Efficiency and ambiguity in an adaptive neural code

Adrienne L. Fairhall, Geoffrey D. Lewen, William Bialek & Robert R. de Ruyter van Steveninck

NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

We examine the dynamics of a neural code in the context of stimuli whose statistical properties are themselves evolving dynamically. Adaptation to these statistics occurs over a wide range of timescales—from tens of milliseconds to minutes. Rapid components of adaptation serve to optimize the information that action potentials carry about rapid stimulus variations within the local statistical ensemble, while changes in the rate and statistics of action-potential firing encode information about the ensemble itself, thus resolving potential ambiguities. The speed with which information is optimized and ambiguities are resolved approaches the physical limit imposed by statistical sampling and noise.

More than forty years ago it was suggested that neural codes may constitute efficient representations of the sensory world[1,2]. Within the framework of information theory, efficient coding requires a matching of the coding strategy to the statistics of the input signals[3]. Recent work shows that the sequences of action potentials from single neurons provide an efficient representation of complex dynamic inputs[4–7], that adaptation to the distribution of inputs can occur in real time[8–10] and that the form of the adaptation can serve to maximize information transmission[10,11]. However, adaptation involves compromises. An adaptive code is inherently ambiguous: the meaning of a spike or a pattern of spikes depends on context, and resolution of this ambiguity requires that the system additionally encode information about the context itself. In a dynamic environment the context changes in time and there is a tradeoff between tracking rapid changes and optimizing the code for the current context. Here we examine the dynamics of adaptation to statistics in motion-sensitive cells in the visual system of a fly, *Calliphora vicina*, and find that different aspects of adaptation occur on timescales that range from tens of milliseconds to several minutes. The speed of adaptation to a new input distribution must be limited by the need to collect statistics. Adaptation of the neural input/output relation to optimize information transmission approaches this theoretical maximum speed. This rapid adaptation of the input/output relation leaves the longer timescales in the response dynamics as a nearly independent channel for information, resolving potential ambiguities of an adaptive code.

Multiple timescales in statistical adaptation

Codes are defined by the response not just to a particular stimulus, but to a statistical distribution of inputs, which provides the context. Statistical adaptation involves matching the neural code to this input distribution. If the input signals are gaussian the only quantities that can govern statistical adaptation are the mean (as in light–dark adaptation in the retina) and the correlation function. Here we consider signals with a fixed, short correlation time and zero mean, so that the only free statistical parameter is the variance. We record action potentials from an identified motion-sensitive neuron in the fly visual system (H1 (refs 12, 13)). The input signal is the angular velocity of motion across the visual field; a sudden change in variance might mimic what happens as a fly makes a transition from cruising flight[14] ($S_{\mathrm{r.m.s.}} \sim 230$ degrees s^{-1}) to acrobatic or chasing behaviour[15] ($S_{\mathrm{r.m.s.}} \sim 1{,}200$ degrees s^{-1}). To track the cell's response to changes in the local stimulus statistics, we modulate a (nearly white) zero-mean signal $z(t)$ in time with a slowly varying amplitude $\sigma(t)$, such that the angular velocity stimulus is given by $S(t) = z(t)\sigma(t)$ (see Methods). Thus the stimulus has a time-varying distribution that is locally characterized solely by its (varying) standard deviation.

A simple probe of adaptation to statistics is a switching experiment[8], shown in Fig. 1a, where the standard deviation switches periodically between σ_1 and σ_2, with a cycle time of T; we choose a new random sequence $z(t)$ in each cycle. Averaging over many cycles we obtain the spike rate as a function of time, which reflects the dynamics of the response to $\sigma(t)$. As in related experiments[8,16], the rate after an abrupt increase in σ first jumps to a high value, then gradually decreases almost to a steady state over the period of the switch (Fig. 1c). It is tempting to identify this transient as the (unique) timescale of adaptation, but as the experiment is repeated with increasing periods T, from 4 to 40 s, the apparent decay time also increases (Fig. 1b). In fact, the decay time scales linearly with the switching period T (Fig. 1c). This scaling holds for the transient after increases but not decreases in variance, suggesting that different mechanisms may be at work in the different directions of adaptation. Similarly, in response to sinusoidal modulations in $\sigma(t)$, the rate is shifted in time with respect to $\sigma(t)$ by a fixed fraction of the modulation period T, for T from 10 to 240 s (data not shown; see ref. 17). The adaptation of the firing rate clearly has access to many timescales[18], and the observed timescale depends on the design of the experiment.

Input/output relations and information transmission

If the spike rate changes when stimuli remain statistically the same—as for adaptation after a switch—the system's rule for generating a spike must be changing; we will quantify this by constructing an input/output relation. Under stationary conditions, the input/output relation of H1 is adapted to the stimulus ensemble in such a way that the stimulus input is rescaled by its standard deviation[10]. It is natural to assume that slow transients in the rate dynamics are indicative of or caused by the progress of this adaptation, so we follow the evolution of the input/output relation after a sudden change in statistics, as in the switching experiment.

Even for statistically stationary signals, the spiking probability can be affected by the history of the motion stimulus in a 400-ms window preceding the spike. Thus 'the stimulus' at time t really is a high-dimensional object describing the whole history $S(t)$ for $t_{\mathrm{spike}} - 400\,\mathrm{ms} < t < t_{\mathrm{spike}}$. To study the input/output relation we require a lower-dimensional description of the stimulus, and, as in ref. 10, we do this by linearly projecting or filtering the signal $S(t)$ to obtain a single number s for each spike time (see Methods). The most relevant projection is a smoothed version of the velocity. Although there are interesting adaptations in these linear filters that are analogous to those previously observed[9,19,20]—possibly also

implicated in information maximization, as has been suggested in other cases[9,21]—our focus here is on the nonlinear relation between the spike-firing probability and the smoothed velocity signal.

