Reliability and statistical efficiency of a blowfly movement-sensitive neuron

ROB DE RUYTER VAN STEVENINCK[1,2]* AND WILLIAM BIALEK[2]

[1] *Laboratorium voor Algemene Natuurkunde, Westersingel 34, 9718 CM Groningen, The Netherlands*
[2] *NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, U.S.A.*

CONTENTS

SUMMARY

We develop model-independent methods for characterizing the reliability of neural spike trains in response to brief stimuli. Through this approach we measure the discriminability of similar stimuli based on the real-time response of a single neuron in much the same way that modern psychophysical techniques measure the discrimination performance of the whole animal. Extending these techniques, we quantify discriminability as a function of time after stimulus presentation, so that it is possible to compare the measured reliability of the neuron to its theoretical limit predicted from signal transduction and noise levels in the sensory periphery. The methods are applied to a wide-field movement-sensitive neuron (H1) in the visual system of the blowfly *Calliphora vicina*, where we also record from the photoreceptor cells that provide the sensory input to H1. From an analysis of neural responses to wide-field stepwise movements of various step sizes we find the following. (1) One or two spikes are sufficient to encode just noticeable differences of approximately one-tenth the angular spacing between photoreceptors, comparable to the hyperacuity régime observed in humans. (2) Discriminability improves upon observation of successive spikes as if the interspike intervals carried independent information. Coding seems orderly and analogue in the sense that we find no indication of information being transmitted in complex combinations of spike intervals. (3) As a result of neural refractoriness the real neuron's performance is significantly better than that of a neuron generating spikes according to a Poisson process at the same firing rate. (4) Over

* Present address: NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, U.S.A.

behaviourally relevant time intervals following the movement step, that is up to about 30–40 ms, the discrimination performance of the neuron is close to that of an ideal observer who extracts movement information from all the photoreceptor cells in the field of stimulation. Beyond this time the neuron's performance relative to the ideal observer decreases significantly.

1. INTRODUCTION

The overall reliability of information processing in an organism can be measured behaviourally by using the ideas of signal detection theory and the methods of psychophysics. This approach has led to a considerable catalogue of data on the performance of human observers at various perceptual tasks. Early papers on this subject (Hecht *et al.* 1942; de Vries 1943; van der Velden 1944; Rose 1948) are concerned mainly with photon counting by the visual system. Their results indicate that the reliability of dark-adapted vision reaches the fundamental limits set by the physics of the stimulus. Even in the absence of these quantitative data, several investigators realized in the 1940s and 50s that understanding the reliability of computation in the nervous system posed significant theoretical challenges. Attempts to perform reliable computations with the available electronic computers certainly posed serious practical problems, and the possibility that the problems of natural and artificial computing are related was explored. Guided by the practical problems of electronic computing, von Neumann (1956) formulated the theoretical problem of 'reliable computation with unreliable components'. Many authors seem to take as self-evident the claim that this is a problem faced by the nervous system as well. Their qualitative picture is of the nervous system as a highly interconnected network of rather noisy cells, in which meaningful signals are represented only by large numbers of neural firing events averaged over numerous redundant neurons. Neurophysiological experiments lend some credence to this view: if the same stimulus is presented repeatedly to a sensory system, the response of an individual spiking afferent neuron is different for each presentation. This apparently has led to a widespread belief that neurons are inherently noisy, and ideas of redundancy and averaging pervade much of the literature. Significant objections to this view have been raised, however (cf. Bullock 1970).

As emphasized by Bullock, the issue of reliability of the nervous system is a quantitative one. Thus, the first problem that should be overcome is to find a way for its measurement. This paper focuses on a restricted but basic question, the reliability of a single neuron, much in the spirit of previous work by Barlow & Levick (1969), Levick *et al.* (1983), Tolhurst *et al.* (1983) and Parker & Hawken (1985). We extend their methods of analysis in an attempt to describe the neuron's reliability in a way that is as model-independent as possible. Among the extensions is a method to study discrimination performance in a time-dependent way.

The second – conceptually more difficult – problem is summarized cogently in Bullock's words 'how reliable is reliable?'. Just quantifying reliability is not enough, and the qualitative question of whether redundancy, averaging, multiplexing, or yet more exotic solutions to von Neumann's problem are relevant to the operation of the nervous system hinges on a quantitative comparison of reliability at the level of single cells with the reliability for the whole system. Essentially, there are two ways to make such a comparison: one can compare the performance of the single cell either with the output or with the input of the whole system. As to the first possibility, if a single cell responds to a certain stimulus as reliably as the animal does in a behavioural experiment, it is difficult to imagine why multiple redundant neurons should be used to encode the same stimulus. Alternatively, if the reliability of a single neuron were to approach the limits set by the sensory periphery, there would seem to be little purpose for the nervous system to use functional duplicates of such a cell. Experiments that address the question from the first point of view are described by Britten *et al.* (1992). These authors report that, if movement stimuli are carefully matched to the receptive field of a neuron recorded from area MT in monkey, the reliability of the recorded nerve cell and of the whole animal in discriminating direction of motion are about equal in most cases. Also, Britten *et al.* observe trial by trial correlations between the behavioural response of the animal and the response of the neuron, which strongly suggests that the monkey is 'listening' to the same cell being studied in the experiment.

In the present paper we take the other approach, and study the reliability of signals both in the retina and in a wide-field movement-sensitive neuron (H1) in the visual system of the blowfly. Signal transfer and noise in the photoreceptor cells are measured by intracellular recording. Based on the measured statistics of photoreceptor noise we compute the limits to the reliability of detection of movement steps for an otherwise noiseless movement detector model. This result is then compared with the reliability of movement information encoded by the H1 neuron, so that we get a measure of the efficiency with which the nervous system uses the total information available at the input in computing a biologically relevant signal. The model used here is a variant of Reichardt's correlator model (Reichardt 1957, 1961), in which movement is computed as the correlation between linearly filtered retinal signals. Under the conditions of our experiment we know that correlation is the optimal strategy for movement discrimination (Bialek 1992); so we are really comparing the performance measured in the neuron with the fundamental limits imposed by the signal quality in the photoreceptor array.

In the H1 neuron, as in most of the neurons of the central nervous system, information about the results of computations is represented in sequences of action potentials. Before we can make meaningful statements

about the reliability of neural computation we must understand the structure of this representation, or 'neural code' (Bullock 1970). Ideally we should like to characterize reliability independent of arbitrary hypotheses about the code, and we approach this in a limited experimental context. Along the way we note several interesting features of the code in H1, but we emphasize that our goal is the characterization of the computation, and not the code. The experimental approach in this paper consists in essence of a direct application of signal detection theory to the trains of neural impulses generated by brief stimuli. Using methods familiar from psychophysics (Green & Swets 1966) we quantify d', the disciminability of different stimuli, and, because neural responses are functions of time, we can extend this to the time-dependent measure of discriminability, $d'(t)$.

In line with the remarks made earlier, the emphasis in this paper is on analysing neural and model responses relative to the moment of stimulation. As that point in time is determined externally and therefore unknown to the fly, we cannot readily translate our results in terms of how the fly itself might interpret its neural responses. Also, in evaluating the reliability of the response we assume that the distribution of the magnitude of possible stimuli is known, and again the fly has no access to that type of knowledge. Both the question of external timing and that of stimulus distribution are instances of a fundamental problem in interpreting neural responses from the point of view of the organism, or, more generally, of evaluating noisy input data. Interpretation of signals in noise inevitably involves finding a balance between observations and prior expectations. In the analysis we shall first assume exact knowledge of the moment of timing. Second, the distribution of stimulus magnitude is kept as simple as possible, and the statistical decision task is always formulated as a choice between two equally probable possibilities.

The above considerations do not invalidate our method of measuring the reliability of the nerve cell or the efficiency of the brain in using the movement information implicitly present in the retina. In this more limited interpretation the fly has the methodological status of a device of which the statistics of input–output behaviour are characterized. Methods for characterizing neuronal reliability under 'real time' conditions closer to those in natural situations are discussed elsewhere (de Ruyter van Steveninck & Bialek 1988; Bialek *et al.* 1991).

2. THEORETICAL FOUNDATIONS

(*a*) *Representation of stimulus and response*

An individual stimulus in our experiment consists of a sudden small displacement of a wide-field pattern. The fly watches the pattern while the response of its H1 neuron is recorded. Steps vary in size and are presented at regular time intervals, long enough to ensure that responses to successive stimuli are independent. In the analysis we therefore treat the stimulus as a point event in time characterized by one parameter, its step size.

For ease of notation we mainly use probabilities $P(.)$ defined on discrete spaces of either responses, stimuli or time. The continuous case is described by probability densities $p(.)$. The neuron's signal is a stochastic process with stimulus-dependent parameters. Its statistical behaviour is described by $P(r)$, which represents the probability of finding a certain response r from the set of possible responses $\{r\}$. From the neurophysiological data we estimate the conditional probabilities $P(r|\xi)$, which describe the chance for a stimulus of step size ξ to generate a response r. In the experiment there was no discernible response until a latency time $t_{\mathrm{lat}} = 15$ ms after stimulus presentation. Spikes fired before this time can be safely attributed to spontaneous activity of the neuron and are excluded from analysis.

Figure 1 summarizes three different approximations to describe the neural response: a succession of time intervals, a time-dependent spike count and a neural firing pattern. In the first (see figure 1*a*) a response r to a stimulus is identified with the set of consecutive time intervals $\{\tau_i\}$. This form of interval distribution should be distinguished from the conventional interspike interval distribution, which is computed without reference to the order of intervals with respect to an external stimulus. We approximate the interval distribution by computing the normalized histogram of consecutive intervals following a large number of stimulus presentations. Most interesting for our purposes are τ_0, and τ_1, and the combination $\{\tau_0, \tau_1\}$. Here τ_0 is the interval from stimulus presentation to the first spike fired after t_{lat}, and τ_1 is the first complete interspike interval following t_{lat}; $p(\{\tau_i\}|\xi)$ represents the conditional probability density for observing the set of intervals $\{\tau_i\}$. The approximation we make here is to consider a finite number of intervals which in this paper will be limited to two.

The time-dependent spike count distribution is a generalization of the method used by Barlow & Levick (1969) to characterize the statistical reliability of a neuron. These authors identified the response with the number of spikes, n, fired in a fixed time window following the stimulus. Here, essentially the same procedure is adopted, but now for time windows of increasing size, as illustrated in figure 1*b*. We denote by $P(n|\xi, t)$ the probability of finding a certain value of n conditional on ξ, at a certain time t. In contrast to the interval description mentioned above, the number of spikes fired can be arbitrarily large, but the method ignores the detailed timing information that might be present in the spike trains.

In principle the neural firing pattern representation yields a statistical description of the response within which every possible spike train is assigned a certain probability. As shown in figure 1*c*, a single response is represented by a sequence $[q_1, q_2, \ldots, q_k]$, or in shorter notation $[q_i]_k$, of binary digits $q_i (i = 1, \ldots, k)$, where $q_i = 1$ and $q_i = 0$ respectively signify the presence or the absence of a spike in time bin i (cf. Eckhorn & Pöpel 1974). Experimentally, the corresponding conditional probability of firing patterns $P([q_i]_k|\xi)$ is determined by counting the number of occurrences of each realization of $[q_i]_k$ for a large number of presentations of stimulus ξ. Time is discretized in this

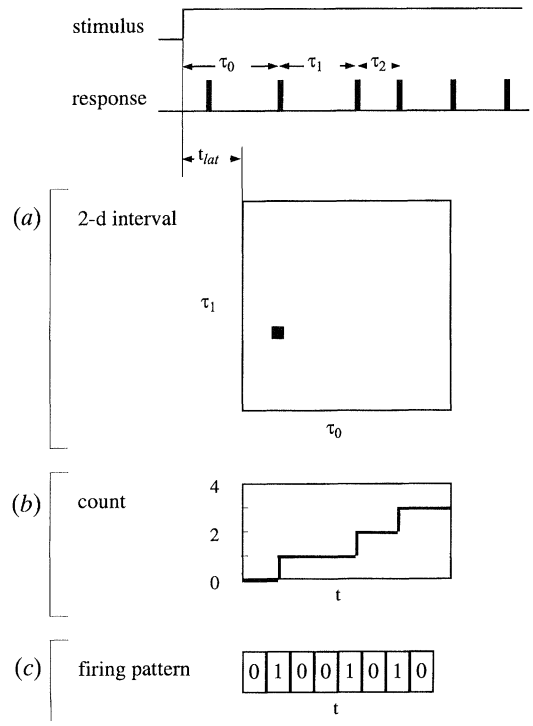

Figure 1. Three methods of describing response statistics. A stimulus is presented at the time of the step in the topmost trace and the neuron's response to this single presentation is shown directly below it. The response consists of a sequence of spikes separated by time intervals τ_1, τ_2 etc. The neural signal over the first $t_{\text{lat}} = 15$ ms following stimulus presentation is ignored in all of the analysis. Thus the first relevant spike is the first one fired after t_{lat} and the interval between stimulus presentation and this spike is denoted by τ_0. The panels (*a–c*) depict three ways of partially representing the single-shot neural response as a function of time. By accumulating these single-shot representations for a large number of neural responses we obtain distributions that are amenable to statistical analysis. (*a*) Here the timing of the first two relevant spikes is recorded by the ordered intervals τ_0 and τ_1 and subsequent intervals are ignored. Each single response generates a $\{\tau_0, \tau_1\}$ pair, the occurrence of which is accumulated in a two-dimensional array. Examples of the resulting distributions are shown in figure 5. By taking the marginals of the two-dimensional distribution we obtain the one-dimensional distributions for τ_0 and τ_1 separately. Examples of these are shown in figure 4. (*b*) The solid line in the box represents the cumulated spike count for a single response as a function of time, starting at t_{lat}. This function of time is accumulated in a series of count histograms ordered along the time axis. In this way, all relevant spikes are taken into account, but in the final distribution (see figure 8) the serial dependence of spikes in single response traces is lost. (*c*) A spike firing pattern is constructed by representing a single trace of response as a binary number. These patterns are accumulated in a 2^n dimensional array, with n the number of bins ($n = 8$ in this example, $n = 12$ in the actual analysis). In this way, all of the information in single responses is preserved in principle. However, owing to the high dimensionality of the representation we are limited to a small number of relatively coarse bins (2 ms in the actual analysis). The resulting distributions can be represented by binary trees, as in figure 10.

approach and for simplicity we use equal time bins of width Δt. Defining $t = k\Delta t + t_{\text{lat}}$ we obtain the time-dependent distribution of firing patterns $P([q_i]_t | \xi)$. The resulting family of probabilities is conveniently described by a tree as shown in figure 10 and described in section 4 (*d*). The time window in this analysis is strongly limited since the number of elements in the set of firing patterns grows as 2^k. In practice we restrict the number of time bins to 12, starting at time t_{lat}. It turns out that meaningful results are obtained with a bin width of 2 ms (see section 2 (*d*)), so that we can get a fairly complete representation of response statistics over a 24 ms window. Apart from the approximations mentioned, we make no assumptions about the way the neuron encodes its signals. As all possible firing patterns are represented by such distributions, this analysis should be quite powerful in revealing any subtle complex code that the neuron might use.

