Complex Spike-Event Pattern of Transient ON-OFF Retinal Ganglion Cells

Martin Greschner,1,2 Andreas Thiel,1 Jutta Kretzberg,1 and Josef Ammermüller1

1Department of Biology AG Neurobiology, Carl-von-Ossietzky Universität Oldenburg, Oldenburg, Germany; and 2The Salk Institute for Biological Studies, San Diego, California

Submitted 26 October 2005; accepted in final form 14 July 2006

INTRODUCTION

All visual information has to be transmitted to the brain by the spike trains of the retinal ganglion cells (RGCs). It is still under discussion, however, which of their response features are used for coding. Since the pioneering work of Lord Adrian (Adrian 1926) on the stretch receptor, the information about sensory stimuli was assumed to be encoded by the mean firing rate. In recent years, however, the importance of spike patterns for neuronal coding has been intensely debated, both relating to patterns of activity across neurons in the ensemble (e.g., Latham and Nirenberg 2001; Schneidman et al. 2003) and in terms of spike timing patterns of individual cells (reviews: Lestienne and Oram 2001; Oram et al. 2002).

For RGCs it was concluded that spike rates could account for most of the encoded information when they are considered on a fine timescale (Koch et al. 2004). In a different study, the temporal response structure of RGCs was found to carry additional information about certain stimulus conditions that cannot be decoded linearly (Passaglia and Troy 2004) as would be the case for a rate code (Theunissen and Miller 1995). Precisely replicating spike doublets and triplets were found to a lesser extent in RGCs than in other parts of the CNS (Lestienne and Tuckwell 1998), but whether these kinds of temporal patterns contain additional information compared with the spike count was doubted even for LGN and V1 (Oram et al. 1999). On the level of RGC ensembles the discussion of spike patterns focuses primarily on the question of whether RGCs encode visual information independently of each other or by correlated activity across cells (e.g., Meister et al. 1995; Nirenberg et al. 2001; Puchalla et al. 2005). Coming from a theoretical background, it has also been proposed that the rank order with which RGC population members respond to visual stimulation could enable a fast stimulus reconstruction (review: van Rullen and Thorpe 2002).

In experimentally measured RGC activity, complex temporal structures are clearly visible. Many RGCs respond to temporally fluctuating light stimuli with discrete spike events, consisting of single spikes or bursts of spikes, superimposed onto low background activity (e.g., Ariel et al. 1983; Balasubramanian and Berry 2002; Berry et al. 1997; Koch et al. 2004; Kuffler 1953; Schwartz 1973). For repetitive stimulation, these highly specific patterns of spike events occur precisely timed with a jitter of a few milliseconds or less (Berry et al. 1997). Spike events are triggered by a broad range of stimuli, including flashes of different spatial extent (Schwartz 1973), random full-field flicker (Berry et al. 1997), and naturalistic stimuli (Meister and Berry 1999; van Hateren et al. 2002). This finding suggests that RGCs use temporal coding in addition to rate information (reviews: Meister and Berry 1999; Victor 1999). Whether the fine structure of spike timing plays a role in coding, however, still needs to be investigated.

Here we focus on retinal coding on the level of spike events. Using stimuli resembling saccadic eye movements (abrupt changes of light intensity followed by periods of constant light intensity), we regularly found that transient ON-OFF RGCs of the turtle retina respond with two or more spike events to each intensity transition. Both the number of spikes and the timing of the events depend in a characteristic way on stimulus intensity. Even though response patterns consisting of multiple spike events were observed in RGCs of lower vertebrates (Schwartz 1973) and mammals (Ariel et al. 1983), only responses consisting of a single spike event were investigated in the context of retinal coding (Balasubramanian and Berry 2002; Berry and Meister 1998; Berry et al. 1997; Keat et al. 2001). Thus the first goal of this paper is to document the
retinal spike-event patterns and to analyze whether they could improve stimulus discrimination.

Our second goal is to develop a minimal model reproducing the temporal structure of the observed spike-event patterns. Because models consisting of a linear filter and a static non-linear transfer function (Meister and Berry 1999; Sakai 1992) fail to perform this task, the ON-OFF RGC subpopulation studied here requires a more complex model. A cascade of multiple filters including a nonlinear gain-control mechanism (van Hateren et al. 2002; Victor 1987) and spike rate adaptation sufficiently explains the origin of spike-event patterns for a wide range of different stimulus conditions.

METHODS

Electrophysiological recording and light stimulation

RGC activity was recorded extracellularly in 10 isolated turtle (Pseudemys scripta elegans) retinae. Animals were killed according to the guidelines of the University of Oldenburg Ethical Committee and to ECC rules (86/609/ECC). We made recordings in isolated retinae as described (Greschner et al. 2002), using the Utah 100-electrode array (Cyberkinetics, Foxborough, MA).

Light stimulation was performed via optical bench by a white LED light source (LXHL-FW6C; Luxeon, San Jose, CA), controlled by a stand-alone programmable microprocessor, which allowed for precise timing. The intensity of the light stimuli was chosen from sets of 17 or 15 intensity levels. Starting with a maximal illuminance of 1,000 lux, the intensity at a given level was half the value of the next higher level. Thus either 4.8 or 4.2 log units were covered, exceeding the range of 2.5–3 log units in natural environments (van Hateren 1997), and probing the cell responses from the detection limit up to saturation. Four types of light stimulation (flashes, flicker, moving gratings, and spots) were used.

1) Light flashes (Fig. 1, A and B): Spatially homogeneous flashes of 200 ms duration were presented at 0.5 Hz.