The measured input/output relations are shown in Fig. 2 for selected points in the cycle. When measured in physical units, the input/output relations in the two halves of the cycle differ strongly: when the dynamic range of the signal is large, significant modulations in firing probability require proportionately large variations in the signal; when the dynamic range is small, a much smaller change in input evokes the same modulation in output. The effect is large: under different adaptation conditions the firing probabilities in response to the same input can differ by orders of magnitude. Thus, the input/output relation is not a static property of the system, but adapts, dynamically, so that the input seems to be measured in units proportional to the local standard deviation.

Strikingly, the differences in the input/output relation are established within the first second after the switch, by which point the input/output relation (and the locally computed linear filter; data not shown) is already very similar to the steady-state form (Fig. 2). More precisely, the transient and steady-state input/output relations differ simply by a scale factor in the spike rate, as seen by comparison with the curves shown in normalized units, where the transient and steady-state curves collapse. This scale factor corresponds to an overall modulation by the time-varying spike rate—the scaling with respect to the input remains the same. Thus the

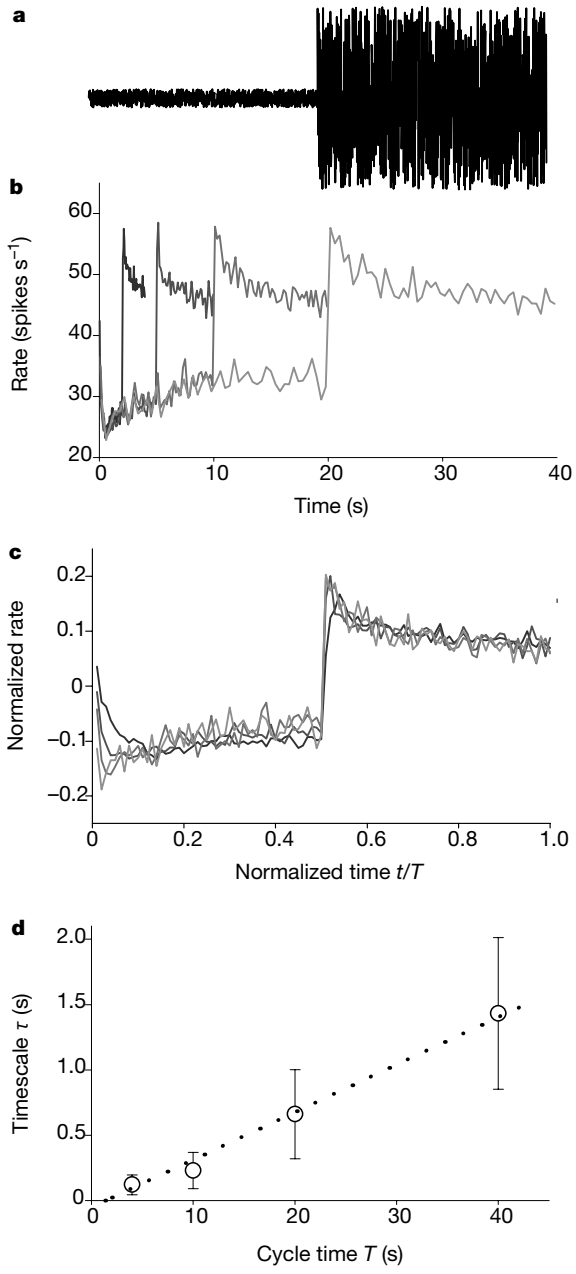

Figure 1 Variance-switching experiment. **a**, The stimulus is a white-noise velocity signal modulated by a square wave envelope that switches between two values, σ_1 and σ_2, with period T. **b**, Averaged rate as a function of time relative to the cycle, for $T = (4, 10, 20, 40)$ s, represented by a progressively lighter grey scale. **c**, The same curves as in **b** but with the time axis normalized by the period T and with the rate normalized to a mean of zero and unit standard deviation. **d**, The decay timescale τ, extracted from the curves of **a** by fitting an exponential to the decay of spike rate after an upward switch in variance, plotted as a function of T.

