

Reverse correlation in neurophysiology

Dario Ringach^a Robert Shapley^b

^a*Departments of Neurobiology and Psychology, Jules Stein Eye Institute, David Geffen School of Medicine,
University of California at Los Angeles, USA*

^b*Center for Neural Science, New York University, USA*

Received 6 April 2003; received in revised form 30 October 2003; accepted 11 November 2003

Abstract

This article presents a review of reverse correlation in neurophysiology. We discuss the basis of reverse correlation in linear transducers and in spiking neurons. The application of reverse correlation to measure the receptive fields of visual neurons using white noise and m-sequences, and classical findings about spatial and color processing in the cortex resulting from such measurements, are emphasized. Finally, we describe new developments in reverse correlation, including “sub-space” and categorical reverse-correlation. Recent results obtained by applying such methods in the orientation, spatial-frequency and Fourier domains have revealed the importance of cortical inhibition in the establishment of sharp tuning selectivity in single neurons.

© 2004 Cognitive Science Society, Inc. All rights reserved.

Keywords: Triggered-correlation; M-sequences; Impulse response; Simple cells; Sub-space reverse correlation; Kernels

1. Introduction—history and background

The article is about reverse correlation in neurophysiology. Reverse correlation is a technique for studying how sensory neurons add up signals from different locations in their receptive fields, and also how they sum up stimuli that they receive at different times, to generate a response. Recently we have developed a variation of reverse correlation that also allows us to study the time evolution of sensory neurons’ responses to different categories of stimuli and have applied this to the study of visual signal processing in the visual cortex. After we introduce the ideas behind the technique, we will discuss various interesting discoveries about

E-mail addresses: dario@ucla.edu (D. Ringach), shapley@cns.nyu.edu (R. Shapley).

0364-0213/\$ – see front matter © 2004 Cognitive Science Society, Inc. All rights reserved.

doi:10.1016/j.cogsci.2003.11.003

sensory function that have come out of experiments that used reverse correlation to study spatial and temporal processing, and some of our new results about the time evolution of categorical responses.

Many sensory neurons can be understood as stimulus-response transducers that are driven by sensory stimulation from the outside world. Such a neuron is quiet or in a background state in the absence of stimulation. Then, when presented with an appropriate stimulus, the neuron is either activated above its background level of activity or in some cases suppressed below background in a more or less consistent manner from one stimulus presentation to the next. When the stimulation ceases, the sensory neuron's activity relaxes back to the background state. This functional description applies to most sub-cortical sensory neurons that have been studied. In the sensory areas of the cerebral cortex there are neurons that behave as sensory transducers according to how we have defined it here, though not all the cortical neurons fit this description. Any cortical neuron that is involved in memory or decisions or action will have some activity that is not stimulus driven, and therefore such a neuron will not fit neatly into the definition of a transducer neuron. The techniques that we will be discussing in this review paper under the heading of "reverse correlation" are only applicable to neurons of the transducer type. These techniques do not apply to cell's whose response depends on memory or decisions or action. Nevertheless, there are many neural transducers that are worth understanding in order to understand how the nervous system works and how neuronal networks can explain aspects of behavior.

2. Linear transducers

The simplest transducers are linear, so let's consider them first. The simplest linear transducers are those with a single input and a single output. We can find out the principles of reverse correlation from these simplest transducers, which we'll call linear-single-input, single-output systems, or LSISOS. As illustrated in Fig. 1a, a LSISOS will respond to a brief pulse of input

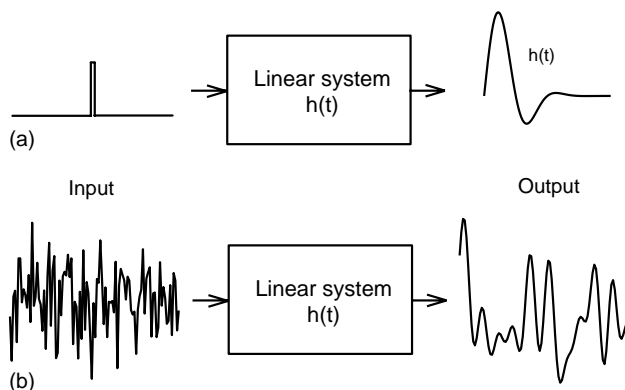


Fig. 1. A linear time-invariant system is characterized by its impulse response, $h(t)$. The figure illustrates two different ways of measuring it. (a) The impulse response can be measured as the response to a brief pulse of unit area, or (b) the impulse response of the system, $h(t)$, can be measured by having a white noise input and cross-correlating the resulting output with the input. The impulse response provides a full characterization of a linear system.