2) Full-field random flicker (Fig. 1D): To minimize adaptation effects and to cover more intensity combinations, sequences of slow full-field flicker were used. A new light intensity was chosen every 500 or every 1,000 ms. Each of the intensity transitions was presented 30 times in random order. Four different stimulus conditions and

FIG. 1. ON-OFF transient ganglion cell subtype displays complex spike-event patterns in response to abrupt intensity changes. In all panels, stimulus intensities are symbolized by gray levels, with lighter gray corresponding to higher light intensities. A: spike-event patterns depend on the intensity of spatially homogeneous light flashes. Spikes of one cell are shown in a raster plot and as a peristimulus time histogram (PSTH). B: arranging series of light flashes according to light intensities reveals how the latencies of the events depend on the flash intensity (see RESULTS). C: similar patterns occur during intensity changes induced by simulated saccades across a contrast border. To simulate saccades, contrast borders of different intensities (indicated by gray level in figure background) were moved over the retina. Gray curve indicates the time course of the contrast border position relative to the receptive field center of the cell. A one-dimensional spike-triggered average (STA) is shown at the right of the figure as a representation of the cell's receptive field; lighter gray indicates higher STA amplitude. D: cell responses to stimulation with intensity steps of 1 s length presented in random order. An example for a short stimulus sequence and responses of one cell are indicated in the top row. Four panels below show the PSTHs of responses to 4 stimulus conditions (see RESULTS) with 17 combinations of light intensities, averaged over 30 trials and 5 cells (bin = 2 ms). Dotted lines connect the PSTHs of the 4 stimulus conditions to corresponding examples of stimulation and cell response (see METHODS for procedure to extract data from flicker experiments). Each row in the PSTH panels shows the color-coded (white = 0 Hz, black = 121 Hz) average spike rates (PSTH) in response to a specific light-intensity transition, starting with the time of the transition. Respective start and end intensities of the transition are symbolized as a gray color image to the left of each figure. Rows showing individual PSTHs for a specific intensity transition are arranged in the figure according to increasing intensity (see METHODS).
corresponding responses (Fig. 1D) were extracted from the slow sequences for systematic analysis and for easier comparison with the flash data.

For condition 1, termed “on-start intensity fixed,” all intensity steps starting at the lowest intensity were used to select intervals from a cell’s complete spike train, starting at the time of intensity transition. To display the responses in a peristimulus time histogram (PSTH; Fig. 1D) or a raster plot (Fig. 5A), the spike train slices were ordered systematically with the 14 different end intensities in increasing order. Start and end intensities are symbolized as gray-scale images next to the raster plots. An analogous procedure was used to extract and systematically with the 14 different end intensities in increasing order. Start and end intensities are symbolized as gray-scale images next to the raster plots. An analogous procedure was used to extract and display the responses to condition 2, termed “off-end intensity fixed,” consisting of transitions from various higher start intensities to the lowest end intensity. These first two conditions correspond to the light-on and light-off phases of the flash stimulation. Complementary to these conditions, the maximum light intensity was used as fixed end intensity for condition 3 (on-end intensity fixed) or start intensity for condition 4 (off-start intensity fixed). Because the light intensity increased in both the first and the third conditions, the RGCs reacted with on responses. The intensity decreases in the second and fourth conditions triggered off responses.

For the calculation of the spike-triggered average (STA) needed in the linear–nonlinear (LN) model, fast sequences with intensity changes every 25 ms were used. Sequences of slow flicker were interleaved with periods of fast flicker (Fig. 3A).

3) Moving gratings to mimic saccadic eye movements (Fig. 1C): Square-wave gratings with the same range of intensities as in the flash and the flicker experiments were moved on the retina to simulate saccades. The movement was generated with a computer-controlled miniature mirror system (Datronic, Rastede, Germany). The saccadic velocity increased linearly during half of the saccade duration and decreased afterward. For the cell shown in Fig. 1C the velocity of the contrast border over the receptive field center was 106°/s (Northmore and Granda 1991) and the amplitude was set to about 29° to suit the dimensions of the recording array and to correspond to the fast component of the optokinetic nystagmus in turtles (Ariel 1997).

4) Light spots to determine receptive field properties: For the receptive field measurements a 200-μm spot was moved randomly over 40 × 40 positions across the area covered by the multielectrode array, and the spatial STA was calculated (Eckhorn et al. 1993). A two-dimensional Gaussian was fitted to the STA to reveal the receptive field center (DeVries and Baylor 1997). The classification of RGC types was based on their center responses; thus the complete area of the multielectrode array was probed with 200-ms flashes of a single 200-μm spot.

Data processing and analysis

Typically, each electrode recorded the activity of several cells with different spike waveforms. We used the supervised k-means clustering software SpikeSorter (Cyberkinetics) for the separation of the spike waveforms. We continued the analysis only for units that had waveforms typical of single-unit activity, which were unequivocal in terms of both amplitude and shape of the action potentials, and showed a clear refractory period in the interspike interval histogram. In general, well-separable multiprototype signals were obtained from 20 to 60 electrodes.

For our analysis of spike-event patterns, we used a subpopulation of transient on-off RGCs. The criteria for the selection of cells were clear on-off responses to centered 200-μm-spot stimulation and the occurrence of spike-event patterns during light-on and after light-off (Fig. 1, A and B). To characterize the spike events quantitatively, we averaged the simultaneously recorded responses from all members of the subpopulation and smoothed them by Gaussian filtering. The spike events were defined as the time interval in which the smoothed data were above a given threshold. The maximum of the smoothed data within this period was taken as the event’s time stamp. Response latencies of the individual cells within the population were determined relative to this time stamp. The width of the Gaussian filter and the threshold for the definition of spike events were adjusted for maximal stimulus estimation results, as described below, but variation of these parameters did not qualitatively influence the final estimation results.