(*b*) *Performance in a discrimination task*

A useful measure of the performance of a neuron, and one that corresponds to psychophysical use, is the reliability with which on average two nearly equal stimuli ξ_1 and ξ_2 can be discriminated on the basis of the difference in neural responses. Such a measure can be computed from the conditional distributions $P(r | \xi_1)$ and $P(r | \xi_2)$. If a particular response r_{obs} is observed it must be decided whether it was caused by stimulus ξ_1 or by ξ_2. The probability of correct decisions is maximized if one uses a maximum likelihood decision rule, so that for equal prior probabilities the outcome is ξ_1 if $P(r_{\text{obs}} | \xi_1) > P(r_{\text{obs}} | \xi_2)$, and vice versa. On average, the probability of correctly identifying step ξ_1 is then

$$P_c(\xi_1) = \sum_{\{r\}} P(r | \xi_1) H[P(r | \xi_1) - P(r | \xi_2)], \tag{1}$$

where $H[.]$ is the Heaviside step function and the summation is over the set of all possible responses $\{r\}$. An interchange of indices 1 and 2 in this expression yields the formula for correct identification of ξ_2. The proportion of correct judgements in an entire experiment in which ξ_1 and ξ_2 are equally probable is then simply $P_c(\xi_1, \xi_2) = [P_c(\xi_1) + P_c(\xi_2)]/2$, which from now on will be referred to as P_c. This analysis is essentially that for a two-alternative forced-choice psychophysical experiment. For convenience we convert P_c into the discriminability parameter d', familiar from psychophysics (Green & Swets 1966), which is the signal to noise ratio (difference in mean divided by the standard deviation) in the equivalent equal-variance gaussian decision problem. So, with $N(x)$ the unit variance cumulative normal distribution function,

$$N(x) = \int_{-\infty}^{x} \exp(-z^2/2)\, dz/\sqrt{2\pi}, \tag{2}$$

d' is related to P_c by

$$d' = 2N^{-1}(P_c). \tag{3}$$

Strictly speaking, d' is only defined in the context of an equal-variance decision problem, whereas the distributions in our experiment are distinctly non-gaussian. Nevertheless, we express our results in terms of d' because in the model treated in section 5 (*c*) d' depends in first approximation linearly on step size.

This is no serious problem because there is a one to one relation between the derived parameter d' and the measured quantity P_c. Hence, if one wishes one can always return to directly measured quantities.

(*c*) *Time-dependent discrimination*

To make the comparison between measured discrimination performance and its theoretical limit we must consider the time dependence of the discriminability function. The reason is that if we allow an arbitrarily long delay it is possible to estimate displacement to arbitrary precision from the signals in the photoreceptor array (see section 5 (*c*)). Because it is pointless for an animal to wait indefinitely for the accuracy to grow, it uses a limited observation interval to make decisions. As we have no independent estimate of this interval, we compare the discriminability as a function of time, $d'(t)$, for the H1 experiment with that for the ideal movement sensor. In doing so, we make the artificial assumption that the timing of the stimulus is known to the observer, as mentioned in the introduction. This poses no problem as long as we are interested only in comparing observed and theoretical performance.

To compute $d'(t)$ for the neuron, the response probability distribution must be specified in a time-dependent way. Both the spike count distribution and the firing pattern distribution described in section (*a*) fulfil this aim by construction, because a new distribution is specified at each successive time bin. Spike interval distributions need some modification: consider the case of single-interval distributions $p(\tau_0|\xi_1)$ and $p(\tau_0|\xi_2)$, defined on the interval $t_{lat} \leqslant \tau_0 < \infty$. The information available on a limited observation interval (t_{lat}, t) during a single trial is either that the first spike was fired at a specific time within the window, or that no spike had yet been fired. Thus, if at a certain time t we wish to assign the observed response to either ξ_1 and ξ_2, we can use the shape of the probability densities $p(\tau_0|\xi)$ for $\tau_0 < t$, combined with the total probability $P(\tau_0 > t|\xi)$. We define a truncated time dependent probability distribution which is zero for $\tau_0 > t$, and which for $\tau_0 \leqslant t$ has the form

$$p(\tau_0|\xi_1, t) = p(\tau_0, \xi_1) + \delta(\tau_0 - t) \cdot \int_t^\infty p(\tau_0, \xi_1)\, d\tau_0 \qquad (4)$$

and a similar one for ξ_2.

Substituting these into equation (1) we can compute d' for each t. This procedure can be generalized to higher-dimensional interval distributions, although the notation becomes more complex because we have to distinguish the different ways in which $\tau_0 + \tau_1 + \ldots$ can exceed t.

(*d*) *Error analysis and data requirements*

In our analysis, in particular for the firing pattern distribution, we make an approximation by discretizing time. The validity of this approximation can be assessed by varying the bin width. It turns out that the results do not change very much if the bins are made smaller than 2 ms. Furthermore, if the analysis is to make sense, stationarity is required, i.e. the probability distribution from which responses to a certain stimulus are drawn should be invariant over the course of the experiment. Also, the distributions, being computed from a finite sample of responses, are subject to statistical error. The statistical error in the final result was estimated by partitioning the data and working out the values of P_c for these partitions separately. The statistical variations in P_c were of the order of 0.01 in the most interesting region of values of P_c, i.e. from 0.6 to 0.9. This results in a typical statistical error of 0.05 in d'. In addition, this analysis revealed no significant trends with time; so we may assume stationarity of the preparation.

Finally, systematic errors can be of importance. Consider for example the comparison of two sparsely filled histograms. In such a case, even if both histograms represent samples from the same distribution, high values of P_c may be computed by the methods described above. These systematic effects were studied both by partitioning the raw data and by varying the bin size of the histograms. This analysis showed that, for the experiment presented here, the systematic errors are of the order of the statistical errors.

3. EXPERIMENTAL METHODS

(*a*) *Preparation and recording*

In this paper we describe one experiment on a female blowfly (*Calliphora vicina*) from which the H1 neuron was recorded extracellularly. In addition, data are presented from intracellular photoreceptor recordings from another specimen of the same species. In both types of experiment the fly was immobilized inside a plastic cylinder and fixed with wax.

For extracellular recording, a tungsten electrode was used, which penetrated the back of the head through a small hole cut in the integument. Care was taken to cut the hole small enough for the wound to seal, preventing desiccation. At regular intervals the fly was fed sugar water or carrot juice. With these precautions reliable recordings could be made for several days. Action potentials were amplified by conventional equipment and electronically discretized. Spike interval times were digitized in 50 μs bins and stored on disk for off-like analysis. Further details can be found in de Ruyter van Steveninck *et al.* (1986).

Intracellular photoreceptor recordings were made with a 150 MΩ glass pipette. The recorded voltage was sampled at 1 kHz, digitized by a 12 bit A/D converter, and stored on disk. For further details, see van Hateren (1985).

(*b*) *Stimulus presentation and measurement of the interommatidial angle*

In the experiments described here the stimulus was a pattern displayed on a Hewlett-Packard 1311A CRT (phosphor P31, with decay time constant 1.2 μs and peak spectral emission at 530 nm). For the experiments on the H1 neuron the pattern consisted of a raster of 2048 vertical lines (spacing 0.029°) which were randomly set dark or bright. The raster had a refresh

rate of 800 Hz, at a mean radiance of 32 mW/(m^2 sr). The fly viewed the screen through a square $(30.5°)^2$ diaphragm. A stationary pattern was produced by cyclically (period 2048) reading binary intensity values from a laboratory-built line-intensity memory in synchrony with the 1.6 MHz vertical deflection signal. Pattern steps were generated at 204.8 ms time intervals. The timing of these steps was locked to the spike interval timer clock: every 4096 periods of the clock signal, the line memory received a pulse that suppressed a preprogrammed (via a PROM) number of read-out pulses to the line-intensity memory. This resulted in a pattern step of a well defined size on the CRT. In the experiment we used a set of 16 steps of different sizes, presented in random order.

In the photoreceptor experiment the same CRT was used, but with a raster of 512 lines at a repetition rate of 3200 Hz. In this case the field was spatially homogeneous, with field intensity set to maximum or zero by a pseudo-random binary signal with a 1 ms time bin. The time-averaged intensity was the same as that in the H1 recordings.

For the computation of the limit to performance in section 5 (*c*) we need the value of the horizontal interommatidial angle α_h, the horizontal sampling basis of the fly visual system. This parameter was measured by determining the Nyquist limit of the retinal sampling raster through the reversed reaction (Götz 1964) in the response of H1 (Zaagman *et al.* 1977). The reversal point was at 2.7° for the frontal visual field of the animal used in the experiment; hence we take $\alpha_h = 2.7°/2 = 1.35°$. In the frontal visual field the ommatidial axes form a regular array of lying hexagons (Beersma *et al.* 1975), from which the solid angle for one hexagon is computed as $2/3(\alpha_h)^2\,(°)^2$. Taking the ratio of the solid angle of the stimulus to that of one hexagon, we estimate the number of illuminated ommatidia to be 442. The movement-sensitive system derives its input from all six members of the R1–6 class of photoreceptors (see review by Wehner (1981)) present in each ommatidium (we neglect the extra sense cells in the equatorial neuro-ommatidia; see Stavenga (1979)); so the numbers of photoreceptor cells that contribute to the response of the H1 neuron in the experiment is about 2652.

4. STEP SIZE DISCRIMINATION BY THE H1 NEURON

(*a*) *Overview of the experiment*

In the experiment the fly observed movement steps, i.e. instantaneous displacements of the whole stimulus pattern, in the preferred direction of the H1 neuron. Steps of 16 sizes in the range 0.12° to 1.92° were presented in random order (see figure 2*a*) at 204.8 ms intervals. The pattern was a random-bar grating with a contrast standard deviation of 0.13 (see section 5 (*c*)). The total effective duration of the recording was about 11 h.

The activity of the neuron averaged over the entire duration of the experiment is shown in the peri-stimulus time histogram (PSTH) of figure 2*b*. The figure shows that the amplitude of the response depends on

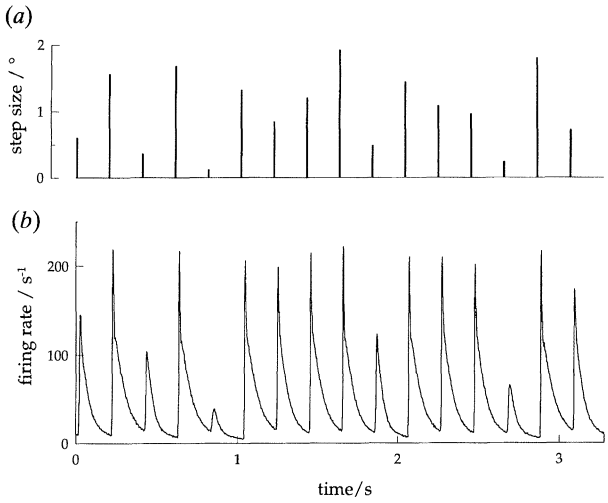

Figure 2. Summary of the stimulus and the spike activity averaged over the entire experiment. (*a*) Bars indicate the occurrences and sizes of 16 successive steps presented at 204.8 ms intervals. Step sizes range from 0.12° to 1.92°. (*b*) The averaged response (PSTH) obtained from 12552 presentations of the complete stimulus sequence.

