equation 2). Figure 8 shows the information of the rate distribution obtained from this computation as a function of the stretch factor λ. The information exhibits a clear maximum at $\lambda = 1$, the value chosen by the system. These results demonstrate unambiguously the implementation of a maximum information principle in a sensory system.

Discussion

The idea that efficient signal processing systems must take into account properties of the distribution of incoming signals goes back many years (Wiener, 1949). One naturally expects to find this principle implemented in neural signal processing systems. In particular, the intermittent structure of natural signals suggests that the nervous system should adapt its strategies for coding and computation to local changes in the low order statistical properties of the sensory environment. Here, we characterized the neural response by a steady-state nonlinear input/output function and observed how this function changes among stimulus ensembles with different variances.

The nonlinear input/output relation used here is analogous to the velocity tuning curve widely used in the study of motion-sensitive visual neurons—it describes the probability of spiking (or spike rate) as a function of the relevant stimulus features. At present, there is considerable interest in situations in which the tuning curve or its spatial analog, the receptive field, is altered by the context in which stimuli are presented. The H1 neuron provides a clear and understandable example: the system stretches or compresses its tuning curve to match the dynamic range of incoming stimuli. This example suggests that, more generally, the tuning curve can change with the statistical properties of the distribution from which stimuli are drawn. In this picture, there is no single tuning curve characterizing the neuron but rather a set from which the neuron can choose the most suitable tuning curve for each context.

The stretching of the input/output relation effectively rescales or normalizes the inputs by their standard deviations. There is at least an intuitive connection between this phenomenon and the appearance of "normalization" in mammalian visual cortex (Carandini and Heeger, 1994), and it has been suggested that cortical normalization helps to provide an efficient code for natural sensory inputs (Simoncelli and Schwartz, 1999). Here, we have provided direct evidence that adaptive rescaling is used to maximize information transmission: among the set of available response functions, the "most suitable" is the one for which the information conveyed about the sensory stimulus is the highest.

The most important difference between differently scaled input/output functions is the distribution of the output symbols they determine under a given stimulus distribution. More specifically, the entropy of these symbols is sensitive to the stretching and compressing of the response. For the conditions explored here, the noise seems to be much less sensitive to the stretching and compression of the response, such that optimizing information transmission is almost equivalent to maximizing the entropy of the output symbols.

The principle of maximizing the entropy of the output symbols to enhance the information capacity of a neuron was discussed by Laughlin (1981) in connection with the responses of large monopolar cells in the fly visual system. In Laughlin's work, the input/output relation was characterized by measuring peak voltage responses to contrast steps, and matching was to the distribution of contrasts found in a natural scene. Presumably, this matching to global statistical properties of the visual world would have occurred over evolutionary or developmental timescales. Although these results have provided considerable inspiration for work on optimal visual coding, there have been few other examples in which such a direct test of optimization itself has been possible.

The present work adds to Laughlin's results in several ways. First, we are able to characterize input/output relations under more natural dynamic stimulus conditions. Second, we can measure directly the noise in the response and hence test for the optimization of information rather than the maximization of entropy. Finally, the results presented here pertain to adaptive processes that take place on much shorter timescales, of the order of seconds or minutes. Such adaptation to the local conditions in the laboratory implies that the system can "learn" a parameter of the signal distribution. This helps it to optimize its operation locally, providing a flexibility to alter its code as the environment changes.

An important issue is the timescale of the adaptation process: how long does it take to adapt to a distribution with a new variance? With rapidly varying signals, we can estimate this timescale by an experiment in which the variance is switched between two values. We find the adaptation time to be several seconds, with an asymmetry between increasing and decreasing variance, as seen also in the vertebrate retina (Smirnakis et al., 1997). Adaptation to the stimulus variance or other statistics requires that the nervous system estimate these statistics, at least implicitly, and reliable estimation takes time; it has been suggested that the dynamics of adaptation are connected with the dynamics of this estimation process (DeWeese and Zador, 1998). Preliminary experiments suggest that there may not be one single timescale for the adaptation process and that the connection of these dynamics to the estimation problem may be more subtle (A. Fairhall et al., personal communication).

The idea that there are multiple timescales for adaptation is supported by the absence of simple exponential recovery even in the earliest experiments on the adaptation of H1 to constant velocity input (de Ruyter van Steveninck et al., 1986) and by the fact that such similar phenomenology is seen for adaptation to the variance of both fast and slow stimuli. This multiplicity of timescales, like the rescaling phenomenon itself, may be another adaptation to the statistics of signals in the natural world.