To test whether the observed response patterns could contribute to intensity estimation, we applied linear discriminant analysis (LDA; for reviews see, e.g., Duda et al. 2001; Fukunaga 1999). LDA is a method to find the linear combination of features that best separates different classes and it was previously used to analyze neuronal data from multielectrode recordings (e.g., Fernandez et al. 2000; review: Nicolelis 1998). In the case of two classes, LDA can be visualized as projecting the data onto a single vector, which allows optimal separation. For a given assignment of the data points to L ≥ 2 classes, the scatter of data points within the classes and the scatter between classes...
are calculated. The method determines a set of linear equations performing a projection of the data set that minimizes the within-class scatter and maximizes the scatter between classes. In our case, the different classes we wanted to separate were 56 intensity transitions that occur during slow flicker stimulation (four conditions with 15 different start or end intensities each; the four transitions that occurred in two different conditions were pooled into one class each; Fig. 1D). Because orthogonality of separation features is not assumed by the LDA method, we were able to use latency and spike count as features for the separation of classes. For each individual RGC the total response and separate spike events were used for analysis (see RESULTS and Fig. 2, insets a–d). Combinations of these features allowed (e.g., Fig. 2, insets c–f) us to determine their relative importance for discrimination. When classification was based on more than one cell, linear combinations of the response features of the individual cells were used. For instance, for the classification shown in Fig. 2e total spike rate and response latency were considered for 20 cells, and thus the optimal linear combination of 40 features was determined. In this example, the data set is transformed by a set of linear equations

$$Z_c = W_{c0} + W_{c1}R_1 + W_{c2}R_2 + \cdots + W_{c20}R_{20} + W_{c21}L_1 + \cdots + W_{c20}L_{20}$$

where $R_j$ is the spike rate and $L_j$ is the latency of cell $j$; $W_{cj}$ is the optimal transformation coefficients; and $W_{c0}$ is a constant for class C. According to these equations, each data point (consisting of all response features of all cells determined in one stimulus trial) is assigned to the class C yielding the largest discriminant $Z_c$ of all classes.

All classification results presented were based on jackknife procedures, including permutations of cells in a population of a given size. In this procedure, all but one of 30 trials were used to determine optimal coefficients for the linear combination of features. The response of the remaining trial was used to estimate the intensity transition by which it was elicited, and the result was compared with the actual stimulus. Repeating this procedure for all trials yields the average percentage of correct classifications.

**Modeling**

We used two different types of models to reproduce the timing of spike events as was found experimentally: a linear–nonlinear model and a more complex cascade model.

Linear–nonlinear (LN) models are widely used to reproduce RGC responses (reviews: Meister and Berry 1999; Sakai 1992) by convolving the stimulus with a linear filter and applying a nonlinear function afterward. The linear filter is obtained as the spike-triggered average (STA) of the stimulus, that is, the average of intervals of the stimulus time course preceding a spike. Here, the STA was calculated in the frequency domain (van Hateren and Snippe 2001). Using a bin width of 2 ms, we averaged 680 intervals, each of 8,192 ms length, of the resulting spike data of the fast full-field flicker stimuli and of the Gauss-filtered stimuli ($\sigma = 5$ ms). Convolution of the linear filter with the stimulus yields a generator potential that is transformed into spike rates by the nonlinearity function. The nonlinearity function was fitted to the experimentally observed spike rates by a parameterized form of the cumulative normal density function (Chichilnisky 2001). A reasonable extension of the LN model to reproduce responses of ON-OFF cells is to consider separate ON and OFF pathways. To obtain distinct STAs for ON and OFF responses we separated the stimulus intervals into two classes before calculating the average. To calculate the ON STA, we used responses to stimulus intervals with the average illumination increasing in subsequent periods of 70 ms. To calculate the OFF STA, intervals with decreasing average stimulus intensity were used. As an additional test, PSTHs for both transitions were inspected to confirm that the cells respond truly to both ON and OFF stimulus transitions. The two-dimensional nonlinearity, which is necessary for modeling distinct ON and OFF pathways, was calculated with the same procedure used for the one-dimensional nonlinearity and linear interpolation of the intermediate values.

The cascade model consists of sequences of low-pass and high-pass filters, a compressive nonlinearity and a rectification, which are described in more detail in RESULTS. The model’s 15 free parameters were adjusted by error minimization using the Nelder–Mead Simplex algorithm (Press et al. 1992). A spike train distance metric was used as an error function, which indicates the difference between the experimental data and the actual model prediction (Aronov 2003). Parameter sets were additionally punished if they predicted a considerably smaller range of contrasts to elicit spikes than that experimentally observed.

The spike timings of the experimental data that were used for error calculation were determined by averaging over several responses to identical stimulation and applying a threshold function, which was adjusted by visual inspection. To reduce the susceptibility of the parameter search to getting stuck in local minima of the error value, new start parameters were randomly chosen repetitively around the current best parameter set. The radius of the start parameter variation was reduced with iteration number. The model’s temporal resolution was set to 1 ms. Model simulations and all analysis routines were performed with Matlab 6.5 (The MathWorks, Natick, MA).