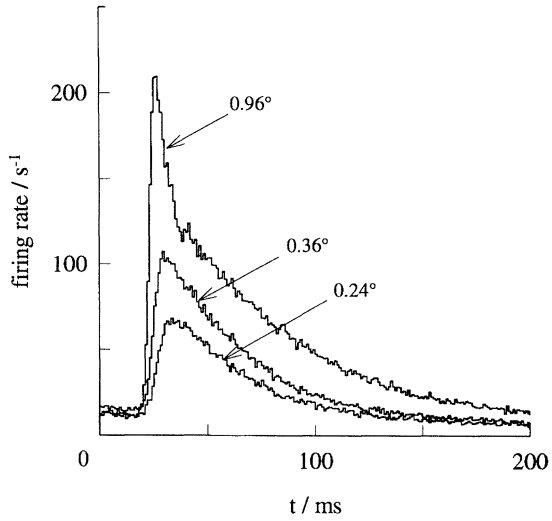

Figure 3. PSTHs for steps of 0.24°, 0.36° and 0.96° shown at 1 ms resolution.

the size of the step, and most strongly so in the region of small steps. For larger steps the responses show a pronounced peak, which reflects a time-locked firing of the first spike.

Most of the results presented below are for 0.24° and 0.36° steps, because there we find the best discrimination performance. Figure 3 shows the PSTHs to these two steps at higher resolution and for comparison a PSTH for a large, 0.96°, step. After a latency of about 20 ms the averaged response rises steeply to a peak and then decays slowly to the resting level. For large steps the peak is very pronounced and followed by a slight undershoot due to refractory effects.

(*b*) *Interval statistics*

(*i*) *Discrimination at long delays*

For the moment we ignore the time dependence of d' and only compute its asymptotic value at long delays for one- and two-dimensional interval histograms.

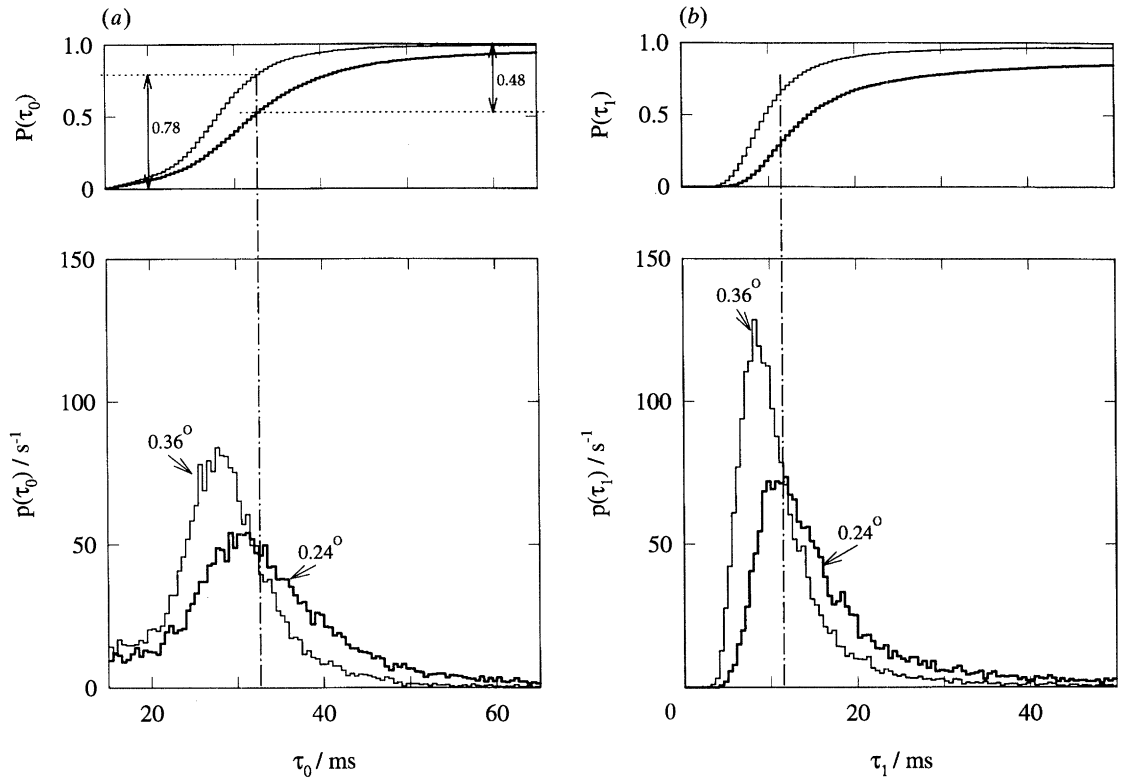

Figure 4. Two pairs of one-dimensional interval distributions, with the corresponding cumulative probabilities on top. Each of the histograms shows results for a step size of 0.24° and 0.36°. (*a*) Timing of the first spike that occurs at least 15 ms after stimulus presentation. (*b*) Histograms for the first interspike interval. See text for further explanation.

Figure 4 shows normalized interval histograms and their cumulative probabilties, for steps of 0.24° and 0.36°. From these the percentage of correct discriminations can be read off graphically. As an example, consider the case of τ_0. The figure shows that for τ_0 less than 32 ms the probability density for the 0.36° step is larger than that for the 0.24° step, but that the converse is true after 32 ms. Thus, using maximum likelihood in a single trial one decides that the step was 0.24° (0.36°) if the observed value of τ_0 is longer (shorter) than 32 ms. With this decision rule, the proportion of correct responses can be read from the cumulative distributions, as indicated by the arrows in the top left panel of the figure. If a step of 0.24° is presented, this will be identified correctly on 48% of the occasions, which is the probability for such a step to generate a first spike after 32 ms. For a 0.36° step, 78% of the first spikes occur earlier than 32 ms. In an experiment in which both these steps occur equally often, the probability of guessing correctly is the mean value of these proportions; thus $P_c = 0.63$. Substituting this in equation (3) we find a value of 0.66 for d' in discriminating a step size difference of 0.12°. So, with the conventional criterion $d' = 1$ for discriminability of two stimuli and the approximation that d' is linear in step size, a difference of 0.18° can be discriminated on the basis of the time delay between stimulus presentation and the first spike fired after t_{lat}.

The two-dimensional interval histograms for $\{\tau_0, \tau_1\}$ conditional on the same pair of steps are shown in figure 5. Figure 5*c* illustrates the maximum likelihood decision map for choosing stimulus step size. The integral of $P(\tau_0, \tau_1 | \xi = 0.24°)$ over the black area is 0.71, while the total probability for a 0.36° step to generate $\{\tau_0, \tau_1\}$ in the grey area is 0.77. The resulting value of P_c is again the average of these two, 0.74, which gives a value of d' close to 1.

Figure 6 presents values of d' for discrimination between step pairs that differ by 0.12°, as a function of the size of the smaller step. The results for discrimination based on the single intervals τ_0 and τ_1 are shown in (*a*). For small steps discrimination based on the interspike intervals τ_1 is better than that for τ_0. For step sizes larger than about 0.8°, discrimination deteriorates for both τ_0 and τ_1. This can be attributed to saturation of the neuron's response: for steps of these sizes it becomes increasingly difficult to generate smaller intervals. For the two-dimensional case shown in (*b*) the best performance is again found for small step sizes, with a just discriminable difference of about 0.1°.

The figure also compares the d' values obtained from the two-dimensional distributions with the pythagorean sum $\sqrt{[d'(\tau_0)^2 + d'(\tau_1)^2]}$, computed from the values of d' for the intervals separately. If for each step size the value of τ_1 is statistically independent of the value of the preceding τ_0, then at least in the gaussian approximation the measured and computed values of d' should agree. The figure shows that the computed data lie somewhat below the measured ones, but on the whole the agreement is quite good. As a check for statistical independence, the correlation coefficients between τ_0 and τ_1 were computed. For all step sizes these coefficients are low, about 0.3 or less, as expected. Thus, apparently the two consecutive intervals can be regarded as statistically independent, and so contribute the information almost independently.

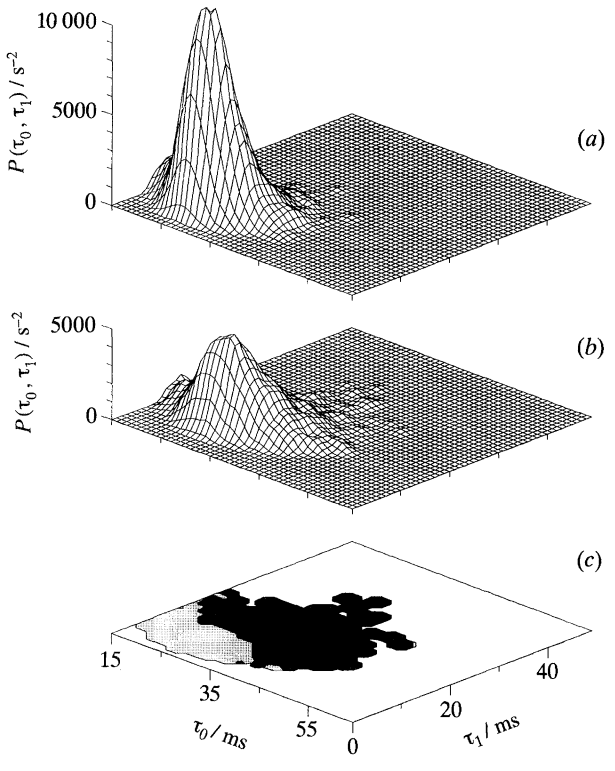

Figure 5. Two-dimensional interval histograms for intervals $\{\tau_0, \tau_1\}$ and step sizes of 0.36° (*a*) and 0.24° (*b*) respectively. (*c*) Decision map for the (0.24°/0.36°) combination, based on histograms (*a*) and (*b*). Interval pairs are assigned to the 0.36° or the 0.24° step on the basis of maximum likelihood. The grey and the black areas correspond to regions with a higher likelihood for the 0.36° and the 0.24° step respectively. For clarity, those bins for which both interval distributions contain five counts or fewer are assigned to neither of the steps.

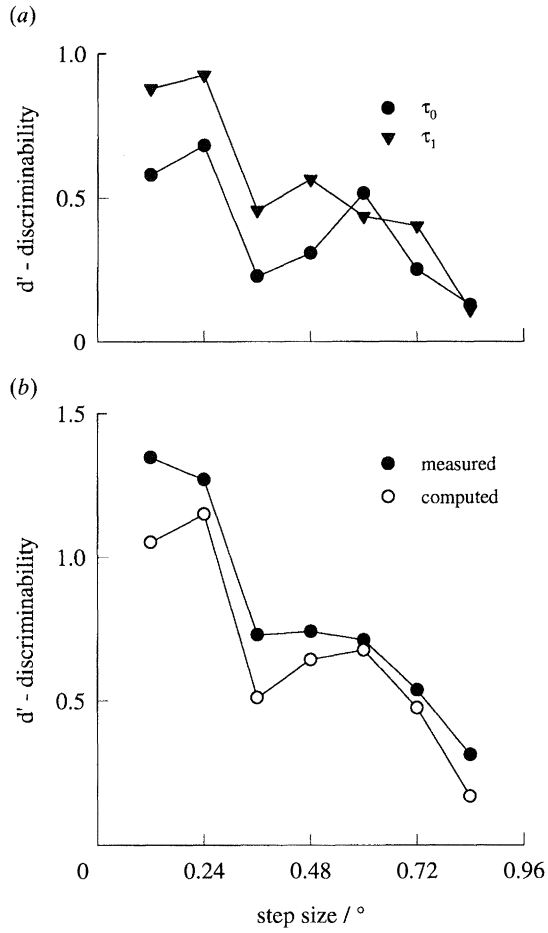

Figure 6. (*a*) Discriminability d' as a function of the size of the smaller step in a discrimination task between pairs of steps that differ by 0.12°. (*b*) As above, but now for the double interval $\{\tau_0, \tau_1\}$. The closed circles represent the values obtained from two-dimensional histograms. Open circles are the pythagorean sums of the d' for both corresponding combinations of single intervals.

(ii) Time-dependent discrimination

By constructing truncated interval distributions (see section 2 (*c*)), d' can be computed as a function of time. Figure 7 shows results for three step size pairs. These are selected to represent the comparisons of two small steps, of a small with a large step and of two large steps respectively. In all figures $d'(t)$ reaches a plateau at a certain time. For the 0.24°/0.36° pair this plateau sets in at $t = 32$ ms for discrimination based on τ_0. This is due to the crossover in the interval distributions for the two step sizes, as can be seen from figure 4. Discrimination performance among the two large steps is relatively bad, while it is best for the 0.12°/0.96° pair. However, if we scale the values of d' for the step size difference, it turns out that the performance is best for the two small steps, as can be seen from figure 12. For the 0.24°/0.36° and 0.12°/0.96° pairs, the performance for the double interval distribution is about twice as good as for the single. In contrast, for the 0.96°/1.44° case, the d' for the first interval alone is about one-third of that for the interval combination, most likely as a result of locking of the first spike.

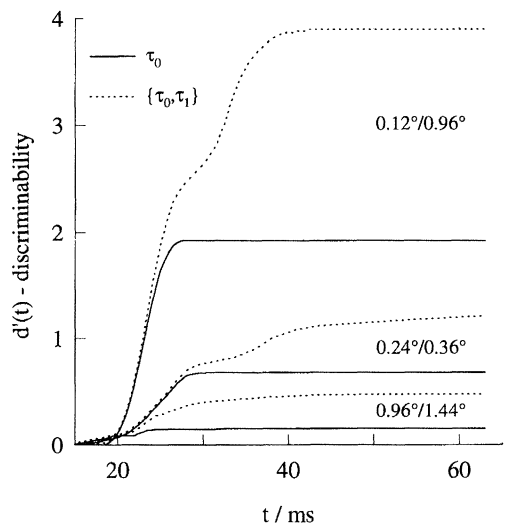

Figure 7. Time-dependent d' for pairwise discrimination of steps with the sizes indicated in the figure.