of unit area with a response $h(t)$, its so-called impulse response. It turns out that if the system is linear, once we know the impulse response, we know all there is to know about how the LSISOS will respond to any input. This is because linear systems obey two principles: homogeneity and superposition. Homogeneity means that the response of a stimulus x that has been scaled in intensity by a factor α has to equal to α times the response to x . Superposition means that the response of the cell to a sum of two stimuli, $x + y$, must equal the sum of the responses to the individual stimuli. Any input can be decomposed into a sum of pulses at different times, of different heights; the response of an LSISOS to this sum is simply the sum of its responses to each of the pulses appropriately scaled and translated in time. This process of summation is usually called convolution and it is the basis for the synthesis of responses of LSISO transducers to any input (e.g., see [Bracewell, 1998](#)). The point of this discussion is that if we can measure $h(t)$, the impulse response, then the LSISOS is understood.

3. Cross correlation with white noise

There are many operational ways to measure the impulse response of an LSISOS. One could simply measure the response to a pulse. Or one could measure a frequency response with a sequence of sinusoidal inputs, and recover the impulse response as the Fourier transform of the frequency response (see [Brodie, Knight, & Ratliff, 1978](#) for a physiological example of this procedure). But for various technical reasons (and also theoretical reasons we will address below) there is some advantage to determining the impulse response by using noisy signals as inputs, and measuring the impulse response by means of cross-correlation. The cross correlation function between the input stimulus $S(t)$ and the output response $r(t)$, in a system that is stationary over time, is defined as $C(t') = \langle S(t)r(t + t') \rangle$, where $\langle \cdot \rangle$ means averaging over time. Suppose the stimulus $S(t)$ is a white noise. This means that the stimulus $S(t)$ has a flat power spectrum (it has the same energy at all frequencies), or equivalently that the autocorrelation of $S(t)$ is an impulse function such that $A(t') = \langle S(t)S(t + t') \rangle = \delta(t')$. Here, $\delta(t')$ is the Dirac's delta function which is obtained as the width of a pulse of unit area tends to zero. Then one can prove that, for an LSISOS driven by white noise:

$$C(t') = h(t') \tag{1}$$

that is, the cross correlation function of the response with its white noise input is equal to the impulse response (see [Bracewell, 1998](#)). This is an important theoretical result about white noise analysis techniques for studying transducers (see [Fig. 1b](#)).

Because it is so important for understanding reverse correlation and its applications, we need to study the cross correlation result in some detail. First, note that in the defining equation $C(t') = \langle S(t)r(t + t') \rangle$, we intend for t' to be understood as the time between when the stimulus is presented and when the response is affected. $C(t')$ will only be non-zero for positive values of t' because the LSISOS's in which we are interested are causal, that is they only respond to past inputs and the response cannot depend on the stimulus at a future time. Next, let us consider how the cross correlation is calculated. One chooses a time offset t' and fixes that value. Then for all stimuli at each time t , one looks ahead to the response at time $t + t'$, and forms the product $S(t)r(t + t')$. Then one averages up all these products to get the correlation

for the time offset t' . However, you could do it differently. After picking t' and fixing it, you could choose to define the running time variable as $u = t + t'$. And then you could choose $r(u)$ and look backward to the stimulus that preceded it by t' , that is $S(u - t')$, and form the product $S(u - t')r(u)$, and then average all these products as before. It is obvious that each of the products is the same whether you choose the stimulus first or the response first, and the sum is the same, and the sum is the cross correlation, $C(t') = \langle S(t)r(t + t') \rangle = \langle S(u - t')r(u) \rangle$. But if one chooses the response first and looks back in time to find the stimulus that preceded it, this is usually termed “reverse correlation.” But it should be clear that what one is doing is calculating the same cross correlation function as if one had chosen the stimulus first and calculated the “forward correlation.”