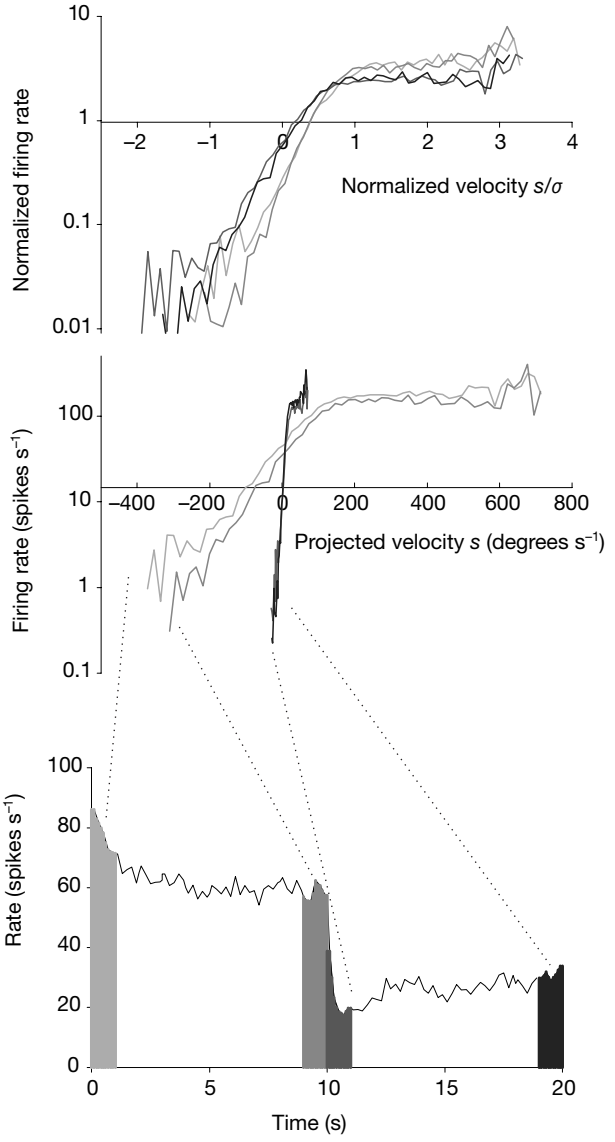

Figure 2 Input/output relations from the switching experiment. Input/output relations were constructed from the 1-s-wide time windows indicated in the rate plot. For the two variance conditions, we compare a window in the initial transient regime after a transition with the final second of the segment, representing the steady state. For a given value, the input/output relation in the transient is identical up to a multiplicative factor to that in the final window, showing that the rescaling of the input by variance is completed in under 1 s. The multiplicative factor stems from the modulation of the rate by the slowly varying adaptive process. The top graph shows the same plots in normalized units. For a given variance, the curves now overlay. The curves for the two different variances are very similar, differing only by a slight shift.

slow, transient change in spike rate after a sudden change in statistics is associated with an evolution of the input/output relation, but only in that the output is multiplied by a factor; the rescaling of the input seems to happen almost immediately on a timescale much shorter than the change in rate. If adaptation involved a single process—a change in neural threshold or the adjustment of inhibitory inputs—then the continuous changes in spike rate during the adaptation transient would be accompanied by a continuous stretching or compression of the input/output relation. Instead the stretching happens rapidly, and scaling of the output is independent of the processes that rescale the inputs. The rate dynamics seem to be a separate degree of freedom in the coding scheme, and may be an independent channel for information transmission.

Rescaling of inputs serves, under stationary conditions, to optimize information transmission[10]; it would be attractive if we could observe this optimization dynamically during the adaptation transient. Previous efforts[16,22] to measure information transmission during adaptation have used the stimulus reconstruction method[23]. Here we use a direct method[6] that estimates the local information transmission rate without any assumptions about the structure of the code and allows us to follow very rapid changes (see Methods). The information transmission as a function of time throughout a switching cycle, with a switch at $t = 0$, is shown in Fig. 3. First, the information per spike is a constant 1.5 bits in the two halves of the experiment, suggesting that maintaining a fixed information per spike may be a design principle of the system[24]. After a downward switch, the information per spike dips briefly and recovers with a time constant of ~40 ms, returning to its maximal level within 100 ms—much faster than the return of the spike rate to steady state. After an upward switch, there is no detectable transient in the information transmission. Adaptation to variance changes requires that the system gather evidence that signals come from a new probability distribution and are not just outliers in the old distribution. The speed of this sampling is limited by the noise level in the system. As discussed below (see Methods) the minimal source of noise in the photoreceptors sets a physical limit of about 40 ms for the adaptation timescale after a decrease in variance given the conditions of our experiments. Adaptation after an increase in variance can happen more rapidly[25]. The recovery timescale that we derive theoretically and observe experimentally is remarkably fast, given that the stimulus correlation time is around 10 ms as a result of filtering by the photoreceptors: the system receives barely a few independent samples from the new ensemble. Both the timescale of the information recovery and the asymmetry between increasing and decreasing variance are consistent with adaptation at a speed near the limit set by physical and statistical considerations.

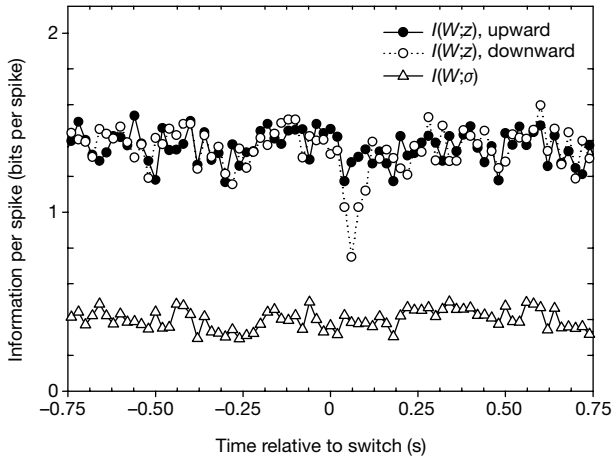

Figure 3 Time dependence of the information transmission. The information/spike was computed using a switching experiment in which the fast component $z(t)$ of the stimulus is a set of 40 fixed, white-noise 'probe' segments of 2 s in duration, presented in random order. These were modulated by a square wave envelope for a period of 4 s. The probes were out of phase with the modulation so that each probe straddled a transition from σ_1 to σ_2. The response of the cell was represented by binary 'words' from ten consecutive 2-ms time bins. The presence of a spike in a single bin is represented by 1, the absence by 0. We compute how much information/word the distribution of words at time t (relative to the switch) conveys about which probe was used, and how much information/word is transmitted about the variance σ (see Methods). The information/word is divided by the mean number of spikes per word to obtain the information/spike. The information/spike is plotted as a function of time surrounding an upward and a downward switch in σ. After a downward switch, there is a dip in the information transmission for a duration of <100 ms; after an upward switch, no change in the information transmission rate is detectable. The steady-state information/spike is the same in the two regimes, despite significant differences in spike rate. The information/spike conveyed about σ is also shown—it remains at a steady rate throughout the experiment.