**RESULTS**

**Complex event patterns**

The transient ON-OFF retinal ganglion cells studied here respond to stimulus transients with two or more discrete spike events. Each event consists of one or a few spikes, but because of the low spontaneous spiking rate, events show up clearly in repetitive trials (Fig. 1A), leading to sharp peaks of higher firing probability in the PSTH. The timing of these events depends reproducibly on the stimulus contrast, as can be seen for a series of flashes of different light intensities superimposed onto identical low-background illumination. When plotting a cell’s responses in the order of increasing flash intensity, the events form a continuous pattern (Fig. 1B) consisting of two main events forming the ON response (after light onset) and two other main events during the OFF response (after the offset of light stimulation). The latency of the first ON event decreases monotonically with increasing contrast, as observed before in several studies (e.g., Kuffler 1953; Levick 1973). The subsequent ON event, however, shows a more complex S-shaped dependency on flash intensities, characterized by a shorter latency for intermediate intensities and longer latencies for low and high intensities. The latency of the first OFF event does not show the simple contrast dependency of the first ON event. It rather resembles the form of the letter “C” because latency is shortest for intermediate contrasts and increases drastically for high contrast. The second OFF event, however, displays a similar nonlinear dependency as the second ON event. A subpopulation of transient ON-OFF RGCs reliably showing such patterns was regularly recorded in all studied retinae. They are characterized by a clear ON-OFF activity in response to centered 200-µm-spot stimulation, and by spike-event patterns during light-on and after light-off as shown in Fig. 1A and B. Event patterns were slightly more pronounced in response to spatially homogenous, full-field light flashes compared with small centered spots.

The spiking patterns also occur in more naturalistic stimulation paradigms such as simulated saccadic eye movements (Fig. 1C). Saccades were simulated using spatial gratings of...
fixed low and varied high intensities. Moving a stimulus region of high light intensity into the receptive field, rather than turning the light on instantaneously as in the flash paradigm, results in a smoother increase of overall light intensity in the receptive field. Nevertheless, quite similar spiking patterns were elicited.

Additionally, we stimulated with random, full-field flicker to minimize adaptation effects and to cover more combinations of intensity transitions. Using low-frequency flicker (Fig. 1D; see METHODS) allowed us to unambiguously assign each spike event to an individual intensity transition. To analyze the data analogously to the flash data, the responses to continuous flicker were split into sequences elicited by separate intensity steps, characterized by the respective intensities before and after the intensity change. To display the responses, these intensity pairs were combined in blocks of identical trials and sorted according to the start and end intensities. Using this method, four different conditions can be distinguished. The first condition (Fig. 1D, ON-start intensity fixed) is equivalent to the ON component of the flash responses; it represents the response to an increase in intensity starting at the background illumination level of the flash experiments. Corresponding to the OFF component of the flash stimulation, the second condition (Fig. 1D, OFF-end intensity fixed) contains the responses to a reduction in light intensity. Responses to both flashes and flicker stimulation elicit qualitatively the same event patterns for ON and OFF responses, even though the light adaptation level of the retina is different as a result of the higher mean light intensity in flicker experiments. Complementary to the two conditions corresponding to the flash experiments, we considered ON stimuli with fixed end intensity and OFF stimuli with fixed start intensity (Fig. 1D), with the fixed intensity being the maximum light level of our stimulus set. Interestingly, under condition 3 “ON-end intensity fixed” the latency of the first event is nearly independent of the contrast, whereas the latency of the second event strongly decreases with decreasing change in intensity. Under condition 4, “OFF-start intensity fixed,” the latency seems to be rather independent of contrast and the large temporal jitter causes the event pattern to be barely visible.

Because transitions of light intensity trigger highly specific spike-event patterns, the question arises whether these events carry additional information about the stimulus.

**Stimulus estimation**

We performed linear discriminant analysis to compare how well different intensity transitions can be discriminated based on features of RGC spike responses. For the first discrimination step, spike events had to be identified. To restrict our analysis to decoding strategies available in a behavioral context, in which the nervous system cannot average over identical repetitions, events were detected by averaging over the population of neurons. A spike event was defined as the time interval in which the population activity exceeded a threshold (see METHODS). In the second step of the stimulus estimation procedure, we determined the following response features for each individual cell: 1) the response latency relative to the population event; 2) the total number of spikes in the response; 3) the time between the onset of the first and second events (relative latency); 4) the number of spikes in the first and the second spike event of the response (Fig. 2, insets). In the third step of the analysis, these response features were used individually (Fig. 2, a and b) or in linear combination (Fig. 2, c–f) to estimate the intensity transition that elicited the response. The percentage of correct classifications was determined in a jackknife procedure, by performing the linear discriminant analysis for multiple trials and several cells and averaging the results (see METHODS).

The percentage values for correct classification given in Fig. 2 may seem small at first glance. The absolute values, however, depend strongly on the number and kind of classes to be distinguished. In the present case, 56 different stimulus classes had to be discriminated (four conditions with different start or end intensities each, Fig. 1D), leading to a chance level of <2% (Fig. 2, dashed line). Because cell responses were saturated for several of these stimulus conditions (Fig. 1D), their responses did not differ significantly and could therefore not be discriminated. Moreover, one should bear in mind that this method is not meant to measure the upper limit of classification performance the brain could achieve by optimally decoding the RGC responses, but rather to show the relative importance of different response features for decoding.