(*c*) *Spike count statistics*

Time-dependent spike count distributions for steps of 0.24° and 0.36° are shown in figure 8, together with the corresponding maximum likelihood decision map.

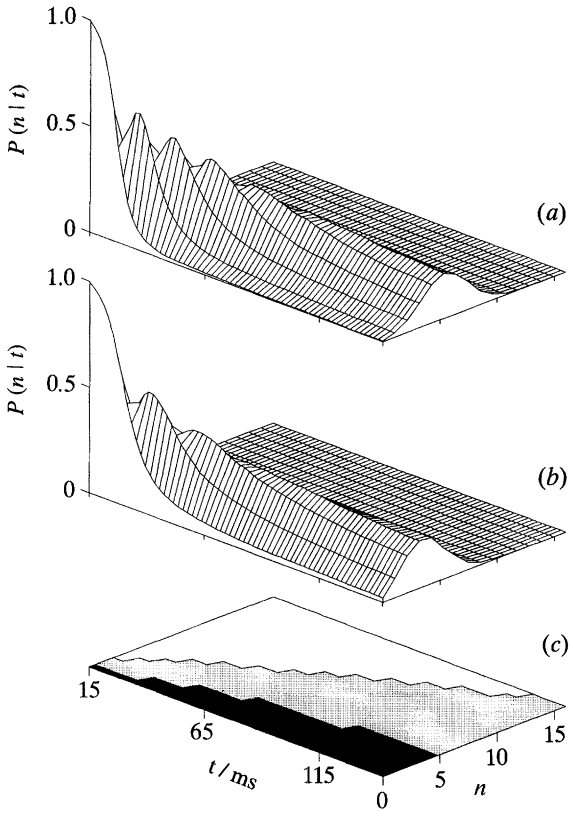

Figure 8. (*a*, *b*) Distributions of the spike count, *n*, conditional on time after stimulus presentation, *t*, for a 0.36° (*a*) and a 0.24° (*b*) step. (*c*) Decision map for the (0.24°, 0.36°) combination. Spike counts are assigned to the 0.36° (grey) or the 0.24° (black) step on the basis of maximum likelihood. The figure might give the false impression that there are dips in the distribution. The impression arises because in the projection shown one looks obliquely at the ridges corresponding to the discrete values taken by *n*.

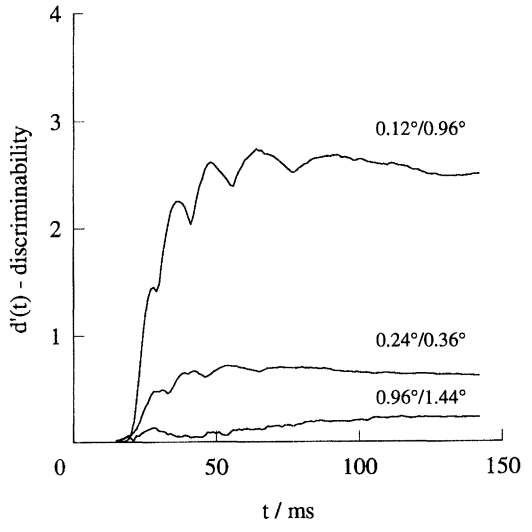

Figure 9. Time-dependent discrimination computed from spike count distributions. Data are shown for the three step size pairs indicated in the figure.

Figure 8*a*, *b* must be interpreted as conditional distributions: for each moment in time, they represent a distribution of spike count up to that time. So, for $t = 15$ ms, the moment at which counting starts, the probability for $n = 0$ equals 1. As time progresses, the probability for $n = 0$ decreases, and flows over to the probability for n to become 1, and so on. The count distributions for the two stimulus conditions are compared for each time bin, and by maximum likelihood a given count at a certain time is assigned to each of the two stimuli. The result of this procedure is depicted in (*c*), which shows that the lower values of n are assigned to the smaller step and the higher values to the larger step, as expected. By computing the associated values of $P_c(t)$ we compute $d'(t)$, shown in figure 9 for three pairs of steps. There are two noteworthy aspects to this result. The first is the fluctuation of $d'(t)$, which is related to the discreteness of the count distributions: from one time bin to the next, the stimulus to which a particular value of n is assigned by maximum likelihood may change from one to the other, as is illustrated by the stepwise changes in the decision map of figure 8. The second is that on a longer time scale the value of d' decreases slowly. This effect is inherent to the analysis of transient responses in terms of spike count: at longer times, the contribution of the neuron's spontaneous activity to the spike count washes out the extra counts in the transient response to the stimulus. The beginning of this effect can be seen from figure 9, but it should be more apparent at longer time scales. This cannot be shown from these data, because the steps in the experiment were presented at 204.8 ms intervals.

The maximum value for d' in figure 9 is about 2.6. If we compare this with the maximum value of nearly 4 found for the interval distributions (see figure 7), it appears that a significant amount of information is lost when only the spike count is taken into consideration.

(*d*) *Firing pattern distributions*

Neural firing pattern distributions for three step sizes are represented by the trees shown in figure 10. These distributions are constructed as follows: starting at the first time bin (15 ms after the stimulus was presented) there are two possibilities: either a spike is fired within the bin or not. These two events are represented by the black bar and the white bar on top of it in bin 1. The length of each of the bars represents the probability of each event, hence the lengths add up to unity. Now the two possibilities in the first time bin create different firing patterns, and for each of the two outcomes in the first bin there are two possibilities in the second bin. In the second bin we then find four areas: the offspring of the black bar in the first bin consists of a (negligibly short) black bar, with a white bar on top. The total length of these two is equal to the length of the black bar in the first bin. The silent events in bin 1 give rise to the other combination of a black and a white bar in the second bin. Thus the lengths of the four bars in bin two represent the probabilities of double firing, of firing followed by silence, of silence followed by firing and of two consecutive silent periods, respectively. This

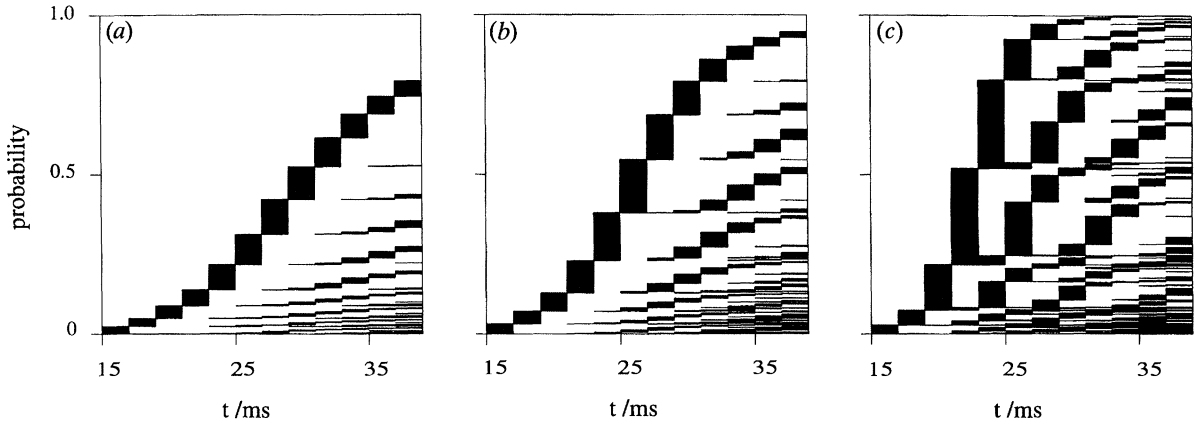

Figure 10. Three representations of firing pattern distribution trees for (*a*) a 0.24° step, (*b*) a 0.36° step and (*c*) a 0.96° step. Each figure displays all the spike sequences recorded in the experiment in response to each of the step sizes, at 2 ms resolution. The sequences are ordered by the arrival times of the spikes, earlier arrivals being treated as more significant.

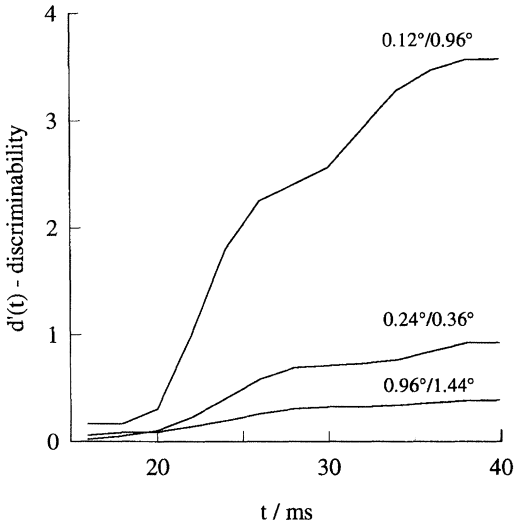

Figure 11. Discrimination performance computed from measured firing pattern distributions such as those in figure 10. The data are presented as a function of time for the three step size pairs indicated.

process of bipartitioning takes place in each subsequent time bin; hence the total number of response categories grows exponentially with time. Here the analysis is limited to 12 bins, or 4096 response categories. Note that the number of presentations, 12552, is of the same order as the number of response categories. We can still do meaningful statistics, however, because a large proportion of response categories hardly ever occurs, for example those containing spikes in adjacent time bins.

As depicted in figure 1, all firing patterns are represented by a binary number. We use this representation in figure 10, ordering the binary numbers from the highest at the bottom to the lowest value at the top, with the more significant digits earlier in time. The effect of the neuron's refractoriness is clearly visible; hence the choice of the time bin seems reasonable. As a check, the performance was computed for 0.5 ms and 1 ms bins as well, and the results for the first 6 and 12 ms respectively were essentially the same as with 2 ms time bins. As outlined above, a set of probabilities is constructed for every time bin, in each of which the sets for the different stimuli can be compared as described in section 2(*b*) to obtain a measure of stimulus discriminability. This is presented in figure 11, where the value of d' is given as a function of time, as it is computed from a number of pairs of firing pattern distributions. After 40 ms the discriminability for the best case is near 1, which agrees with what was found for discrimination based on interval distributions.

At first sight it may seem contradictory to construct a continuous measure from a signal in which information is encoded in the form of discrete points in time. Note, however, that the neuron's response can in effect be considered continuous because not only the occurrence but also the non-occurrence of a spike in a certain time window carries information. This symmetry is expressed in a natural way by representing the firing pattern density in the form of a tree. Furthermore, the analysis does not suffer from the shortcomings of the spike count distributions mentioned in section 4(*c*): $d'(t)$ is a monotone non-decreasing function, because each time bin of the firing pattern distribution takes into account the complete evolution of firing statistics in time.

(*e*) *Comparing coding schemes and stimulus pairs*

Here we treat two separate issues. The first is the comparison of the different assumed coding principles. To this end we plot the $d'(t)$ curves for these cases in the same graphs in figure 12. The second issue is the comparison of discrimination performance for various step size pairs. To make this comparison meaningful, the values of d' for each of these pairs must be normalized. Therefore, in figure 12 the values of d' are scaled to unit (i.e. 1°) step size difference. The scaling factors were computed from equation 9.

It is clear from the figures that at short times the discriminability curves are similar for all coding schemes. This makes sense, because at short times the only information present is based on the timing of the first spike, and for observation windows containing at most one spike there is obviously no real difference

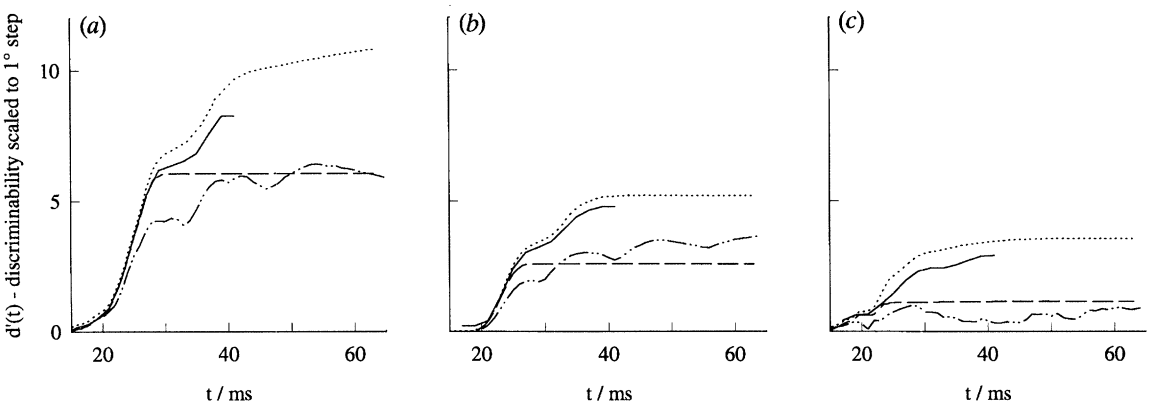

Figure 12. Compilation of $d'(t)$ curves presented above, but all scaled to a 1° step size difference to facilitate comparison among stimulus pairs. (*a*) 0.24°/0.36°; (*b*) 0.12°/0.96°; (*c*) 0.96°/1.44°. Dashed lines, τ_0; dotted lines, $\{\tau_0, \tau_1\}$; dot-dashed lines, spike count; solid lines, firing pattern.

between rate and timing codes. Discrimination based on spike count is less efficient than the other methods when information from the second interval starts to come in, and it loses out considerably at longer times. Over the time window shown, $d'(t)$ based on path distributions is close to, but somewhat less than, that based on the $\{\tau_0, \tau_1\}$. This discrepancy is due to the larger bin size of the path distribution.