Resolving ambiguity

The fast recovery of information about the rapid variation in the

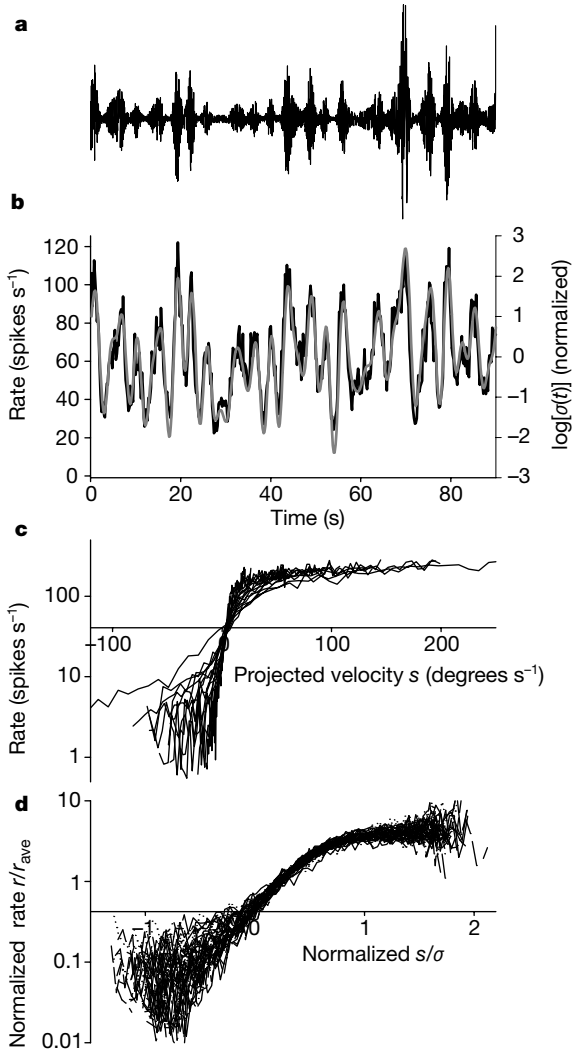

Figure 4 A stimulus with randomly modulated variance. **a**, The stimulus is a white-noise velocity signal modulated by an exponentiated random gaussian process, with correlation time 3 s and cycle length 90 s. **b**, Averaged rate. Overlaid is the logarithm of $\sigma(t)$ for this experiment, normalized to zero mean and unit standard derivation. **c**, Input/output relations from selected windows of width 3 s throughout the experiment. **d**, The same input/output relations as in **c**, rescaled along the x-axis by the local standard deviation, and along the y-axis by the average rate.

stimulus means that one of the possible problems of adaptation is solved, that is, the slow relaxation of the spike rate does not correspond to a similarly long period of inefficient coding. However, the rapid adjustment of the neuron's coding strategy could exacerbate the problem of ambiguity: if the cell encodes a continuously rescaled version of the input signal, how can an observer of the spike train know the overall scale of the stimulus? Although it is possible that other neurons in the network might encode the scale factor, our information-theoretic analysis suggests that H1 itself retains this information; at least in the context of the switching experiment the spike train provides information about the variance at a nearly constant rate per spike (Fig. 3). However, it would be useful if one could extract this information to generate a running estimate of the stimulus variance. An approach to this problem is shown in Fig. 4, where we deliver stimuli in which the variance itself varies at random with a correlation time of 3 s. The result is an intermittent stimulus that is locally gaussian, capturing some of the statistical properties of natural signals in vision[26] and in audition[27]. Under these conditions the spike rate provides an almost linear readout of the logarithm of the variance. At the same time, as before, the system shows continuous adaptation: the input/output relations were calculated in bins throughout the experiment (Fig. 4c), and when the input is normalized by the local standard deviation, the curves coincide (Fig. 4d). This suggests that fine structure of the normalized signal is encoded by the precise placement of spike patterns, generating the information $I(W;z)$ that we measure in Fig. 3, whereas the normalization factor or overall dynamic range of the signal is coded by slower variations in spike rate.

With a sudden change in variance, however, the spike rate remains ambiguous because the representation of the variance is confounded with the time since the last sudden change. Indeed, the stimulus-dependent decay time of the rate could be viewed as coding the time since the switch[28]. Although the rate may not encode the variance well, this information is present in other statistics of the spike train, as hinted at in ref. 29. Instead of averaging to obtain the time-dependent spike rate, we can collect, for example, all of the interspike intervals, sorted by the order in which they occur after the switch. More than a second after the switch, when the interval distributions (Fig. 5) have reached a steady state, these steady-state distributions for the two variances are distinct—the mode interval is shifted (Fig. 5); these and other data[10] indicate that the interval distribution provides a 'fingerprint' for the variance, an example of distributional coding[30]. Figure 5 further shows that by the second or third interval, the distributions are nearly indistinguishable from the steady state. Thus, even while the rate is changing slowly, providing ambiguous information about the stimulus variance, the interspike intervals become typical of the new variance after just a few spikes. We can construct a simple readout scheme for deciding whether the variance is large or small (see Methods), and this readout reaches reliable discrimination with just a few spikes (Fig. 5c), corresponding to a decision time of only 43 ms after a switch to larger variance or 120 ms after a switch to lower variance. These timescales are comparable to those for recovery of information about the normalized stimulus (Fig. 3), and the limit set by the need to gather statistics about the noisy inputs. We conclude that within 100 ms the system both adjusts its coding strategy to optimize information about the rapid (normalized) stimulus variations and provides an unambiguous signal about the overall stimulus scale.