In Fig. 2 the classification performance based on different response features is shown for an increasing number of cells. For all response features and their combinations, classification performance improved drastically with increasing cell number and was not saturated for a population size of 20 cells. We found no significant differences in classification performance of the individual cells included in the study. Comparing the performance based on different response features for a constant number of cells shows the relative importance of the features. Latencies of the first spike response (Fig. 2a) and total spike count (Fig. 2b) both allow stimulus discrimination far above chance level, even if only small numbers of cells are considered. This finding is remarkable insofar as the classification is based on only a narrow range of different spike counts elicited by the set of stimuli. For the classification based on latencies, no external triggers are needed, but the small differences between the latencies of individual neurons and the population response are sufficient as the basis for stimulus discrimination. However, compared with using the spike count or the relative latency of the total response, the ability to discriminate between different stimulus conditions is significantly improved by using features specific for the two separate spike events. For instance, estimation performance was increased by 31.0% (SD 3.5) by using the relative latency of the second event in addition to the relative latency of the first event (Fig. 2, a compared with c). The estimation based on spike count was improved by 24.4% (SD 3.9) by considering spikes separately for the two events instead of summing them together (Fig. 2b, b compared with d). Using the relative latencies of both events was superior to the total, summed rate count (Fig. 2, c compared with b). The best estimation result was obtained using both latencies as well as separate rate counts from each event (Fig. 2f). These results show that the second spike event indeed adds information about the stimulus. Moreover, we can conclude that considering temporal information in addition to spike rates yields optimal classification results.

**Evaluation of simple models**

To approach an explanation for the spike-event patterns, we tried to develop a computational model to reproduce this
behavior as simply as possible. Linear–nonlinear (LN) models are a common approach to describe RGC responses (Meister and Berry 1999; Sakai 1992) by convolving the stimulus with a linear filter to obtain a generator potential that is transformed nonlinearly into a continuously varying spike rate (Fig. 3A, inset, LN model). The time reverse of the filter intuitively represents the stimulus feature to which the cell responds best, leading to higher amplitudes in the generator potential if this feature is strongly represented in the stimulus. The generator potential is transformed into the spike response by the nonlinearity function, capturing properties of spike generation such as rectification and saturation.

For our LN model, we calculated the linear filter and the nonlinear function based on experimentally determined RGC responses to fast, random full-field flicker. In the raster plot and the PSTH of the experimental data, the double events of the patterns are clearly visible, especially during the slow flicker periods that were interleaved with fast flicker stimulation (Fig. 3A, experiment). The resulting LN model with the linear filter and the nonlinear function shown in Fig. 3B yields a relatively poor reproduction of the experimental data. It correctly generates most events during fast flickering stimulation, but fails in predicting the second main event during slow flicker (Fig. 3A, LN model).

One reason for the poor performance of the LN model is the filter characteristic necessary to model responses of ON-OFF cells, which are maximally excited both by an increase and by a decrease in light intensity. This behavior is difficult to represent in a single filter because the ON and OFF components partially cancel each other. The remaining filter (Fig. 3B) results from the difference in signal transduction speed between the OFF and ON pathways and their slight imbalance, rather than representing the main stimulus features triggering spikes.

To overcome the problem of interactions between ON and OFF components of the filter, a reasonable elaboration of the LN model is to use two separate filters (Fig. 3C). Therefore we separated the spike-triggered stimulus into an ON and an OFF class before averaging. In combination with a two-dimensional (2D) nonlinearity (Fig. 3D), this 2D LN model predicts the events more accurately and is able, in principle, to produce two events in response to a single step (Fig. 3A, 2D LN model). When looking at the responses in more detail, however, it becomes apparent that the small second event in the generator potential sometimes passes the non-
linearity when no spike event is visible in the experimental data (Fig. 3A, 2D LN model).

In principle, the LN model could be further elaborated to accomplish a better reproduction of the experimental findings. If a quadraphasic instead of biphasic linear filter is used, the generator potential more reliably produces a second peak, which can be transformed into spike events if it crosses the nonlinearity threshold. However, the model still predicts monotonically increasing latency with decreasing stimulus contrast for both events instead of the specific S-shaped latency behavior of the second event. Therefore instead of improving the LN model by additional components, we chose to construct a model composed of filter cascades, which turned out to nicely reproduce the experimentally observed spike-event patterns (Fig. 3A, cascade model).

Cascade model

Like LN models, cascade models of retinal information processing first deduct the time course of a generator potential from the stimulus and then transform this generator potential into spike responses. In contrast to LN models, these steps can consist of a sequence of several operations in cascade models. Using as few cascade operations as possible, we found that three main steps (Fig. 4) are required to reproduce the time course of the observed spike-event responses for the four different stimulus conditions introduced in Fig. 1D. The first two of these steps lead to a generator potential, which can be transformed reliably into two spike events. In these two steps, ON and OFF components are calculated separately. The third step translates the combined ON and OFF generator potentials into spikes.

1) The time course of the stimulus is transformed into separate ON and OFF components of the generator potential by applying the sequence of operations shown in the left half of Fig. 4: First, a low-pass filter smooths the intensity transitions of the stimulus and causes the generator potential to rise faster with increasing contrast of the intensity transition, ultimately leading to decreased latency of the first spike event with increasing contrast. The nonlinear compressive function enables the model to respond over the whole range of stimulus intensities. In the OFF pathway, the signal is inverted after being compressed. For the sake of simplicity, we chose to achieve the correct latency of the first event with a simple delay, rather than with a stronger cascade of low-pass filters in combination with a second threshold or a sigmoidal nonlinearity. After the time delay, ON and OFF components of the generator potential are filtered with a high-pass filter that is implemented by subtracting a low-pass filtered version of the signal from the signal itself. This procedure defines the total duration and the transient character of the cell’s response. The subsequent second low-pass filter ensures a smooth characteristic of the generator potential even for conditions in which the compressive nonlinearity has diminished the effect of the first low-pass filter. This sequence of processing steps corresponds to classic cascade models for separately simulated transient ON and OFF cells. The resulting generator potential is shown in gray in Fig. 5B for the different light intensities in the four stimulus conditions we consider. For each of the intensity combinations, this preliminary generator potential consists of only a single peak, with the amplitude and the slopes of rising and falling phases depending on the stimulus contrast.