At 40 ms after the step, the number of spikes fired on average is 3.2 for the 0.96° step and 2.0 and 1.3 for 0.36° and 0.24°, respectively. Thus it is not surprising that d' for trees is not better than for interval distributions. If we use 4 ms bins we can compute trees over a longer time interval, resulting in 5.2, 3.4 and 2.3 spikes on average for the same cases at 60 ms after the step. For 4 ms bins the dependence of d' on time shows a positive slope at large times, especially for the 0.96°/1.44° pair, but the price paid is a lower performance at small times because there the spike rate is too high for the bin size. The slope at long delays is far less than the slope at around 30 ms, so that even if we combine the data for the 2 ms and 4 ms case in the most optimistic way, the value of d' at 60 ms may only be 10 or 20% higher than that based on two-dimensional interval histograms. In principle there is enough information in the stimulus to encode step sizes with a high precision after a long waiting time, as can be seen from figure 19*a*. The neuron does not seem to encode the stimulus with this precision, however, as the values of d' from the firing pattern distributions are off by roughly an order of magnitude from the theoretical precision. In this sense, and with the caveats mentioned earlier, the analysis of firing pattern distributions seems to exclude the use of a complex coding strategy.

If we compare across step size pairs, and consider the double interval case, it is clear that scaled performance is best for the 0.24°/0.36° pair at long delays. The asymptotic performance is roughly one half and one third of this for the other step pairs. In contrast, the steepness of the $d'(t)$ curves for the different combinations is of equal order of magnitude for the three cases. This is important, because theory predicts an upper bound to the slope of this curve, as will be shown in section 5.

(*f*) *Effects of relative refractoriness*

For a Poisson process, the simplest mathematical model of stochastic point processes, the probability of generating an event is just proportional to the instantaneous rate, and does not depend on the occurrence of previous events. The instantaneous rate of the Poisson process may change over time, in which case it is called inhomogeneous. By contrast, neural action potentials are subject to serial correlation, because it is impossible for a neuron to generate a spike very closely following a previous one (absolute refractoriness) while the threshold for spike generation remains elevated for some time longer (relative refractoriness). In attempts to model neural data these effects are often neglected, and mainly because of its mathematical convenience the Poisson process is used as an approximation.

A priori it is not clear to what extent and under what conditions such approximations are valid or useful. One way to approach this question is to analyse how the serial correlation in the neuron's firing affects discrimination performance. This can be done simply by comparing the performance based on measured distributions to the performance computed for distributions derived for inhomogeneous Poisson processes. Using the rate functions given by the PSTHs of figure 3 we compute the distributions of τ_0 and τ_1 for 0.24° and 0.36° steps shown in figure 13. These should be compared to those at the bottom of figure 4. For τ_0 the distributions for the actual neuron and for the Poisson process do not differ markedly, because their shape is mainly determined by the steep increase in rate at about 20 ms. The distributions for τ_1, however, are dramatically different. For the Poisson process, the maximum of these histograms lies at $\tau_1 = 0$, while figure 4 shows that the generation of the second spike by the neuron is suppressed for over 5 ms, shifting the peak to the right. The important point here is that the refractoriness is relative, which means that the peak for the larger stimulus step is shifted less than the peak for the smaller one. The effect of this is that the two interval distributions are pulled apart further, and as a consequence are more discriminable than they would

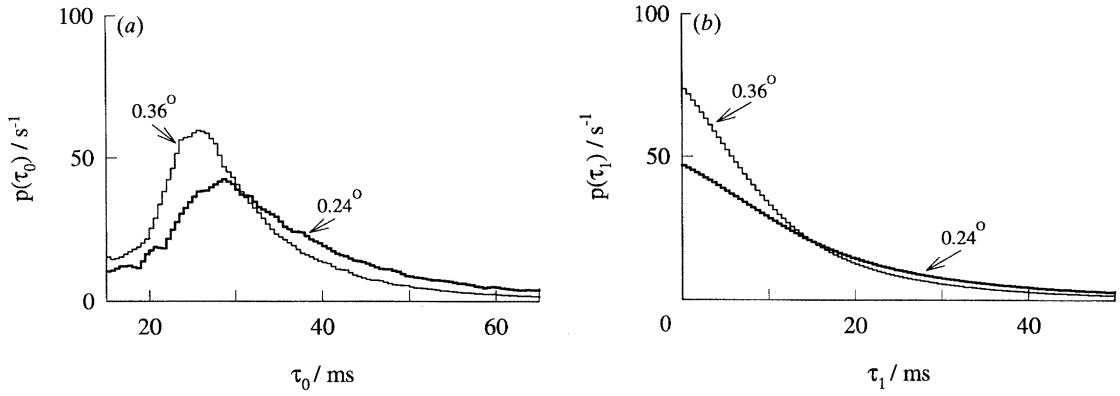

Figure 13. One-dimensional interval distributions for steps of 0.24° and 0.36°, computed for inhomogeneous Poisson processes with rates given by the PSTHs of figure 3. (*a*) Timing of the first spike that occurs at least 15 ms after stimulus presentation. (*b*) Histograms for the first interspike interval. These data should be compared with the measured interval distributions shown in figure 4.

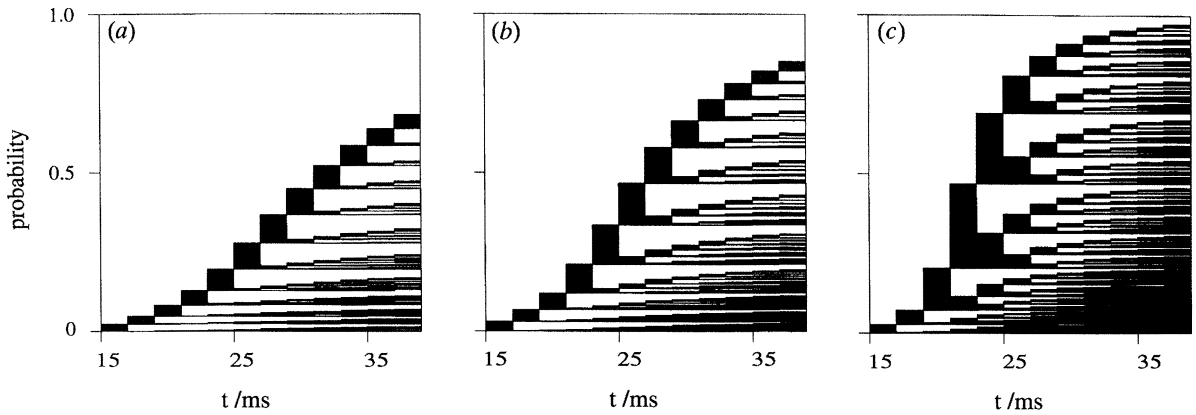

Figure 14. Three examples of computed firing pattern trees for an inhomogeneous Poisson process with a rate given by the measured firing rates of figure 3: ((*a*) 0.24° step; (*b*) 0.36° step; (*c*) 0.96° step). These model results should be compared to those in figure 10.

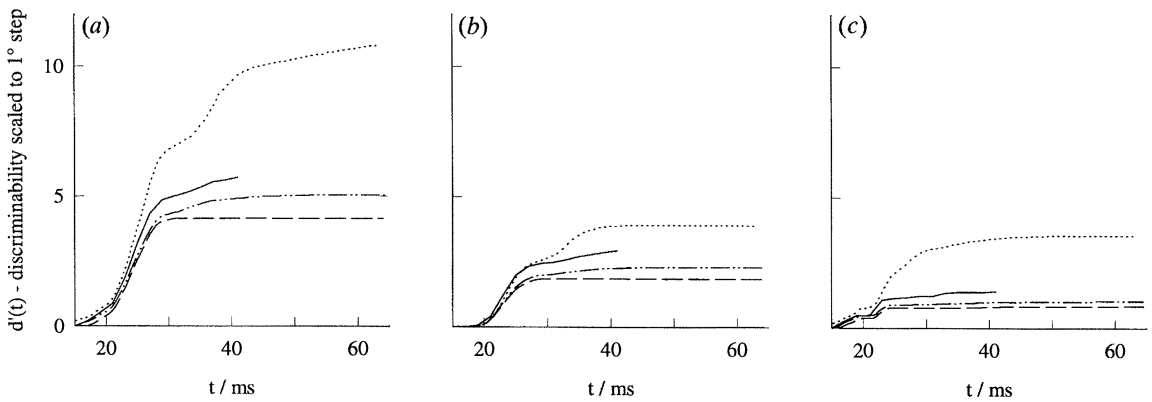

Figure 15. Comparison of discriminability based on one- and two-dimensional interval distributions and firing patterns for the Poisson firing model (dashed lines, τ_0; dot-dashed lines, $\{\tau_0, \tau_1\}$; solid lines, firing pattern). Data are expressed as $d'(t)$ for the same three step size pairs as used in figure 12: (*a*) 0.24°/0.36°; (*b*) 0.12°/0.96°; (*c*) 0.96°/1.44°. To facilitate comparison among stimulus pairs the data are scaled to a step size of 1°. To facilitate comparison with discriminability as measured for the neuron, the dotted line represents the data for the measured performance based on $\{\tau_0, \tau_1\}$.

have been in the absence of refractoriness. Computed firing pattern distributions are shown in figure 14, which should be compared with figure 10. The trees in figure 10 look less dense than those in figure 14, owing to the neuron's suppression of short intervals.

Figure 15 shows scaled discrimination as a function of time for the Poisson one- and two-dimensional histograms and firing pattern distributions, with the same step combinations as in figure 12. As a reference, the figure also shows $d'(t)$ for the measured $\{\tau_0, \tau_1\}$ case. Comparison of figures 12 and 15 shows clearly that discrimination performance based on the first spike is somewhat reduced for the Poisson process. There is a drastic reduction, however, in the contribution of τ_1 to

d' as expected from figure 13. A similar effect is seen for the firing patterns. From the point of view of the fly itself this must be quite important because, as pointed out earlier, internally it has access only to the information in interspike intervals like τ_1 and subsequent intervals, and not to the time lapse τ_0 between an external stimulus event and the first spike.

One way to interpret these results is that the real neuron is more efficient than the Poisson process in using a fixed number of spikes to encode information. This is in line with the observation that the information content of shorter interspike intervals rises rapidly when the interval gets shorter (de Ruyter van Steveninck & Bialek 1988). It is perhaps important to stress that the comparison we make is based on use of a fixed net spike rate that was measured from a real neuron subject to refractoriness. In the Poisson model we use this rate function but we switch off the serial correlation. Something different would happen if we could so to speak switch off refractory effects in the nerve cell while keeping the generator potentials unchanged. In that case the model's spike rate would increase compared with the neuron's, and, at the expense of a greater number of spikes, the model's discrimination performance would improve.

5. RETINAL LIMITS TO RELIABILITY OF MOVEMENT DETECTION

Any inference made by the visual system is limited in its accuracy by the information available at the level of the retina. This in turn is determined by signal transduction and noise in the individual photoreceptors, by the photoreceptor point spread function and array spacing, and by various physical properties of the stimulus.

For the simple stimuli used here it is relatively easy to determine the theoretical bound on step size discrimination based on the photoreceptor signal quality. We compute this limit for Reichardt's (1957, 1961) correlation model of movement detection. This describes the measured movement response for a small step very well (cf. figure 18), and has been very successful in describing a wide variety of phenomena in biological movement detection, both in fly (Reichardt & Poggio 1976; Buchner 1984) and in humans (van Santen & Sperling 1984). Also, it can be shown that correlation is the optimal computational strategy for movement detection under the conditions of our experiment (Bialek 1992; Potters & Bialek 1994). The measured signal transfer of the photoreceptors, combined with the known geometry of the stimulus and the optics of the visual system, determine the signal input to the model. The noise input is taken directly from the measured photoreceptor noise power spectrum.

(*a*) *Spatial properties of the photoreceptor array*

For the spatial sampling basis we take the horizontal projection of the interommatidial angle, $\alpha_h = 1.35°$ (see section 3 (*b*)). The horizontal is the only coordinate of importance in the computation, because the stimulus consists of vertical bars. The point spread function of the photoreceptor is approximated by a gaussian (Götz 1965) with a full width at half maximum of 1.2° (Smakman *et al.* 1984). The corresponding standard width, β, of the gaussian is then $1.2°/\sqrt{(8 \ln 2)} = 0.51°$. For the computations presented here we work with the one-dimensional point spread function

$$M(\phi) = \exp(-\phi^2/2\beta^2)/\beta\sqrt{2\pi}, \quad (5)$$

which has unit gain for spatially homogeneous patterns.