It may seem paradoxical that the spike rate changes over a timescale of seconds when the interspike interval distribution reaches steady state in roughly 100 ms. This can be understood from the structure of the interval distributions, which have three main features: (1) an exponential tail at long intervals; (2) a peak at a preferred interval; (3) a refractory 'hole' at small intervals. A small change in the weight of the tail, or in its slope, can cause significant changes in spike rate, thus the interval distribution can appear almost stationary while the rate changes, and conversely counting spikes (to measure the rate directly) can obscure large changes in the structure of the distribution at shorter times. Both the preferred interval and the shape of the refractory hole depend strongly on the stimulus, in particular on its variance, and these are the features that provide the basis for rapid, reliable discrimination of the variance change (Fig. 5). These effects can be seen even in the response of simple conductance-based models for single neurons responding to injected currents (B. Agüera y Arcas; personal communication) and

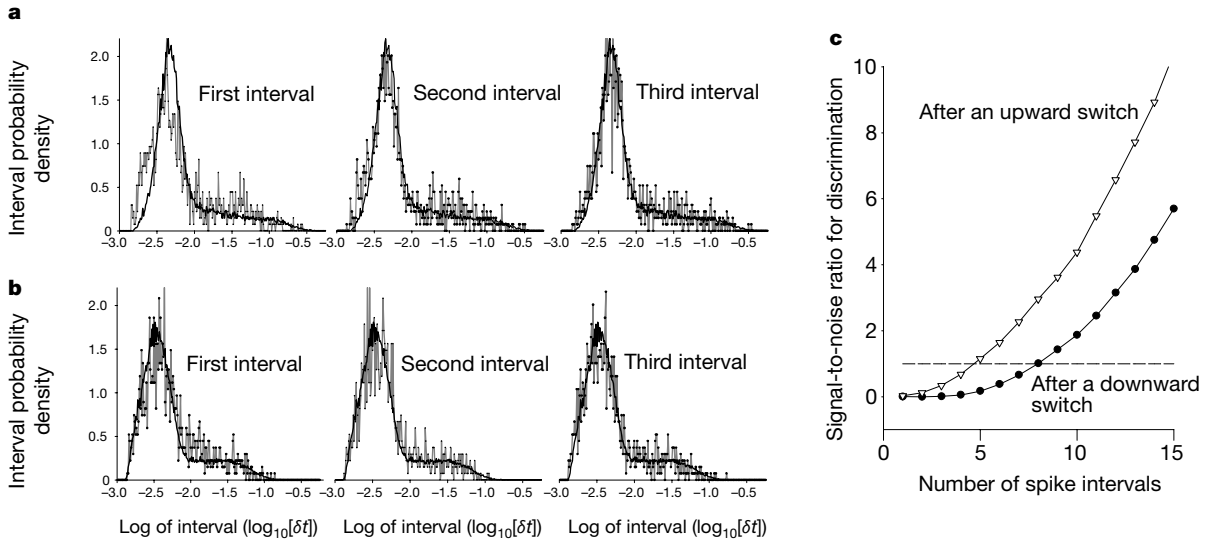

Figure 5 Discrimination using the statistics of interspike intervals in a switching experiment. Intervals were measured for the first several spikes after a switch in the variance, both upward and downward. The intervals shown are all measured in logarithmic units. **a**, The evolution of the interspike-interval probability density is shown for the first, second and third spike intervals after a downward switch, compared with that of all intervals collected after 1 s has elapsed since the switch (the steady-state distribution, in black). The densities for the individual, sequenced intervals are shown in grey. **b**, After an upward switch, the interval density strongly resembles the steady-state density by the first interval. **c**, The signal-to-noise ratio for discrimination using n spike intervals (see Methods) as a function of the number of intervals after an upward and a downward switch.

in networks of model neurons[31]. This emphasizes that the relative probability of short intervals need not be determined by a fixed absolute or relative refractory mechanism, but can respond to input signals and provide information able to resolve the ambiguities introduced by adaptation.

Discussion

These results suggest a multilayered coding scheme that allows the spike train to convey information about the stimulus through several channels independently: the timing of individual spikes or short spike patterns encodes stimulus features that are normalized to the stimulus ensemble, the statistics of interspike intervals on slightly longer timescales encode the stimulus ensemble, and the spike rate can carry information about the changes in ensemble on yet longer timescales. Although we are not yet able to identify specific mechanisms at work in the visual system of the fly to produce these dynamics, any mechanism must display the existence of multiple timescales, from the <100 ms recovery of information (Fig. 3) to the minute-long relaxation of the firing rate (Fig. 1). Recent biophysical studies[32] indicate that such a range ($\sim 10^3$) of timescales can arise even in a single cell from the complex inactivation kinetics of NaII channels. Although multiple timescales could arise also from network dynamics, it is appealing that important aspects of adaptive coding could be implemented by cellular mechanisms. A complementary question concerns the readout of information, for example from the distribution of intervals. This is a classical problem in neural coding, and it has been demonstrated that single neurons are sensitive to the temporal pattern of synaptic inputs[33]. Recent work shows how the diverse dynamics of synapses can extract and transmit different aspects of a spike train[34,35], providing specific mechanisms for readout of a distributional code.