2) To ensure that two spike events are generated in response to a light-intensity transition, the generator potential needs to be biphasic, with each of the two peaks triggering one event. To obtain a dip in the time course of the generator potential (Fig. 5B, black lines), the signal is fed into a divisive low-pass filtered feedforward loop.

3) In the last step, the positive parts of the ON and OFF pathways are finally summed and fed into the spike-generating

---

**FIG. 4.** Minimal cascade model of the ON-OFF transient ganglion cell reproducing the timing of spike-event patterns: \( s(t) \) is the intensity time course of the light stimulus in arbitrary units from \( 2^0 \) to \( 2^{16} \); LP1 is a first-order low-pass filter, with the time constant \( \tau \); NL1 is a compressive nonlinearity of the form: output = \( (2/\pi) \) atan \((k \times \text{input})\) (van Hateren et al. 2002); the signal of the OFF pathway in the lower half of the schema is inverted; \( \Delta t \) is a delay of \( t \) ms; LP2 is a first-order low-pass filter with the time constant \( \tau_2 \), in a subtractive feedforward loop, which thereby forms a high-pass filter; LP3 is a cascade of 2 first-order low-pass filters with the time constant \( \tau_3 \); LP4 is a first-order low-pass filter, with the time constant \( \tau_4 \) and an amplitude \( b_1 \), the output of which is multiplied as \( 1 - \) output (output \( > 0 \)) to the signal in a feedforward loop; \( [\ ] \) depicts the half-wave rectification of the signal, returning only positive values. Spike-generating mechanism consists of a threshold function (depicted by \( \times \), a comparator with the logical output 1 if the value of the preceding model exceeds the control value of the spiking mechanism, and 0 otherwise) and 2 leaky integrators; first-order low-pass filters LP5 and LP6, with time constants \( \tau_5 \) and \( \tau_6 \), respectively; and an amplitude \( b_2 \) for LP5; \( T_0 \) is a constant additive signal added to the threshold; \( r(t) \) is the model response. Different parameters in the ON and OFF pathway are indicated by indices. Gray and black arrowheads indicate the processing steps for which the time course of the generator potential is shown in Fig. 5B.
mechanism. This mechanism consists of a threshold, a leaky integrator with a fast time constant modeling the refractory period, and a slow leaky integrator providing a threshold adaptation (Lankheet et al. 1989). If the generator potential exceeds the threshold, a spike is produced and the threshold is temporarily increased by the two integrator mechanisms.

The modeling results and the spike responses of the RGC are compared in Fig. 5A. For simplicity, we did not include a stochastic component in the model, and thus all trials produce identical results. The model simulates the temporal characteristics of the experimental responses quite well for the four stimulus conditions.

In the first condition (Fig. 5A, ON-start intensity fixed) the latency of the first spike event displays the normal decrease for higher stimulus contrasts. This effect results from a combination of the first low-pass filter output, which is smaller in amplitude and less steep for higher stimulus contrasts, and the threshold function. In condition 2 (Fig. 5A, OFF-end intensity fixed) the latency of the first event is longer at high contrast steps than at intermediate ones, as can also be seen in the OFF response of the flash paradigm (Fig. 1B). This behavior is a consequence of the saturating nonlinearity affecting the low-pass filtered signal. For combinations of high-start and low-end intensities, the first low-pass filter generates a smoothly decreasing generator potential, in which slight differences are canceled out when the signal is compressed by the nonlinearity. In consequence, the inverted OFF component increases sluggishly during OFF responses (condition 2), whereas the ON component rises steeply during ON stimulation (condition 1) with the same difference in light intensities. Thus even though the time constant of the low pass in the OFF pathway is distinctly shorter, the spike threshold is crossed later in condition 2. The delay introduced by this interaction of the low-pass filter and the nonlinearity depends on the start intensity in condition 2, shortening the latency of the first OFF event with decreasing intensity. For smaller intensity steps, however, this effect is outweighed by the counteracting mechanism of a shallower generator potential slope for lower contrasts, which also produces the increase in latency in condition 1.

In both conditions "ON-start intensity fixed" and "OFF-end intensity fixed," the latency of the second event shows the S-shaped contrast dependency, as can also be seen in the flash data in Fig. 1B. The timing of the second event is mainly caused by the dip in the generator potential produced by the divisive loop (Fig. 5B, after div. loop). At lower stimulus contrasts, this dip is less pronounced and the second event occurs earlier. However, the threshold adaptation of the spike generation prevents the second event from getting too close to the first event. Therefore the latency of the second event increases again when the first event is delayed by lower contrasts. In condition 2, the latency of the second event shows an additional abrupt change at high contrast conditions. Here, the dip is least distinct and the occurrence of the second event is mainly dependent on the threshold adaptation. The abrupt changes in the latency of the second event are caused by the transition between the regime in which the threshold adaptation...
mechanism defines the latency and the regime of the divisive loop.