(*b*) *Photoreceptor signal transfer and noise*

Figure 16 shows the impulse response $g(t)$ and the noise power density spectrum W of a blowfly photoreceptor. These results were obtained by measuring the response of an impaled photoreceptor cell while it was stimulated by a series of 100 identical pseudo-random sequences of light intensity values as described in

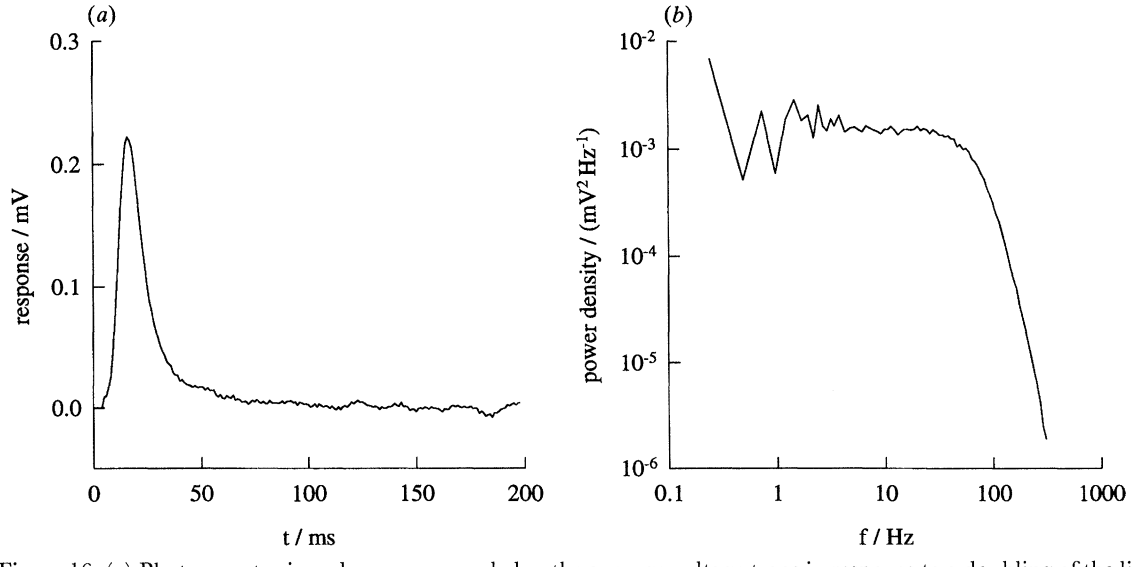

Figure 16. (*a*) Photoreceptor impulse response scaled as the average voltage trace in response to a doubling of the light intensity during 1 ms. (*b*) Photoreceptor noise power density spectrum.

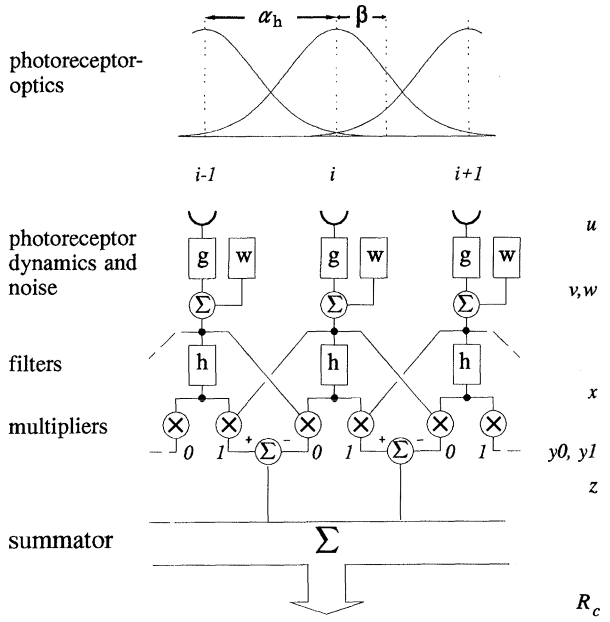

Figure 17. Two neighbouring Reichardt correlators. The gaussian functions on top represent the spatial point spread functions of the photoreceptors characterized by the width β of the point spread function, with an angular separation α_h between neighbourhing photoreceptors. The light flux into each photoreceptor is converted to a voltage by the filters g representing the cell's transduction dynamics. Gaussian noise is added to the signal of each receptor by the independent noise sources w. In the computations the filters h have a negative-exponential impulse response with a time constant of 20 ms.

section 3. At the contrast values used the photoreceptor is linear to a good approximation (Leutscher-Hazelhoff 1975), which means that the averaged response waveform can be described as the convolution of the stimulus contrast with the photoreceptor's impulse response. As convolution in the time domain corresponds to multiplication in the frequency domain, the impulse response is found by taking the ratio of the Fourier transforms of the average response trace and the stimulus sequence and inverse-transforming the result. The noise power density was obtained by averaging the noise power density spectra of all recorded voltage traces after subtraction of the average.

Using shot noise analysis, one can show that, for an ideal photon counter and Poisson photon statistics, the contrast power transfer function divided by the noise power density should be equal to the photon arrival rate. Analysis of the photoreceptor data shows that, up to about 20 Hz, this ratio is close to the $9.8 \times 10^3\ s^{-1}$ photon conversion rate estimated from the stimulus luminance, the optics and the quantum efficiency of phototransduction (de Ruyter van Steveninck 1986). This means that in this frequency range the photoreceptor acts as a nearly ideal photon counter.

(*c*) *The response to a small movement step*

For simplicity we assume that the stimulus pattern makes only one step at $t = 0$. The task of estimating a movement step involves the comparison of the signals present in the photoreceptor array after $t = 0$ with those before $t = 0$. This implies that a replica of the photoreceptor signals must be stored in retinotopic memory elements. It is also intuitively clear that it takes time to make accurate judgements: photons arrive stochastically, and so the change in light flux at $t = 0$ in each individual photoreceptor can be estimated reliably only after a sufficient number of photons has been accumulated. For the simple task we analyse here, namely estimation of the size of one step given its occurrence time and given that this is the only stimulus that occurs on $t = (-\infty, \infty)$, the accuracy can be made arbitrarily high if we are willing to wait long enough to accumulate photons. Clearly this makes no sense for a biological sensor that has to operate in real time. However, in the experiment as it is set up, the time derivative of $d'(t)$ is ultimately bounded by the number of photons that are collected per unit time by the photoreceptors. So we should look at the measured and the predicted slope of $d'(t)$ to make a physically meaningful comparison.

The movement response is computed by an ensemble of elementary movement detectors (Buchner 1976), two of which are shown in figure 17. This configuration is a modification of Reichardt's correlation model for movement detection (Reichardt 1957, 1961). The most salient difference is that the temporal averaging at the output is replaced by summation over an ensemble of spatially distinct units. For this configuration we compute both the ensemble response to a movement step and the output noise power spectrum.

The conditions of the experiment allow an important simplification: the spatial correlation length of the stimulus pattern is much smaller than the angular sampling basis α_h. This means that, for small step sizes, correlations among signals from next-nearest neighbour photoreceptors are negligibly small compared with those among nearest neighbours, so that correlator outputs in different areas of the visual field are statistically independent. Because the pattern moves as a whole, the outputs of different correlators have equal expectation values. The total movement signal can thus be obtained simply by adding the signals of all nearest-neighbour correlators, and the output noise power spectrum of the correlator array is found as the sum of the power spectra of different elementary correlators. In this section and the following we compute the expectation values of the response, and the noise power spectrum, of one elementary correlator. There are two inputs to each correlator, corresponding to different directions of view. In the eye of the fly, six photoreceptors share the same direction of view, and their signals are combined by neural superposition (Kirschfeld 1967). In computing signal and noise for a single correlator we therefore simply use inputs that can be thought of as the sum of six photoreceptor signals, each with an independent noise source. This effectively increases the signal to noise amplitude ratio at the correlator inputs by a factor $\sqrt{6}$.

The elements h in the model of figure 17 represent the retinotopic memory referred to earlier. They are modelled as first-order low pass filters, i.e. $h(t) = (1/\tau) \exp(-t/\tau)$, with $\tau = 20$ ms. This value of τ is chosen to fit the model's response to that of the H1

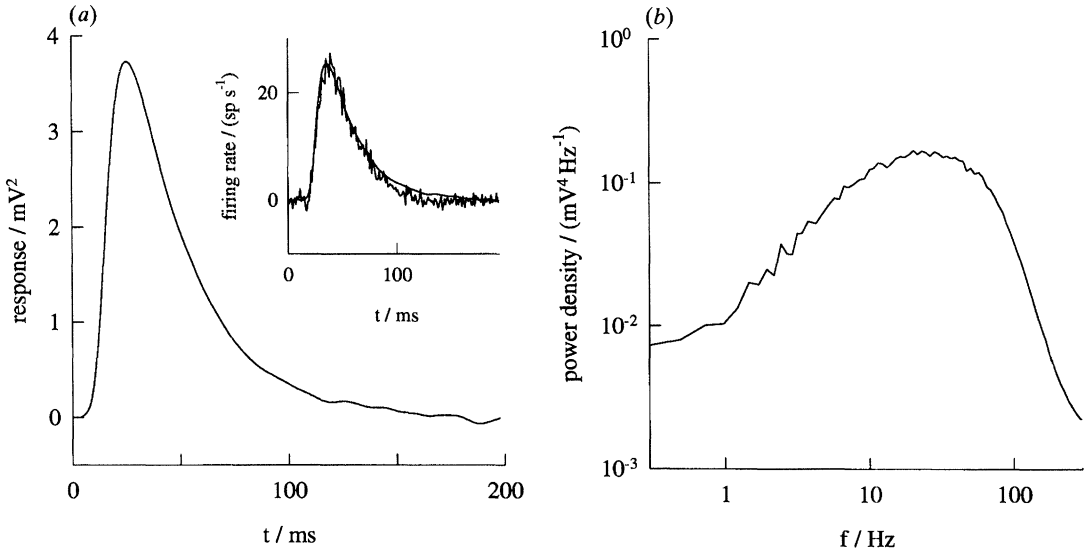

Figure 18. (*a*) Response of a single model correlator to a 0.12° movement step. Inset: the computed response (smooth line), shifted 10 ms to the right and scaled to fit the PSTH (ragged line) in response to a 0.12° movement step. (*b*) Noise power density spectrum at the output of a single model correlator. The results are computed for the conditions pertinent to the experiment. Each of the two correlator inputs receives the summed signals of six photoreceptors, in accord with neural superposition in the fly's eye (see text). The units along the vertical axes are somewhat unusual because the output signal of the model correlator is the product of two signals each with the dimension of voltage.

neuron for a small step (see inset, figure 18*a*). If τ is allowed to become very large, the step response reaches a constant level different from zero at large times. As the noise level stays finite, single-step displacement can be estimated with arbitrarily high precision if one waits long enough. This problem is avoided by giving τ a realistic value. However, the exact value of τ is not very relevant for our purposes, because we shall compare the measured and computed values of d' only for the initial rising slope of the response, and the steepness of this is determined by the steepness of the photoreceptor response.

As illustrated in the figure, the input light intensities are denoted by u_i, the photoreceptor signals by v_i and the output of the filter h by x_i. These signals all depend linearly on the stimulus intensity. After the multipliers we have the signals $y0_i$ and $y1_i$ and the output z_i of an elementary correlator, which is equal to $y1_i - y0_{i+1}$. The photoreceptor impulse response is denoted by $g(t)$ and its step response by $G(t)$, and it is convenient to define the convolution of $g(t)$ and $h(t)$ separately as $\psi(t)$, with the corresponding step response $\Psi(t)$. Note that $G(t)$ and $\Psi(t)$ represent responses to a unit step in light intensity; they should not be confused with the model's response to a movement step. The direct current (DC) gains of $g(t)$, $h(t)$ and $\psi(t)$ are represented by g_0, h_0 and ψ_0 respectively; thus $\psi_0 = g_0 h_0$. Finally, ensemble-averaging is denoted by $\langle \ldots \rangle$, and signals just before and after $t = 0$ carry the superscripts $^-$ and $^+$ respectively. For the signals v and x we now find:

$$v^+(t) = u^- g_0 + (u^+ - u^-)\, G(t),$$
$$x^+(t) = u^- \psi_0 + (u^+ - u^-)\, \Psi(t). \tag{6}$$

For an elementary correlator i, we have $y1_i = v_{i+1} x_i$, and $y0_{i+1} = v_i x_{i+1}$, and the difference z_i is found by substituting (6):

$$z_i(t) = (u_i^- u_{i+1}^+ - u_{i+1}^- u_i^+)\, [\psi_0 G(t) - g_0 \Psi(t)], \tag{7}$$

in which the first factor on the right side depends on the spatial geometry and on the stimulus, and the second on the dynamics of the filters. To find the output of an ensemble of correlators, the first factor, A, is ensemble-averaged: $\langle A \rangle = \langle (u_i^- u_{i+1}^+ - u_{i+1}^+ u_i^+) \rangle$. The two terms in this equation are two values of the spatial autocorrelation function of the stimulus pattern as seen through the photoreceptor point spread function, at different distances. If we neglect the DC level in the stimulus pattern, which carries no movement information anyhow, this autocorrelation, Θ_{ss}, is found as

$$\Theta_{ss}(\phi) \approx \Phi_b \exp(-\phi^2/4\beta^2)/2\beta\sqrt{\pi}, \tag{8}$$

where $\Phi_b = 0.029°$ is the bar width of the stimulus binary random bar pattern, and the approximation is good because $\Phi_b \ll \beta$. The variance of the stimulus contrast after spatial filtering by the photoreceptor is found directly as the value of (8) for $\phi = 0$. Substituting Φ_b and β we find $\Theta_{ss}(0) = 0.016$; so the standard deviation of contrast values is about 0.13.