Many features of adaptation in H1 have analogues in the mammalian visual system. Adaptation to the distribution of image velocities in H1 is most closely analogous to the adaptation to contrast distributions in retinal ganglion cells[8,36], where many different gain-control processes span multiple spatial and temporal scales[37,38]. As with H1, statistical adaptation in retinal ganglion cells is observed at the output of a network; however, recent work shows that aspects of this adaptation can be seen in the response of individual ganglion cells to injected currents[39], strengthening the connection to ion channel dynamics. Several nonlinear effects in neural responses in visual cortex have been described in terms of a 'normalization' model[40] that has a conceptual connection to the rescaling of the input/output relation in H1. Although the phenomena in cortex usually are not described as adaptation to the distribution of inputs, it has been emphasized that the form of these nonlinearities may be matched to the distribution of natural scenes, so that as in H1 their function would be to enhance the efficiency of the neural code[41]. The present work shows that such effects can occur very rapidly, and that this rapidity is expected if the system is statistically efficient, so that nearly instantaneous events may nonetheless reflect adaptation to the distribution of inputs.

In traditional demonstrations of adaptation (including in H1 (refs 9, 19)) a long-duration adapting stimulus is presented, followed, after some delay to allow a return to baseline, by a standard probe signal. This adapt-and-probe design has the advantage of showing explicitly that adaptation involves a memory of the adapting stimulus. The difficulty is that it requires a clear separation of timescales between the adaptation process (or its decay) and the transient response to the probe. Instead we find that the nonlinear input/output relation of H1 can change on a timescale comparable to its transient response, and that adaptation involves an apparently continuous range of timescales from tens of milliseconds to minutes. Furthermore, there is evidence for adaptation to the variance and correlation time of specific stimulus features (motion) as well as to the more general correlation properties of the visual input[9]. Adaptation to statistics over multiple timescales is almost impossible to characterize fully in an adapt-and-probe design, and so here and in ref. 10 we have tried to characterize the response of the neuron to stimuli as they appear in their statistical context. Statistical adaptation can equally well be considered as an example of context dependence in the neural code. According to information theory efficient codes are always context-dependent in this sense, there are limits to how rapidly context can be established, and there are tradeoffs between context-specific coding and ambiguity about the context itself. The neural code for visual motion in H1 seems to be optimal in all these senses, encouraging us to think that information theory may provide a framework for thinking about context dependence of neural coding and computation more generally. □

Methods

Preparation

A fly *C. vicina*, immobilized with wax, viewed a Tektronix 608 display through a circular diaphragm (diameter: 14 horizontal interommatidial spacings), while action potentials were recorded extracellularly. Data in each figure come from a single fly, but all experiments were duplicated on several flies, yielding similar results.

Stimulus

A computer-controlled random, vertical bar pattern, with frames drawn every 2 ms, moves horizontally with velocity $S(t) = \sigma(t)z(t)$, where $z(t)$ is a sequence of unit variance, uniformly distributed random numbers, drawn independently every $\tau_z = 2$ ms. The standard deviation, $\sigma(t)$, varies on a characteristic timescale $\tau_\sigma \gg \tau_z$.

Input/output relation and dimensionality reduction

The input/output relation is the probability distribution $P(\text{spike}|\text{stimulus})$. Accumulating the full distribution is impractical owing to the high dimensionality of the time-dependent stimulus history before each spike. Previous analysis of H1 shows that only a few dimensions control spiking probability. Reverse correlation[4,42] identifies one of these dimensions: spikes occur at times $\{t_i\}$, and we compute the spike-triggered average stimulus,

$$h(t) = \sum_{i=1}^{N} S(t - t_i)/N$$

We normalize $h(t)$ to $h_n(t)$ to produce unit gain for a constant velocity. The projections, $s_i = \int dt\, h_n(\tau)S(t_i - \tau)$, of the stimulus preceding each t_i, onto the relevant stimulus dimension h_n, provide samples of the response-conditional distribution[26] $P(s|\text{spike})$. The input/output relation follows from Bayes' rule, $P(\text{spike}|s) = P(s|\text{spike})P(\text{spike})/P(s)$, where $P(s)$ is the (gaussian) distribution of all projected stimuli from the experiment and $P(\text{spike})$ is the mean spike rate.

Physical limits to the speed of adaptation

Adaptation to a change in variance of the input signal requires enough time to make a reliable measurement of that variance. This trade between speed and reliability depends on what the system can assume about the nature of the inputs. Suppose we observe a signal $x(t)$ for a time τ and must decide from which of two gaussian distributions it was drawn. For large times τ the probability of error in this decision is dominated by an exponential decay, $P_{\text{error}} \propto \exp(-k\tau)$, with

$$k = \frac{1}{4}\left(\int \frac{d\omega}{2\pi}\left[2 - \frac{X_+(\omega)}{X_-(\omega)} - \frac{X_-(\omega)}{X_+(\omega)}\right]\right)^2 \left(\int \frac{d\omega}{2\pi}\left[1 - \frac{X_+(\omega)}{X_-(\omega)}\right]\right)^{-1}$$

where $X_+(\omega)$ and $X_-(\omega)$ are the power spectra of the two gaussian processes with standard deviations σ_+ and σ_- respectively[43,44]. Here, $x(t)$ is the brain's internal representation of the angular velocity stimulus, containing components both from the presented signal and from noise. Comparing the contrast signals generated by our visual stimulus with photoreceptor voltage noise measured under comparable conditions[45] we find that the signal-to-noise ratio is very high up to some critical frequency, $f_c \approx 50$–100 Hz, above which the effective contrast noise rises steeply. This suggests that we approximate $X_\pm(\omega)$ as signal-dominated for $\omega < 2\pi f_c$ and noise-dominated for $\omega > 2\pi f_c$. As the two signals are gaussian, their power spectra have the same shape but different variance; this difference is large ($\times 10^2$) under the conditions of Fig. 4.