Procedures

Neural Recording

A female blowfly was immobilized, and a small hole was cut in the back of the head, close to the midline on the right side. Through this hole, a tungsten electrode was advanced into the lobula plate. This area, which is several layers back from the compound eye, includes a group of large motion detector neurons with wide receptive fields and strong direction selectivity. We recorded spikes extracellularly from one of these, the contralateral H1 neuron (Franceschini et al., 1989; Hausen and Egelhaaf, 1989). A simple threshold discriminator converted the spikes into spike times digitized at a 10 μs resolution.

Stimulus Generation

The stimulus was a rigidly moving bar pattern of average intensity, about 100 mW/($m^2 \cdot sr$), displayed on a Tektronix 608 high brightness

display. The bars were oriented vertically, with intensities chosen at random to be one of two values. The fly viewed the display through a round 80 mm diameter diaphragm showing ~30 bars. Frames of the stimulus pattern were refreshed every 2 ms, and with each new frame, the pattern was displayed at a new position. This resulted in an apparent horizontal motion of the bar pattern that is suitable for exciting the H1 neuron. The pattern position was defined by a pseudorandom sequence simulating a motion trajectory drawn from a probability distribution. The sequence was then multiplied by a different constant in each experiment to give a different value of the standard deviation. Filtering was used to obtain the slowly varying stimulus ensemble.

Information Estimates and Optimality Test

The information that a neuron transmits about the sensory input can be computed by finding the entropy of the distribution of output symbols and subtracting the entropy of the noise or, equivalently, the entropy of the output, given that the input is fixed. In the limit that noise is small (as suggested by the results in Figure 3), the distribution of output symbols can be found by taking the distribution of inputs P(s) and passing it through the input/output relation r(s),

$$P(r) = \sum_{s \to r} P(s), \tag{1}$$

where the sum runs over all signals, s, that are mapped into the response r through r(s). To compute the noise entropy, we need more than the standard characterization of response variance (as in Figures 3c and 3d); we need the entire distribution of the response variability. Since the input/output relation r(s) is invertible, it doesn't matter if we view the distribution of noise as being dependent on s or on r, and so we can write the transmitted information as

$$I = S[P(r)] - \sum_{r} P(r) S[P(n|r)], \tag{2}$$

where S[P] is the entropy of the distribution P, and P(n|r) is the distribution of the noise given the mean response. Note that the distribution of signals P(s) is controlled in our experiment, whereas the response function r(s) is measured, as in Figure 1b, and similarly we measure the distribution P(n|r). Since none of these distributions have any simple analytic form, the entropies were estimated directly by binning the experimental data rather than by any analytic approximation. Our data sets were sufficiently large that we could make small bins and verify that changes in binning had a negligible effect on the results.

We have found that the response function r(s) has the flexibility to change the stimulus scale by a stretch factor, λ; choosing different values for λ results in a differently scaled response function, and this, in turn, affects the value of the entropies in Equation 2. We used the empirical forms of r(s) and P(s) and simulated the stretch factor λ artificially, as illustrated in Figure 8, by the replacement $r(s) \to (\lambda s)$. The measured response function corresponds to $\lambda = 1$; stretching the response by λ and using Equations 1 and 2, we can find the entropy that would result from a different choice of λ. In Figure 8, the squeezed response functions were extrapolated to a constant at the high end of firing rate, and it was checked that a linear extrapolation gives similar results in the optimality test. As noted in the text, the variation of I with λ is dominated by the entropy of responses, with the noise entropy nearly constant, such that optimizing information transmission is almost the same as maximizing the output entropy.

Stimulus Dimensionality Reduction

Once adaptation processes have reached a steady state, neural responses depend on the history of the stimulus in a finite fixed time window. Previous experience with H1 suggests that, in response to rapidly varying velocity signals, spiking is correlated with velocity signals within an integration time of 100 ms or less. In our experiments, a new velocity is defined every 2 ms, and so in principle, the firing rate could depend on 50 different parameters describing the details of the motion trajectory within one integration time. Taking into account the correlation time of 10 ms, the firing could still depend on as many as 10 parameters. We would like to identify a small number of features, or stimulus dimensions, that are most relevant to the neural responses.

In the simplest model, the spike rate depends on the velocity trajectory, as seen through a filter. We can think of the many parameters describing the motion trajectory as being the dimensions of a vector space, such that each possible stimulus trajectory is a vector. Then, the simplest model is that the firing rate depends only on the projection of this vector along one single direction in space. This special direction corresponds to the form of the filter that the system uses for smoothing or averaging. If this simplest model is correct, then the reverse correlation technique (de Boer and Kuyper, 1968; Rieke et al., 1997) can find the one special direction in stimulus space. More precisely, if we choose input signals from a distribution corresponding to Gaussian white noise, then the average stimulus preceding a spike (spike-triggered average stimulus) "points" in the special direction that describes the smoothing filter. This analysis assumes a priori that only one projection is important; it therefore cannot provide any evidence for or against this hypothesis. Furthermore, if two projections are important—smoothed versions of velocity and acceleration, for example—the reverse correlation method confounds these different stimulus dimensions.