For a movement step of size ξ the value of $\langle A \rangle$ is

$$\langle A(\xi) \rangle = \Theta_{ss}(\alpha_h - \xi) - \Theta_{ss}(\alpha_h + \xi), \tag{9}$$

which is proportional to ξ if ξ is small. This result was used to scale the values of d' for each step size pair in figures 12 and 15. The output of an ensemble consisting of N elementary correlators is then

$$R_c(t, \xi) = N[\psi_0 G(t) - g_0 \Psi(t)] \times [\Theta_{ss}(\alpha_h - \xi) - \Theta_{ss}(\alpha_h + \xi)], \tag{10}$$

and all quantities appearing in the expression are known by direct measurement, except for $h(t)$. The resulting form of the response is shown in figure 18*a*, for a step size $\xi = 0.12°$. The inset in this figure shows the measured PSTH for a 0.12° movement step, and for comparison the computed response, scaled to fit the PSTH. Also the computed response is shown shifted

10 ms to the right to account for latency times in computation and signal transport within the nervous system. This total delay time is also used in comparing measured and computed responses, as in figure 19.

With the 20 ms time constant for $h(t)$ and the measured integration time of the photoreceptor (about 13 ms), the rising flank of the model response is determined primarily by the photoreceptor's dynamics, while its decay is mostly governed by the shape of $h(t)$. The inset shows that the model's output corresponds quite well to the measured H1 response.

(*d*) *Correlator output noise power*

The inputs to correlator i consist of two independent fluctuating signals w_i and w_{i+1} due to photoreceptor noise, each superimposed on temporal DC terms $v_i = u_i g_0$ and $v_{i+1} = u_{i+1} g_0$ which represent the average photoreceptor depolarization levels. For the computation of the correlator output noise we consider the stimulus-induced signals u and with it the photoreceptor depolarizations v to be constant in time. This means that we neglect the changes in v induced by the movement step. As the pattern contrast is small and we are interested in small step sizes here, such an approximation is reasonable. Because the stimulus pattern has spatial structure, the signals v are in general different for different photoreceptors. For the computation of the total output noise, the variance in the ensemble $\{v_i\}$ of time-averaged photoreceptor voltages can be considered a component in the photoreceptor noise spectrum at zero frequency. Therefore the ensemble-averaged noise power density spectrum for each input to the correlator is

$$N_{\mathrm{pr}}(\omega) = W(\omega) + 2\pi\Theta_{\mathrm{ss}}(0)\, g_0^2\, \delta(\omega), \tag{11}$$

with $W(\omega)$ the measured photoreceptor noise power density and $\Theta_{\mathrm{ss}}(0)\, g_0^2$ the variance of the time-averaged depolarizations over the ensemble of photoreceptors.

For uncorrelated input signals, the output power density spectrum of a multiplier is the convolution of the power density spectra of its input signals (Mircea & Sinnreich 1969),

$$N(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} N_0(\omega_1)\, N_1(\omega-\omega_1)\, \mathrm{d}\omega_1. \tag{12}$$

The dynamics and the structure of an elementary correlator are represented by the transfer kernel (Wiener 1958; Poggio & Reichardt 1973),

$$B(\omega_1, \omega-\omega_1) = \tilde{h}(\omega_1) - \tilde{h}(\omega-\omega_1), \tag{13}$$

where $\tilde{h}(\omega)$ is the transfer function of the filter's impulse response $h(t)$. The two terms on the right are from the two branches of the correlator. Each of these terms depends on only one frequency, because one of the components in each branch is all-pass. By combining (12) and (13), the ensemble-averaged correlator output noise power spectrum is found as

$$N_{\mathrm{c}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} N_0(\omega_1)\, N_1(\omega-\omega_1)\, |B(\omega_1, \omega-\omega_1)|^2\, \mathrm{d}\omega_1, \tag{14}$$

where N_0 and N_1 are uncorrelated noise power spectra given by equation 11. The result of (14) is shown in figure 18*b*. Each of the interacting input noise terms in equation (14) is the sum of a noise term and a stimulus-induced variance, as given by equation (11).

(*e*) *Detection and discrimination*

Because the final result will be derived for the sum of a large number of correlator outputs, we can apply the central limit theorem and approximate the correlator output noise as gaussian. Then, from the shape of the response and the autocorrelation of the correlator output noise, the detection performance can be obtained in a standard way. Here we follow Van Trees (1967) and compute d' as

$$d' = \sqrt{[R_{\mathrm{c}}^T(t')\, \Theta_{\mathrm{cc}}^{-1}(t,t')\, R_{\mathrm{c}}(t)]}, \tag{15}$$

in which $R_{\mathrm{c}}(t)$ is a vector describing the computed response, given by (10); $\Theta_{\mathrm{cc}}^{-1}(t,t')$ is the operator inverse of the movement detector's output noise autocorrelation $\Theta_{\mathrm{cc}}(t)$, which itself is obtained as the inverse Fourier transform of $N_{\mathrm{c}}(\omega)$ in equation (14). Note that both the response and the covariance are measured in discrete bins, and can therefore be treated as a vector and a matrix respectively. The value of d' can now be found by computing $\Theta_{\mathrm{cc}}^{-1}(t,t')$ for increasing time windows to obtain d' as a function of time. The total response is found by multiplying the ensemble-averaged response by the number, N, of correlators. The noise output signals of the different correlators are independent, because, as can be seen from figure 17, one can add the correlator outputs going from left to right in the figure, and each new output contains one new independent noise source from the input on the right. Therefore the ensemble noise power is found by multiplying the result for one correlator by N, and so d' follows the square-root law. In figure 19*b* $d'(t)$ is shown as a function of time for an ensemble of 442 correlators, in agreement with the conditions of the experiment.

It is perhaps helpful to consider how the correlator's output signal to noise ratio scales with input contrast and with the incoming photon flux. If the photon flux absorbed by each photoreceptor has a mean rate λ, then the average signal power in the photoreceptor is proportional to $\Theta_{\mathrm{ss}}(0)\, \lambda^2$, where $\Theta_{\mathrm{ss}}(0)$ is the stimulus contrast variance as in equation (8). The output signal power of the correlator should scale simply as the square of the photoreceptor power. The photon shot noise power is proportional to λ. In the conditions of our experiment the output noise level is dominated by cross-interactions between stimulus-induced signals and noise, rather than by noise–noise interactions. This means that the correlator output noise power scales approximately as the product of input signal power and input noise power, thus as $\Theta_{\mathrm{ss}}(0)\, \lambda^3$. Consequently, the output signal to noise power ratio is proportional to $[\Theta_{\mathrm{ss}}(0)\, \lambda^2]^2 / [\Theta_{\mathrm{ss}}(0)\, \lambda^3]$, or $[\Theta_{\mathrm{ss}}(0)\, \lambda]$. This represents a signal to noise power ratio; hence d' should be proportional to the square root of this quantity. In particular this means that d' is proportional to contrast and has a square-root dependence on the photon capture rate. For very low contrasts or photon capture rates the output noise power is dominated by noise–noise interactions. Then the correlator output noise

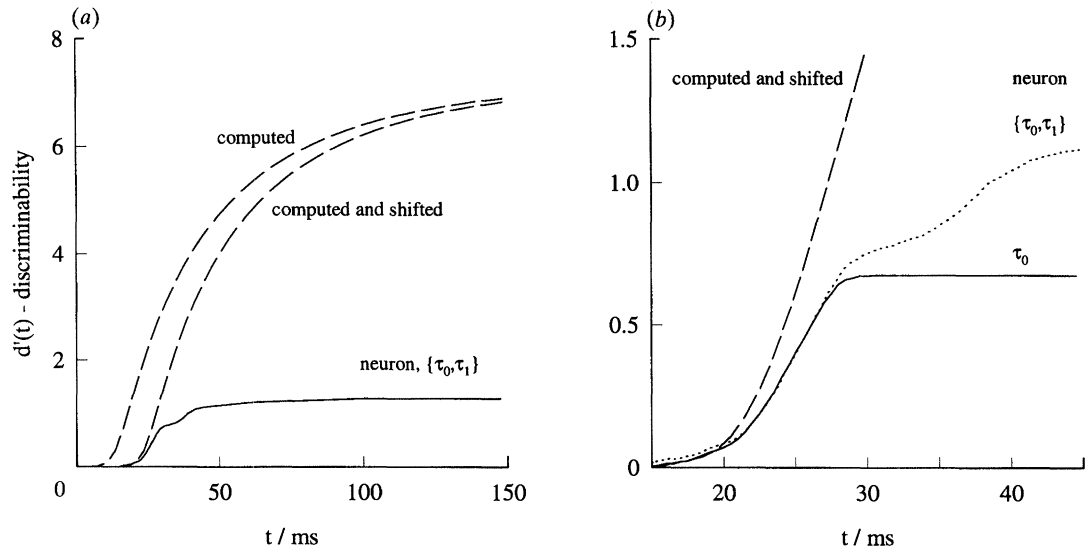

Figure 19. (*a*) Comparison of the discrimination performances of the neuron and the model movement detector. Solid line : $d'(t)$ for the step pair 0.24°/0.36°, based on $\{\tau_0, \tau_1\}$. Left dashed line: $d'(t)$, computed from the noise spectrum together with the spatial and temporal transfer functions of fly photoreceptors, and the number of photoreceptors involved in the response. As shown in figure 18 the neuron's response is delayed with respect to the model by 10 ms. The same delay was applied to the model, and the result shown as the second dashed line. (*b*) The measurements and the shifted computation shown at higher resolution. Solid curve: $d'(t)$ for τ_0 only. Dotted curve: $d'(t)$ for the double interval $\{\tau_0, \tau_1\}$. Dashed curve: $d'(t)$ for the model, shifted over 10 ms. The maximum slope of the neuron's discriminability curve is about half that of the computed curve. The region over which the information is carried primarily by the first spike is from about 20 to 28 ms.

power is proportional to λ^2, so that d' scales as $\Theta_{ss}(0)\,\lambda$. In other words, in this régime d' is proportional to the square of the contrast and is linear in the photon flux.

(*f*) *Comparison of measured and theoretical performance*

In figure 19 the computed limit to step discrimination is compared with the performance of H1 in the same task. To judge neural performance against the theoretical limit we shift the computed result by a 10 ms delay as derived from the fit of the model to the neuron's response (see inset, figure 18). Figure 19*b* shows first of all that at long times the neuron performs much worse than the model. This is not too surprising since the model's asymptotic discrimination performance is determined by our choice of model filter time constants. The asymptotic performance can be made arbitrarily high by choosing long enough time constants, as pointed out in section 5(*c*). As also pointed out there, the more meaningful thing to do is to compare the slopes of the $d'(t)$ curves. To see this in more detail the data are shown magnified in figure 19*b*. In particular, in the region where the information of the first spike comes in, the maximal slope of the H1 curve is about half that of the theoretical curve. After that the curve levels off somewhat and rises again when the information from the second spike is used. At 45 ms, when most of the information of the second spike is used, the neuron's d' is about one-third of the model's.

As reasoned above, d' should scale as the square root of the photon flux. During the fast rise in d' the neuron is off by about a factor of two in the computation of the movement signal; so there the brain uses 25% of the effective number of photons absorbed in the visual field. So, for this time window the randomness of H1's response is determined to an appreciable extent by photoreceptor noise. As a rough estimate, when the photoreceptor signal is integrated over an 8 ms time window most of the resulting photoreceptor noise is due to the random arrival of photons (see section 5(*b*)). Therefore, in the statistical decision task formulated here, the performance of H1 is well within an order of magnitude from that of an ideal wide-field movement sensor limited by photon shot noise in the 2652 photoreceptor cells, or 442 elementary correlators, stimulated in the experiment. Of course this is only true over a short time window and, as the figure shows, the neuron's asymptotic performance is far below the theoretical limit. We shall go into this in the next section.

6. DISCUSSION

We wish to remind the reader once more that from the point of view of the animal our analysis of reliability of H1 may be rather artificial, because most of the computations are done on response distributions specified relative to the timing of an external stimulus. In real life the fly does not know the timing of stimuli generated independently of the animal, and therefore cannot use the information carried by the timing of the first spike on this one neuron by itself. Of course the fly does have access to the information stored in spike intervals, either on the single neuron, or among different cells. If we just consider the single cell we see that the spike interval τ_1 generally gives somewhat better discrimination than τ_0, of course at the expense of a longer waiting time. Interestingly, the encoding of

step size in interspike intervals seems much more efficient for the real neuron than for a Poisson model, given equal firing rates.