With these approximations we find $k = f_c/2$, so that the fastest possible relaxation time is about 20–40 ms. Additional noise sources in the visual pathway will necessitate longer times for reliable adaptation. Here the error probability is asymmetric: the prefactor of the exponential is larger for misidentifying a small variance signal as a large variance, predicting that transients will be easier to observe after a decrease in variance. An alternative is to consider continuous estimation of the variance, in which case the asymmetry between increasing and decreasing variance appears as a difference in the timescale of relaxation rather than as a prefactor[25].

Information transmission

We calculate information conveyed about the stimulus $S(t) = z(t)\sigma(t)$ as a function of time. At time t, z is one of N probe sequences $\{z_j\}_{j=1,\ldots,N}$ and σ one of $\{\sigma_1, \sigma_2\}$. The cell's responses are expressed as binary 'words'[6] (see Fig. 3 legend), discretized at 2 ms, which is approximately the cell's absolute refractory time. Reliable sampling of word distributions requires a compromise between long words and high resolution. Much larger data sets allow finer discretization, possibly yielding higher information rates[6], but are unlikely to affect our conclusions qualitatively. Interpreted strictly, our results apply only to the case of 2 ms resolution.

With $P_\sigma(W(t)|z)$—the distribution of word responses at time t for a given probe, z, modulated by σ—the information/word about the probe is

$$I_\sigma(W(t);z) = H[P_\sigma(W(t))] - \sum_{j=1}^{N} P(z_j) H[P_\sigma(W(t)/z_j)]$$

where H is the entropy of the word distribution:

$$H[P(W(t))] = -\sum_{W} P(W(t)) \log_2[P(W(t))]$$

Similarly, the information that the distribution of words provides about $\sigma(t)$ is computed from $P_z(W(t)/\sigma)$:

$$I_z(W(t);\sigma) = H[P_z(W(t))] - \sum_{j=1}^{2} P(\sigma_j) H[P_z(W(t)|\sigma_j)]$$

We average this quantity over all probes z_j to obtain the average information, $I(W(t);\sigma)$.

Discrimination

We wish to determine whether a sequence of n interspike intervals $(\delta_1, \delta_2, \delta_3, \ldots, \delta_n)$ is the response to a stimulus with standard deviation σ_1 or σ_2. Optimal discrimination requires computing the log-likelihood ratio,

$$D_n = \log[P((\delta_1, \delta_2, \delta_3, \ldots, \delta_n|\sigma_1)|P(\delta_1, \delta_2, \delta_3, \ldots, \delta_n|\sigma_2)]$$

but the multidimensional distributions again are difficult to sample. We obtain a lower bound to discrimination performance using a simpler decision variable which would be optimal if there were no significant correlations among successive intervals,

$$D_n = \sum_{i=1}^{n} \log[P(\delta_i/\sigma_1)/P(\delta_i/\sigma_2)]$$

We measure the reliability of discrimination by the signal-to-noise ratio, $\mathrm{SNR} = \langle D_n \rangle^2 / \langle D_n - \langle D_n \rangle \rangle^2$.

Received 13 March; accepted 5 July 2001.