Condition 3 (Fig. 5A, on-end intensity fixed) and condition 4 (Fig. 5A, off-start intensity fixed) are both characterized by very similar response patterns for a large range of contrasts. This is caused by saturation of the cell at the upper bound of the actual range of contrast adaptation. The amount of saturation in the model is decreased by the compressive nonlinearity that has to be fitted to suit all stimulus conditions. As can be seen from Fig. 5A, the response range of the model’s second event is slightly too small because the model fails to produce spikes at the lowest contrast level for which the real cell still responds. Simply changing the properties of the compressive nonlinearity, however, would result in an impaired fit for condition 1. To correct for this, the model could be expanded by a second nonlinearity after the subtractive luminance control. The less-precise spike timing of the experimental data in condition 4 (Fig. 5A, off-start intensity fixed) can be explained by the interaction of low-pass filtering and threshold nonlinearity. A shallower rising generator potential leads to more jitter in the timing of the spike threshold crossing if noise is present. The latency of the second event is again primarily dependent on the threshold adaptation, which also results in less-precise timing of the experimental data.

The amount of low-pass filtering required in the model has to be divided into two components: On the one hand, only a low-pass filter applied before the nonlinearity function can introduce a contrast dependency of the first event’s latency in conditions 1 and 2, but not in conditions 3 and 4. On the other hand, the time constant of a single low-pass filter before the nonlinearity function cannot be sufficiently long because this would cause the latency of the first event to be too long in condition 4 and in condition 2 at high contrast. The exact position of the second low-pass filter, before or after the subtractive luminance control, has no influence.

The ability of the model to simulate the timing of both spike events for all of our stimulus conditions is summarized in Fig. 6. The event latencies of the five cells shown in the histogram of Fig. 1D do not significantly differ from the timing of events predicted by the model. Thus the model well captures the temporal structure of transient ON-OFF cell responses.

**DISCUSSION**

In this study we have shown that a subpopulation of transient ON-OFF RGCs responds to each rapid intensity change with two distinct spike events. Because these spike events depend on the stimulus contrast in a characteristic way, they can improve stimulus discrimination. Occurrence of these spike-event patterns cannot be explained with linear–nonlinear models which are commonly used for retinal information processing, but require additional nonlinear steps of computation.

**Spike-event patterns occur under natural viewing conditions**

We have shown that transient ON-OFF RGCs produce clear and reliable spike-event patterns. One might wonder whether these patterns are only an effect of an artificial stimulation. However, light conditions similar to slow full-field flicker also occur under natural viewing conditions. During the redirection of gaze arising from saccades the light intensity within the receptive field of many RGCs changes rapidly, whereas between two saccades receptive fields are predominantly stimulated homogeneously. This holds true even in the presence of fixational eye movements because natural images contain mostly low spatial frequencies (Field 1987; van der Schaaf and van Hateren 1996). Moreover, responses consisting of more than one spike event also occur when light stimuli of a smaller spatial extent are used, although they show up less clearly (Ariel et al. 1983; Schwartz 1973; unpublished observation). When the time course of light-intensity changes during saccades was simulated explicitly, we found spike-event patterns similar to those elicited by flashes or full-field flicker (Fig. 1). Furthermore, in VI of primates no systematic difference was detected between cell responses to external changes in light intensity and to intensity changes introduced by real saccades (Gawne and Martin 2002), suggesting that RGC signals are similar in both conditions.

**Spike-event patterns can improve retinal coding**

The classic debate of neural coding has focused on whether neurons use a rate code or a temporal code. For RGCs evidence for the importance of the temporal structure of the response prevails. In particular, the occurrence of distinct spike events is hardly in accordance with the assumption of a continuously varying spike rate carrying all relevant information about the stimulus (Berry and Meister 1998; Berry et al. 1997; Meister and Berry 1999). In agreement with this observation, we found that temporal response aspects of transient ON-OFF RGCs carry additional visual information not conveyed by the average firing rate. For the analysis of temporal aspects of retinal coding, we focused exclusively on response latencies. Whether the fine temporal structure of spike timing within the events is important for stimulus classification remains to be analyzed. In this study, we showed that an ensemble of transient ON-OFF cells carries information about the stimulus in their latencies,
yielding a similar classification performance as obtained for spike counts. Combining latency information and spike counts significantly improves the classification, suggesting that the two types of coding contain information partly complementary to each other. This result fits well with the hypothesis proposed by Thorpe and colleagues, that the rank order of first spikes elicited by members of a population allows faster stimulus estimation than a rate code (Gautrais and Thorpe 1998; review: van Rullen and Thorpe 2002). As proposed by the rank-order hypothesis we investigated in our study a code based on relative latencies (measured as the time difference between response onsets of population members) rather than external triggers to determine absolute latencies (defined as the time difference between stimulus onset and response onset), which are unknown to the nervous system. We found that even during full-field stimulation, the individual cell responses are able to carry information by their slightly differing latencies. One can clearly claim that spatially inhomogeneous stimulation would enhance these differences and a code based on relative latency would become more robust.

Taking each spike event into account separately, instead of using the entire response, improves the ability to discriminate between different light-intensity transitions (Fig. 2). This finding does not necessarily mean that the function of the retinal mechanism producing these patterns is solely temporal coding; the patterns could also be a by-product of a contrast adaptation mechanism. However, we do propose that the additional information coded by the second event can be used for stimulus estimation. One can even argue that if the second event were not used for stimulus estimation, its influence would have to be actively suppressed because, otherwise, it would misleadingly signal an intensity change that did not actually occur. Therefore patterns of spike events could play a significant role in retinal coding. A drawback of using spike-event patterns, however, is the long delay between the events (around 50 ms for contrasts typically occurring under natural viewing conditions). Thus stimulus estimations building the basis for fast reactions should rather rely on the first spike event, whereas the second event could improve stimulus estimation over a longer time period. Moreover, if the responses of transient ON-OFF RGCs are combined with responses of other cell types, presumably more accurate stimulus estimation results can be obtained than those based on one cell type alone (Warland et al. 1997). Because spatial stimulus configuration has less influence on the responses of transient ON-OFF cells than on those of sustained cells, the transient cells might be responsible for encoding contrast information by producing a fast first response and a slower but additionally informative second spike event.