If spiking is related not only to one but to a few stimulus dimensions, then this reduced dimensionality may be found by characterizing the shape of the region in stimulus space that a spike "points to." A strategy for quantitating this intuition was suggested in earlier work (de Ruyter van Steveninck and Bialek, 1988). Instead of computing the average stimulus that precedes a spike, we compute the covariance matrix of the fluctuations around the average, defined as

$$\begin{aligned} C_{spike}(\tau,\tau') &= \langle s(t_{spike} - \tau) \cdot s(t_{spike} - \tau') \rangle - \langle s(t_{spike} - \tau) \rangle \cdot \langle s(t_{spike} - \tau') \rangle \\ &= C_{prior}(\tau, \tau') + \Delta C(\tau, \tau'), \end{aligned} \tag{3}$$

where

$$C_{prior}(\tau,\tau') = \langle s(t - \tau) \rangle \langle s(t - \tau') \rangle \tag{4}$$

is the correlation matrix of the visual stimulus itself. If the stimulus is Gaussian, and if the firing rate is determined by a small number, K, of its projections, then the matrix $\Delta C(\tau,\tau')$ has rank K. Moreover, the eigenvectors of this matrix, which are associated with the nonzero eigenvalues, correspond to linear combinations of the relevant vectors (filters). Identifying these directions provides us with a coordinate system that spans the set of relevant projections. Note that these directions are not the principal components either of C_{prior} or of C_{spike}.

In practice, we collected 120 ms segments of the stimulus waveform surrounding each spike in the experiment, sampled at the 2 ms resolution of the experiment. From these samples, we computed C_{spike}, and by sampling the distribution of stimuli without reference to spiking events, we computed C_{prior} and then formed the difference ΔC. This has units of (stimulus)^2, so if we normalize by the stimulus variance, ΔC is dimensionless. For a typical experiment ($\sigma = 90°$/s), we find that the second largest eigenvalue of ΔC is 14% of the leading one in magnitude, while the third is only 3%, and the rest are <0.6%. The emergence of only two dominant dimensions from the data provides strong evidence that the spiking of H1 is sensitive to only two projections of the time-dependent velocity signal.

As mentioned in the text, we interpret the two leading stimulus projections as smoothed velocity and smoothed acceleration. Accordingly, we normalize the first filter such that for a constant velocity, the value of the filtered signal is the same constant. We normalize the second filter such that for a velocity growing with a constant acceleration, the filtered signal is the constant acceleration.

Constructing the Nonlinear Input/Output Relations

The input/output relation gives the spike rate, or probability of spiking, as a function of the stimulus value. In the case of a rapidly varying stimulus, we have seen that there are two stimulus values relevant for spiking in H1, the projections s_1 and s_2. Defining $P(\text{spike}|s_1)$ as the probability of spiking conditional on the first projection of the stimulus having the value s_1, it follows from Bayes's rule that

$$\frac{P(\text{spike}|s_1)}{P(\text{spike})} = \frac{P(s_1|\text{spike})}{P(s_1)}. \quad (5)$$

Recalling that the spike rate is proportional to this probability,

$$r(s_1) \propto P(\text{spike}|s_1), \quad (6)$$

we find

$$\frac{r(s_1)}{r_{av}} = \frac{P(s_1|\text{spike})}{P(s_1)}, \quad (7)$$

where r_{av} is the average spike rate. The conditional probability distribution $P(s_1|\text{spike})$ describes the stimulus projection, s_1, that a spike "points to" (de Ruyter van Steveninck and Bialek, 1988). To sample this distribution directly from the data, we look back at the stimulus every time that we observe a spike and find the value of the stimulus projection on the first filter. The distribution $P(s_1)$ is known: it is the distribution of stimuli presented in the experiment. Finally, having estimated the two distributions, $P(s_1|\text{spike})$ and $P(s_1)$, we form the ratio as in Equation 7 to give the nonlinear input/output relation. The same procedure can be followed in the two-dimensional space (s_1,s_2), resulting in a two-dimensional input/output relation $r(s_1,s_2)$. In this work, we considered for simplicity the two projections of this function, $r(s_1)$ and $r(s_2)$.

Acknowledgments

We thank Geoff Lewen for his contributions to the experiments and Simon Laughlin for the many discussions that stimulated this work.

Received August 24, 1999; revised April 28, 2000.

References

Adrian, E. (1928). The Basis of Sensation (London, UK: Christopher's).

Atick, J. (1992). Could information theory provide an ecological theory of sensory processing? In Princeton Lectures on Biophysics, W. Bialek, ed. (Singapore: World Scientific), pp. 223–289.