1. Attneave, F. Some informational aspects of visual perception. *Psych. Rev.* **61,** 183–193 (1954).
2. Barlow, H. B. in *Sensory Communication* (ed. Rosenbluth, W. A.) 217–234 (MIT Press, Cambridge, Massachusetts, 1961).
3. Laughlin, S. B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* **36c,** 910–912 (1981).
4. Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, Massachusetts, 1997).
5. Berry, M. J., Warland, D. K. & Meister, M. The structure and precision of retinal spike trains. *Proc. Natl Acad. Sci. USA* **94,** 5411–5416 (1997).
6. Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. & Bialek, W. Entropy and information in neural spike trains. *Phys. Rev. Lett.* **80,** 197–200 (1998).
7. Reinagel, P. & Reid, C. Temporal coding of visual information in the thalamus. *J. Neurosci.* **20,** 5392–5400 (2000).
8. Smirnakis, S., Berry, M. J., Warland, D., Bialek, W. & Meister, M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386,** 67–73 (1997).
9. de Ruyter van Steveninck, R. R., Bialek, W., Potters, M., Carlson, R. H. & Lewen, G. D. in *Natural and Artificial Parallel Computation: Proc. of the Fifth NEC Res. Symp.* (ed. Waltz, D. L.) 21–41 (SIAM, Philadelphia, 1996).
10. Brenner, N., Bialek, W. & de Ruyter van Steveninck, R. Adaptive rescaling maximizes information transmission. *Neuron* **26,** 695–702 (2000).
11. Wainwright, M. Visual adaptation as optimal information transmission. *Vision Res.* **39,** 3960–3974 (1999).
12. Francheschini, N., Riehle, A. & le Nestour, A. in *Facets of Vision* (eds Hardie, R. C. & Stavenga, D. G.) 360–390 (Springer, Berlin, 1989).
13. Hausen, K. in *Photoreception and Vision in Invertebrates* (eds Ali, M.) 523–559 (Plenum, New York, 1984).
14. Schilstra, C. & van Hateren, J. H. Blowfly flight and optic flow. I. Thorax kinematics and flight dynamics. *J. Exp. Biol.* **202,** 1481–1490 (1999).
15. Land, M. F. & Collett, T. S. Chasing behaviour of houseflies (*Fannia canicularis*). *J. Comp. Physiol.* **89,** 331–357 (1974).
16. Clague, H., Theunissen, F. & Miller, J. P. Effects of adaptation on neural coding by primary sensory interneurons in the cricket cercal system. *J. Neurophysiol.* **77,** 207–220 (1997).
17. Fairhall, A. L., Lewen, G., Bialek, W. & de Ruyter van Steveninck, R. R. in *Advances in Neural Information Processing Systems 13* (eds Leen, T. K., Dietterich, T. G. & Tresp, V.) 124–130 (MIT Press, Cambridge, Massachusetts, 2001).
18. Thorson, J. & Biederman-Thorson, M. Distributed relaxation processes in a sensory adaptation. *Science* **183,** 161–172 (1974).
19. de Ruyter van Steveninck, R., Zaagman, W. H. & Mastebroek, H. A. K. Adaptation of transient responses of a movement-sensitive neuron in the visual system of the blowfly *Calliphora erythrocephala*. *Biol. Cybern.* **54,** 223–226 (1986).
20. Borst, A. & Egelhaaf, M. Temporal modulation of luminance adapts time constant of fly movement detectors. *Biol. Cybern.* **56,** 209–215 (1987).
21. van Hateren, J. H. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *J. Comp. Physiol. A* **171,** 157–170 (1992).
22. Warland, D. *Reading Between the Spikes: Real-time Processing in Neural Systems.* Thesis, Univ. California at Berkeley (1991).
23. Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. & Warland, D. Reading a neural code. *Science* **252,** 1854–1857 (1991).
24. Schneidman, E., Brenner, N., Tishby, N., de Ruyter van Steveninck, R. & Bialek, W. in *Advances in Neural Information Processing Systems 13* (eds Leen, T. K., Dietterich, T. G. & Tresp, V.) 159–165 (MIT Press, Cambridge, Massachusetts, 2001).
25. deWeese, M. & Zador, A. Asymmetric dynamics in optimal variance adaptation. *Neural Comp.* **10,** 1179–1202 (1998).
26. Ruderman, D. L. & Bialek, W. Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* **73,** 814–817 (1994).
27. Nelken, I., Rotman, Y. & Yosef, O. B. Response of auditory-cortex neurons to structural features of natural sounds. *Nature* **397,** 154–156 (1999).
28. Hopfield, J. J. Transforming neural computations and representing time. *Proc. Natl Acad. Sci. USA* **93,** 15440–15444 (1996).
29. de Ruyter van Steveninck, R. & Bialek, W. Real-time performance of a movement sensitive in the blowfly visual system: information transfer in short spike sequences. *Proc. R. Soc. Lond. Ser. B* **234,** 379–414 (1988).
30. Perkel, D. & Bullock, T. H. Neural coding: a report based on an NRP work session. *Neurosci. Res. Prog. Bull.* **6,** 3 (1968).
31. Wang, Y. & Wang, W. D. Information coding via spontaneous oscillations in neural ensembles. *Phys. Rev. E* **62,** 1063–1068 (2000).
32. Toib, A., Lyakhov, V. & Marom, S. Interaction between duration of activity and time course of recovery from slow inactivation in mammalian brain Na^+ channels. *J. Neurosci.* **18,** 1893–1903 (1998).
33. Segundo, J. P., Moore, G. P., Stensaas, L. J. & Bullock, T. J. Sensitivity of the neurones in *Aplysia* to temporal pattern of arriving impulses. *J. Exp. Biol.* **40,** 643–667 (1963).
34. Markram, H., Gupta, A., Uziel, A., Wang, Y. & Tsodyks, M. Information processing with frequency-dependent synaptic connections. *Neurobiol. Learn. Mem.* **70,** 101–112 (1998).
35. Gerstner, W., Kreiter, A., Markram, H. & Herz, A. Neural codes: firing rates and beyond. *Proc. Natl Acad. Sci. USA* **94,** 12740–12741 (1997).
36. Meister, M. & Berry, M. J. The neural code of the retina. *Neuron* **22,** 435–450 (1999).
37. Shapley, R. M. & Victor, J. D. The contrast gain control of the cat retina. *Vision Res.* **19,** 431–434 (1979).
38. Brown, S. & Masland, R. Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nature Neurosci.* **4,** 44–51 (2001).
39. Kim, K. J. & Rieke, F. Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J. Neurosci.* **21,** 287–299 (2001).
40. Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17,** 8621–8644 (1997).
41. Simoncelli, E. P. & Schwartz, O. in *Advances in Neural Information Processing Systems 11* (eds Kearns, M. S., Solla, S. A. & Cohn, D. A.) 166–172 (MIT Press, Cambridge, Massachusetts, 1999).
42. Wiener, N. *Extrapolation, Interpolation and Smoothing of Time Series* (MIT Press, Cambridge, Massachusetts, 1949).
43. Feynman, R. P. & Hibbs, A. R. *Path Integrals and Quantum Mechanics* (McGraw Hill, New York, 1965).
44. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
45. de Ruyter van Steveninck, R. R. & Laughlin, S. B. The rate of information transfer at graded-potential synapses. *Nature* **379,** 642–645 (1996).

Acknowledgements

We thank B. Agüera y Arcas, T. Adelman and N. Brenner for discussions, and R. Petersen and N. Ulanovsk for comments on the manuscript.

Correspondence and requests for materials should be addressed to A.L.F. (e-mail: adrienne@research.nj.nec.com).