Independent of these considerations, the question of performance of the fly's nervous system in computing movement step sizes is an interesting one, even if we do not know how the fly uses the information in H1. The analysis given here yields a valid description of this performance in a well defined, but restricted experimental context. Within this framework the results can be compared quantitatively to the theoretical limits computed on the basis of the information available in the retina. The quantity of fundamental interest here is the steepness of the slope of $d'(t)$, because that is limited by the incoming photon flux. If we make the comparison, we find that H1 is reasonably efficient in using the movement information available at the periphery over a short time interval of 5–10 ms, depending on the strength of the stimulus. The analysis of the real-time reliability based on truncated interval distributions and on firing patterns (see section 4(*e*)) reveals that much of the information is contained in the time of firing of the first and the second spike. As shown in figure 3, the latency from stimulus presentation to spike generation decreases for increasing step size. Consequently, if we limit the analysis to a fixed number of spikes the scaled asymptotic value of d' must be lower for the pair of large than for the pair of small steps, because the slope of the d' curve is bounded by the finite photon flux. This is demonstrated clearly by figure 12. However, for all three step size pairs shown in the figure the slope of the neuron's discrimination performance is within a factor of two from the theoretical value. This demonstrates that the statistical efficiency of the fly's movement detector does not depend very much on the stimulus strength over the range of step sizes used here.

It should not be too surprising that the neuron performs well only over a short time interval and does not reach the values for d' computed from the model at large delays (figure 19*a*): the experimental stimulus is not very natural, and in real-life conditions the fly is likely to see movement changing continuously (de Ruyter van Steveninck & Bialek 1988; Bialek *et al.* 1991). In such circumstances it might be better not to wait very long to get an accurate estimate of the stimulus at one point in time, but rather to update rough estimates as fast as possible. Such a view is supported by behavioural evidence: a chasing fly tracks the leading fly with a delay of about 30 ms (Land & Collett 1974), corresponding to the time at which the measured $d'(t)$ levels off. Additional support comes from the analysis of firing pattern distributions, especially for larger step sizes, which does not indicate the existence of a complex time-consuming code to specify step sizes. Finally we notice that the analysis of reliability in real-time estimation tasks (Bialek *et al.* 1991) shows that the 'effective noise level' for movement estimation by H1 is equivalent to a discrimination threshold of about 0.1° in a 30 ms window, comparable to what we find here.

A step size difference of 0.12° can be discriminated with d' close to unity, with use of the timing information of just one spike from one neuron. For the blowfly visual system this angular difference is of the order of one-tenth of the photoreceptor spacing, well within the hyperacuity régime. A similar result was found for monkey cortical cells by Parker & Hawken (1985). Geisler (1989) compiles data on a variety of human photopic visual tasks and compares human performance with an ideal observer who has access to all the photons entering within the aperture of the cones in the retina. The values Geisler computes for the fraction of photons used are below 1 %. These low values may be due in part to a low statistical efficiency of the cones. A hint in this direction is that human scotopic performance seems much more efficient: Savage & Banks (1992) report a highest value of 33 % for the efficiency in a contrast discrimination task. They specify the efficiency, as we do, with respect to the signal to noise ratio of the photoreceptors involved in the visual task and hence the value they report reflects the efficiency with which the brain processes the information present in the retinal voltages. The values of 25 % we report here for the single neuron compares well with the efficiency of the whole human observer. Similar values were obtained by Barlow (1980) and Tapiovaara (1990) for symmetry recognition tasks. This may mean that, given the right conditions, visual information processing really approaches the ideal observer limits, independent of what animal the brain belongs to. Indeed, this notion of near-optimal neural performance may be more generally applicable (Barlow 1981; Bialek 1992).

Given that neurons operate with limited dynamic range and that the brain approaches optimal performance, the nervous system must employ adaptive strategies to keep the computations it performs near to the optimum. It is well known that biological systems adapt their sensors to ambient conditions, and in the fly this has been interpreted as a strategy for improving coding strategies (Laughlin 1989). There is no reason to believe that adaptation should stop there, and in fact there are strong theoretical indications that the visual system should adapt its computational strategies to higher-order statistics of the visual stimulus to perform movement detection in an optimal way (Potters & Bialek 1994). An experimental example is found in H1 itself, which exhibits a form of movement-specific adaptation where the dynamics of the response adapts to the speed at which a stimulus pattern moves (Maddess & Laughlin 1985; de Ruyter van Steveninck *et al.* 1986). Moreover, the response dynamics of H1 adapts to the time interval between stepwise movement stimuli (Zaagman *et al.* 1983). Most probably due to this effect it is possible to present steps with the relatively short intervals used here, without overlap of the responses.

7. CONCLUSION

Over a limited time window, from about 20 to 30 ms after the stimulus step, and roughly corresponding to the firing of the first and the second spike, the H1 neuron effectively uses much of the information

available in the sensory periphery. This means that, for that time window, the fly's brain approaches an ideal noiseless movement sensor. When the information from the second interval is used, at about 45 ms after the step in the example in figure 19, the neuron's performance is about one-third of that of the ideal. Over time the neuron's performance does not seem to reach much higher values of d' and tends to deviate more from the theoretical limit. The measured values of d' seem reasonable, but are never spectacularly high (up to about 4, see figure 7).

Together these findings make sense if we realize that the typical behavioural response time of the fly is of order 30 ms. In other words, the fly seems to use a strategy of making estimates fast, using the input information efficiently over a short time interval only. It would make little sense to try and improve the accuracy of an estimate much beyond the behavioural response time, because by measuring a stimulus and turning in response the fly itself generates another visual movement stimulus, which in turn should be measured by its movement sensors.

On the time scales of interest in the experiment, photoreceptor noise is dominated by photon shot noise which means that the reliability of H1's output must be determined to a large extent by the quality of the optical signal. There is no neuroanatomical or neurophysiological evidence for massive redundancy in arthropod nervous systems. More specifically, for the fly visual system it is known that H1 is unique in its combination of visual field and preferred direction of movement (Hausen 1982), and from the results presented here we may begin to understand why: it makes little sense to use functional duplicates of any neuron that performs close to perfectly when compared with the noise levels inherently present in the stimulus. It remains to be seen to what extent this conclusion can be generalized, to higher contrasts and light intensities for example, but one should at least be cautious in interpreting the variability of response of a single neuron in terms of noise generated within the nervous system itself.

We are greatly indebted to Hans van Hateren for his help in getting the photoreceptor data and to Ben Pijpker for constructing the stimulus pattern generator. Henk Zaagman provided helpful advice and most of the infrastructure. This work was supported in part by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Barlow, H.B. 1980 The absolute efficiency of perceptual decisions. *Phil. Trans. R. Soc. Lond.* B **290**, 71–82.

Barlow, H.B. 1981 The Ferrier Lecture. Critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond.* B **212**, 1–34.

Barlow, H.B. & Levick, W.R. 1969 Three factors limiting the reliable detection of light by retinal ganglion cells of the cat. *J. Physiol., Lond.* **200**, 1–24.

Beersma, D.G.M., Stavenga, D.G. & Kuiper, J.W. 1975 Organization of visual axes in the compound eye of the fly *Musca Domestica* L. and behavioural consequences. *J. comp. Physiol.* **102**, 305–320.

Bialek, W. 1992 Optimal signal processing in the nervous system. In *Princeton lectures in biophysics* (ed. W. Bialek), pp. 321–401. Singapore: World Scientific.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R. & Warland, D. 1991 Reading a neural code. *Science, Wash.* **252**, 1854–1857.

Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A. 1992 The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**(12), 4745–4765.

Buchner, E. 1976 Elementary movement detectors in an insect visual system. *Biol. Cyber.* **24**, 85–101.

Buchner, E. 1984 Behavioural analysis of spatial vision in insects. In *Photoreception and vision in invertebrates* (ed. M.A. Ali), pp. 561–622. New York: Plenum Press.

Bullock, T.H. 1970 The reliability of neurons. *J. gen. Physiol.* **55**, 565–584.

Eckhorn, R. & Pöpel, B. 1974 Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. *Kybernetik* **16**, 191–200.

Geisler, W.S. 1989 Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* **92-2**, 267–314.

Götz, K.G. 1964 Optomotorische Untersuchung des visuellen Systems einiger Augmenmutanten der Fruchtfliege *Drosophila*. *Kybernetik* **2**, 77–92.

Götz, K.G. 1965 Die optischen Übertragunseigenschaften der Komplexaugen von *Drosophila*. *Kybernetik* **2**, 215–221.

Green, D.M. & Swets, J.A. 1966 *Signal detection theory and psychophysics*. New York: Wiley.

van Hateren, J.H. 1985 The Stiles–Crawford effect in the eye of the blowfly, *Calliphora erythrocephala*. *Vision Res.* **25**, 1305–1315.

Hausen, K. 1982 Motion sensitive interneurons in the optomotor system of the fly. I. The horizontal cells: Structure and signals. *Biol. Cyber.* **45**, 143–156.

Hecht, S., Shlaer, S. & Pirenne, M.H. 1942 Energy, quanta, and vision. *J. gen. Physiol.* **25**, 819–840.

Kirschfeld, K. 1967 Die Projektion der optischen Umwelt auf das Raster der Rhabdomere im Komplexauge von *Musca*. *Expl Brain Res.* **3**, 248–270.

Land, M.F. & Collett, T.S. 1974 Chasing behaviour of houseflies (*Fannia canicularis*). A description and analysis. *J. comp. Physiol.* **89**, 331–357.

Laughlin, S.B. 1989 The role of sensory adaptation in insect compound eyes. *J. exp. Biol.* **146**, 39–62.

Leutscher-Hazelhoff, J.T. 1975 Linear and non-linear performance of transducer and pupil in *Calliphora* retinula cells. *J. Physiol., Lond.* **246**, 333–350.

Levick, W.R., Thibos, L.N., Cohn, T.E., Catanzaro, D. & Barlow, H.B. 1983 Performance of cat retinal ganglion cells at low light levels. *J. gen. Physiol.* **82**, 405–426.

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

Mircea, A. & Sinnreich, H. 1969 Distortion noise in frequency-dependent nonlinear networks. *Proc. Inst. Electron. Engrs* **116**, 1644–1648.

von Neumann, J. 1956 Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata studies* (ed. C.E. Shannon & J. McCarthy), pp. 43–98. Princeton: Princeton University Press.

Parker, A. & Hawken, M. 1985 Capabilities of monkey cortical cells in spatial-resolution tasks. *J. opt. Soc. Am.* A **2**, 1101–1114.

Poggio, T. & Reichardt, W. 1973 Considerations on models of movement detection. *Kybernetik* **13**, 223–227.

Potters, M. & Bialek, W. 1994 Statistical mechanics and visual signal processing. *J. Phys., Paris*, **4**, 1755–1775.

Reichardt, W. 1957 Autokorrelations-Auswertung als

Funktionsprinzip des Zentralnervensystems. *Z. Naturf.* **12b**, 448–457.

Reichardt, W. 1961 Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory communication* (ed. W.A. Rosenblith), pp. 303–317. Cambridge: MIT Press.

Reichardt, W. & Poggio, T. 1976 Visual control of orientation behaviour in the fly, part I. A quantitative analysis. *Q. Rev. Biophys.* **9**, 311–375.

Rose, A. 1948 The sensitivity performance of the human eye on an absolute scale. *J. opt. Soc. Am.* **38**, 196–208.

de Ruyter van Steveninck, R.R. 1986 *Real-time performance of a movement-sensitive neuron in the blowfly visual system.* Thesis, Rijksuniversiteit Groningen, The Netherlands.

de Ruyter van Steveninck, R.R. & 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 **234**, 379–414.

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

van Santen, J.P.H. & Sperling, G. 1984 Temporal covariance model of human motion perception. *J. opt. Soc. Am.* A **1**, 451–473.

Savage, G.L. & Banks, M.S. 1992 Scotopic visual efficiency: constraints by optics, receptor properties and rod pooling. *Vision Res.* **32–4**, 645–656.

Smakman, J.G.J., van Hateren, J.H. & Stavenga, D.G. 1984 Angular sensitivity of blowfly photoreceptors: intracellular measurements and wave-optical predictions. *J. comp. Physiol.* **155A**, 239–247.

Stavenga, D.G. 1979 Pseudopupils of compound eyes. In *Handbook of sensory physiology VII/6A* (ed. H. Autrum), pp. 357–439. Berlin, Heidelberg, New York: Springer-Verlag.

Tapiovaara, M. 1990 Ideal observer and absolute efficiency of detecting mirror symmetry in random images. *J. opt. Soc. Am.* A **7**, 2245–2253.

Tolhurst, D.J., Movshon, J.A. & Dean, A.F. 1983 The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785.

Van Trees, H.L. 1967 *Detection, estimation, and modulation theory. Part I.* New York: Wiley.

van der Velden, H.A. 1944 Over het aantal lichtquanta dat nodig is voor een lichtprikkel bij het menselijk oog. *Physica* **11**, 179–189.

de Vries, Hl. 1943 The quantum character of light and its bearing upon threshold of vision, the differential and visual acuity of the eye. *Physica* **10**, 553–564.

Wehner, R. 1981 Spatial vision in arthropods. In *Handbook of sensory physiology VII/6C* (ed. H. Autrum), pp. 287–616. Berlin, Heidelberg, New York: Springer-Verlag.

Wiener, N. 1958 *Nonlinear problems in random theory.* Cambridge: MIT Press.

Zaagman, W.H., Mastebroek, H.A.K. & Kuiper, J.W. 1977 Receptive field characteristics of a directionally selective movement detector in the visual system of the blowfly. *J. comp. Physiol.* **116**, 39–50.

Zaagman, W.H., Mastebroek, H.A.K. & de Ruyter van Steveninck, R.R. 1983 Adaptive strategies in fly vision: on their image-processing qualities. *IEEE Trans. Syst. Man Cyber. SMC* **13**, 900–906.

Received 12 September 1994; accepted 4 November 1994