Attneave, F. (1954). Some informational aspects of visual perception. Psych. Rev. *61*, 183–193.

Barlow, H.B. (1961). Possible principles underlying the transformation of sensory messages. In Sensory Communication, W.A. Rosenblith, ed. (Cambridge, MA: MIT Press), pp. 217–234.

Borst, A., and Egelhaaf, M. (1987). Temporal modulation of luminance adapts time constant of fly movement detectors. Biol. Cybern. *56*, 209–215.

Brenner, N., Strong, S., Koberle, R., Bialek, W., and de Ruyter van Steveninck, R.R. (2000). Synergy in a neural code. Neural Comput., in press.

Carandini, M., and Heeger, D.J. (1994). Summation and division by neurons in primate visual cortex. Science *264*, 1333–1336.

de Boer, E., and Kuyper, P. (1968). Triggered correlation. IEEE Trans. Biomed. Eng. *15*, 169–179.

de Ruyter van Steveninck, R., and Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc. R. Soc. Lond. B Biol. Sci. *234*, 379–414.

de Ruyter van Steveninck, R., and Bialek, W. (1996). Optimality and adaptation in motion estimation by the blowfly visual system. In Proceedings of the IEEE Twenty-Second Annual Northeast Bioengineering Conference, J.K.-J. Li and S. S. Reisman, eds (Piscataway, NJ: IEEE), pp. 40–41.

de Ruyter van Steveninck, R.R., Zaagman, W.H., and Mastebroek, H.A.K. (1986). Adaptation of transient responses of a movement-sensitive neuron in the visual system of the blowfly *Calliphora erythrocephala*. Biol. Cybern. *54*, 223–236.

de Ruyter van Steveninck, R.R., Bialek, W., Potters, M., and Carlson, R.H., and Lewen, G.D. (1996). Adaptive movement computation by the blowfly visual system. In Natural and Artifical Parallel Computation, Proceedings of the Fifth NEC Research Symposium, D.L. Waltz, ed. (Philadelphia: SIAM), pp. 21–41.

de Ruyter van Steveninck, R.R., Lewen, G.D., Strong, S.P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. Science *275*, 1805–1808.

DeWeese, M., and Zador, A. (1998). Asymmetric dynamics in optimal variance adaptation. Neural Comput. *10*, 1179–1202.

Franceschini, N., Riehle, A., and le Nestour, A. (1989). Directionally selective motion detection by insect neurons. In Facets of Vision, R.C. Hardie and D.G. Stavenga, eds. (Berlin: Springer-Verlag), pp. 360–390.

Hausen, K., and Egelhaaf, M. (1989). Neural mechanisms of visual course control in insects. In Facets of Vision, R.C. Hardie and D.G. Stavenga, eds. (Berlin: Springer-Verlag), pp. 391–424.

Laughlin, S.B. (1981). A simple coding procedure enhances a neuron's information capacity. Z. Naturforsch. *36c*, 910–912.

Maddess, T.M., and Laughlin, S.B. (1985). Adaptation of the motion-sensitive neuron H1 is generated locally and governed by contrast frequency. Proc. R. Soc. Lond. B Biol. Sci. *225*, 251–275.

Meister, M., and Berry, M.J. (1999). The neural code of the retina. Neuron *22*, 435–450.

Nelken, I., Rotman, Y., and Yosef, O.B. (1999). Response of auditory-cortex neurons to structural features of natural sounds. Nature *397*, 154–156.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). Spikes: Exploring the neural code (Cambridge, MA: MIT Press).

Ruderman, D.L., and Bialek, W. (1994). Statistics of natural images: scaling in the woods. Phys. Rev. Lett. *73*, 814–817.

Shapley, R.M., and Victor, J.D. (1979a). The contrast gain conrol of the cat retina. Vision Res. *19*, 431–434.

Shapley, R.M., and Victor, J.D. (1979b). Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells. J. Physiol. *290*, 141–161.

Shapley, R.M., and Victor, J.D. (1980). The effect of contrast on the non-linear response of the Y cells. J. Physiol. *302*, 535–547.

Simoncelli, E.P., and Schwartz, O. (1999). Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In Advances in Neural Information Processing Systems, Eleventh Edition, M.S. Kearns et al., eds. (Cambridge, MA: MIT Press), pp. 153–166.

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.

Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., and Bialek, W. (1998). Entropy and information in neural spike trains. Phys. Rev. Lett. *80*, 197–200.

van Hateren, J.H. (1997). Processing of natural time-series of intensities by the visual system of the blowfly. Vision Res. *37*, 3407–3416.

Wiener, N. (1949). Time Series (Cambridge, MA: MIT Press).