**Modeling spike-event patterns**

In the second part of the paper, our goal was to describe the origin of the spike-event patterns as simply as possible with a computational model. A linear–nonlinear model in the style of previous models for retinal information processing (reviews: Meister and Berry 1999; Sakai 1992) failed to reproduce our experimental data. Even though a model of this type (Keat et al. 2001) can account for the single spike events occurring in ON and in OFF RGCs, it is not suitable for the more complex situation of multiple spike events occurring in ON-OFF cells. The timing of the second event of the spike response is not directly coupled to the time of the light-intensity transition, but depends on the stimulus parameters in a nonlinear way. Thus the cell responses cannot be reproduced by a linear filter model that implicitly assumes a rate code (Theunissen and Miller 1995). Thus the inability of the LN model to reproduce the spike-event patterns provides further support for our hypothesis that RGCs use temporal information to encode visual stimuli.

Unlike the LN model, the more complex cascade model is able to reproduce the time structure of the spike events. It is built of standard components often used in models of the early visual system or psychophysical models (Carpenter and Grossberg 1981; Snippe et al. 2000; van Hateren et al. 2002). In contrast to the physiologically inspired model we recently developed (Thiel et al. 2006), the components of the cascade model do not have exact single counterparts in the retina on a cellular level because each processing step integrates several physiological stages of the retinal pathway. Nevertheless, by studying the influence of individual processing steps on the response patterns, it is possible to ascribe certain characteristics of the patterns, which may at first seem complex, to simple well-known retinal mechanisms.

Retinal mechanisms such as phototransduction, the pathways of the inner segments, and synaptic transmission slightly delay the light signal. This delay is modeled by two low-pass filters, which have to be arranged around the compressive nonlinearity. The combination of low-pass filters and a compressive or saturating nonlinearity function can be regarded as a representation of the photoreceptor (Carpenter and Grossberg 1981; Hennig et al. 2002; van Hateren 2005). In particular, the depletion of cGMP (cyclic guanosine 3’,5’-monophosphate) in response to light makes up a large part of the nonlinear behavior. In combination with the delay and the inverter, this first part of the model can also be seen as representing aspects of the transfer to the bipolar cells, including the different receptors of ON and OFF pathways. The OFF pathway displays a steeper nonlinearity and a smaller delay in the model, as was also found experimentally (Zaghoul et al. 2003). The high-pass filter forms the transient character of the cell’s response. This mechanism can be interpreted as implementing a subtractive luminance adaptation. On a cellular level, this could be provided partly by inhibitory feedback from amacrine cells (Dong and Werblin 1998; Maple and Wu 1998; Thiel et al. 2006; Zhang and Slaughter 1995), by glutamate receptor desensitization (Łukasiewicz et al. 1995), or by varying channel kinetics (Awatramani and Slaughter 2000). The divisive loop can be seen as contrast gain control, which is physiologically well established in the retina (e.g., Baccus and Meister 2002; Shapley and Victor 1978; Werblin and Copenhagen 1974). In our cellular model, contrast gain control and the occurrence of two spike events could be attributed mainly to inhibitory input of fast transient amacrine cells on the RGCs (Thiel et al. 2006).

The spike-generating mechanism involves two processes: 1) a component with a time constant of about 1 ms simulates the absolute and relative refractory period, and 2) a second component, nearly 50-fold slower, models a spike rate adaptation, for which several mechanisms have been discussed, e.g., the modulation of the calcium-gated potassium currents or the voltage-gated high-threshold potassium currents, and the recovery of the fast sodium current (Benda and Herz 2003;
Lankheet et al. 1989). As shown above, this slow component plays an important role for the timing of spike-event patterns.

We found that at different stages of the model additional components could produce more precise predictions, but we always chose the simplest alternative. For example, we chose a deterministic spike-generating mechanism and did not include a stochastic component in the model, even though stochastic effects are known to play a significant role in retinal information processing and stimulus estimation based on RGC responses (e.g., Dhingra and Smith 2004; Dunn et al. 2006; van Rossum et al. 2003). Moreover, we used a fixed compressive nonlinearity and fixed time constants in the feedback loops. For a given parameter set the model is thereby limited to a certain adaptation state (Snippe et al. 2000; van Hateren et al. 2002), but different stages of adaptation can be simulated by parameter adjustment. Although we did not rigorously prove that our cascade model is the minimal model reproducing the timing of spike-event patterns, we were able to assign necessary functions to all of the model’s components and to the order of processing steps.

We can conclude that the timing of spike events originates to a large extent from a combination of gain control loop and spike rate adaptation. The temporal structure of the resulting spike-event patterns, irrespective of whether they are by-products of retinal processing, adds information about the visual stimulus situation not carried by spike rates. Thus spike-event patterns could play an important role in retinal coding also under natural viewing conditions.

ACKNOWLEDGMENTS

We thank M. T. Ahlers for construction of the programmable LED stimulation system and J. Shelley for critical reading of the manuscript.

GRANTS

This work was supported by a grant from the European Community Project CORTIVIS (Studies on the Feasibility of a Visual Cortical Prosthesis) and Deutsche Forschungsgemeinschaft Grants AM 70/10 and KR 2925/1.

REFERENCES


J Neurophysiol • VOL 96 • DECEMBER 2006 • www.jn.org