There are computational advantages to calculating the cross correlation in the reverse manner when studying neurons that fire nerve spikes, and it is this technical advantage that has inspired many papers that calculate the cross correlation by the reverse correlation procedure (de Boer & Kuyper, 1968; Jones & Palmer, 1987; McLean, Raab, & Palmer, 1994; Reid & Shapley, 1992; Ringach, Sapiro, & Shapley, 1997; Ringach, Hawken, & Shapley, 1997, among others). As we show below the main reason for the computational efficiency of “reverse correlation” in spiking neurons is that, in these case, the response can be considered a sequence of 0s and 1s (with 0s representing times when the cell did not spike and 1s representing times when the neuron spiked). Because there are usually a much smaller number of 1s than 0s in the sequence, and because accumulating multiplications of the input by zeros is not very helpful, one can restrict the calculation to those times when spikes are present. This makes the time complexity of the calculation linear in the number of spikes. However, for neurons that produce only slow potential responses to visual stimulation, for example, the horizontal cells in the retina, it is quite clear that the “reverse” and “forward” correlation are two names for the same calculation. An illustration of the results of such a calculation applied to horizontal cells in the catfish retina by Ken Naka and his colleagues is shown in Fig. 2 (from Naka, Chan, & Yasui, 1979). Here the horizontal cell (H cell) is acting approximately as an LSISOS around each mean level of illumination, and the cross correlation of the H cell’s response with a white-noise-modulated light allowed Naka and colleagues to calculate the impulse response of the H cell at each light adaptation level. These impulse responses reveal the slow time course of temporal integration in the retina. The weak overshoot of the response is an indicator of some amount of temporal adaptation to the stimulus even in the outer retina. This figure also illustrates that the time course of the impulse response is influenced strongly by the mean level of illumination—this is thus a very quantitative way of characterizing retinal light adaptation.



Fig. 2. Impulse response of a horizontal cell in the retina at different light adaptation levels as measured by Naka et al. (1979).

4. Spiking neurons

A neuron that fires spikes is not an LSISOS. The spike threshold makes it a highly nonlinear system, and we need to take this into account. Let us consider the system that has an LSISOS with impulse response $h(t)$ followed by a spike encoding mechanism, such as a Poisson spike generator. The spike encoder is one example of a class of possible nonlinear transducer stages. In this case [Bussgang's theorem \(1952\)](#) established that if one performs a cross correlation of the output of such a system with white noise input, one will obtain the result that the cross correlation $C(t')$ is proportional to $h(t)$:

$$C(t') = \langle S(t)r(t + t') \rangle = Kh(t) \quad (2)$$

where K is a constant that depends on the details of the nonlinearity in the spike encoder and the variance of the white noise signal. For neuroscience this is a very important result because it means we can use cross correlation to study the neural interactions that drive spiking neurons.

Now we can point out how the reverse (or triggered) correlation technique works for this case of an LSISOS followed by a spike encoder. All the more sophisticated spatio-temporal studies are just generalizations of this result. To calculate $C(t')$ we can choose to assemble the products of stimulus and response that make up the grand sum by first noticing the fact that the response is non-zero only at the times of occurrences of nerve spikes. So $C(t')$ is simply the finite sum of terms like $r(u_i)S(u_i - t')$ where u_i is the time of occurrence of the i -th spike. That is, for each spike we look back in time t' earlier and take the stimulus then and add it to the grand sum for $C(t')$ (because we assign unit height to each nerve spike). Then we get the reverse correlation equation ([de Boer & Kuyper, 1968](#)):

$$C(t') = \langle S(t)r(t + t') \rangle \approx \frac{1}{N} \sum_{i=1}^N S(u_i - t') \quad (3)$$

As above, we note this is simply the cross correlation function and in this case will equal $h(t')$, the impulse response of the LSISOS front end. One has to be aware, however, that not any type of spike encoder will lead to this result. A spike encoder that fires every time when the signal crosses a threshold (in the upward or downward direction), will result in a different expression ([de Boer & Kuyper, 1968](#)). However, for a more realistic model of a spike encoding mechanism that generates an inhomogeneous Poisson process whose rate is controlled by the membrane potential, [Eq. \(3\)](#) works.

5. Multi-input systems and the spatio-temporal impulse response

While LSISOS's are relevant to the study of temporal processing in audition, the field in which reverse correlation was introduced into neuroscience ([de Boer & Kuyper, 1968](#)), in visual neuroscience the nature of spatial interactions and spatial summation is equally if not more important. There have been several extensions of the reverse correlation technique to the study of neural systems that are characterized by spatial and temporal summation and integration. In a sense such systems can be viewed as having multiple inputs from each spatial

location in the receptive field, and a single output. A linear transducer of this type could be called a Linear Multiple Input Single Output System, i.e., LMISOS. Such a system can be characterized by its spatio-temporal impulse response, $h(x, y, t)$ (Brodie et al., 1978; Yasui, Davis, & Naka, 1979), a concept that was suggested in the highly original work by Rodieck (1965). The spatio-temporal impulse response will predict the response of a LMISOS to any arbitrary sequence of images. One way of seeing this intuitively is that $h(x, y, t)$ specifies the temporal impulse response for every position in the spatial field (x, y) . Since we are considering linear systems only here, the response to a temporal sequence of images will be the sum of the responses elicited by each of the locations in the image. Each of these separate responses can be calculated from the temporal sequence of intensity values of the image at each location, say for instance location (x, y) and the temporal impulse response at that location, $h(x, y, t)$. So once we know $h(x, y, t)$ for a LMISOS, we can predict its responses. Another way of understanding the spatio-temporal impulse response, $h(x, y, t)$ is that it is a temporal sequence of maps in the (x, y) plane that describes the time evolution of the neuron's spatial distribution of sensitivity, as we will illustrate in Fig. 5.

The operational question becomes, how can we recover $h(x, y, t)$ experimentally for such LMISOS's? There is a natural way to do this extending the notion of reverse correlation to input stimuli that are not simply sequences of numbers as a function of time, but rather sequences of visual images. Suppose each picture element, denoted (x, y) , is modulated in time by a white noise stimulus. Further, suppose all picture elements are modulated around the same mean luminance. And further, suppose that the cross correlation between the luminance values between any two positions on the image is zero, meaning that all the picture elements in the image are uncorrelated. Such an ideal stimulus is *spatio-temporal white noise*, and it is approximated by the "snow" one can see on a TV set that is not tuned to any program channel. If one were to use this stimulus as the input to a neuron that is an LMISOS, and were to perform cross correlations between each of the inputs at each value of (x, y) and the output, one could determine the spatio-temporal impulse response $h(x, y, t)$. In this case:

$$C_{x,y}(t') = h(x, y, t')$$

Now suppose, as is usually the case, that we need to estimate $h(x, y, t)$ for a neuron that fires spikes. So it cannot be viewed as an LMISOS but rather it could be modeled as an LMISOS followed by a nonlinear spike encoder. As with single-input systems, we can invoke Bussgang's Theorem and state that for a spiking neuron:

$$C_{x,y}(t') = Kh(x, y, t')$$

Thus for spiking neurons also, we can recover the spatio-temporal impulse response by calculating the cross correlation between the stimulus and the neuron's spike train when the neuron is stimulated by spatio-temporal white noise.

It is in the calculation of the cross correlation of spatio-temporal white noise with neuronal spike activity that reverse correlation has been used most effectively. One can calculate $C_{x,y}(t')$ by taking each spike in the spike train, and finding the image that preceded it by t' milliseconds, and then adding this image into an image accumulator (adding up the pixel intensities across images). One can do this for all values of t' . In this way one builds up a sequence of maps in (x, y)

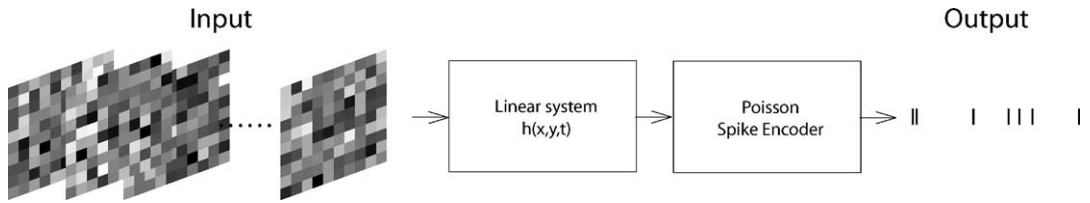


Fig. 3. Simple cortical cells can be modeled as a linear system acting on a spatio-temporal volume followed by a spike encoder. The impulse response of the linear system, $h(x, y, t)$, can be measured via reverse correlation when the system is stimulated with spatio-temporal white noise.

that describes spatial sensitivity as a function of time offset between stimulus and response (Fig. 3).

Erich Sutter provided one of the first demonstrations of how white noise analysis can be used to measure the spatio-temporal impulse response in cat visual cortex (Sutter, 1975). He devised an ingenious device where the (x, y) position of a dot on an oscilloscope was recorded on an analog tape and the responses of the neurons were recorded on a second analog tape. To compute the cross-correlation between stimulus and response the tapes would be delayed one with respect to the other by a few milliseconds and then played back. The average stimulus before a spike was computed by repeatedly exposing photographic film during the occurrence of the spikes (photographic summation), by having the spikes trigger the exposure of the camera. By repeating this procedure for different lags between the stimulus and response tapes Sutter calculated the entire spatio-temporal response for a simple cell in cat area 17. He observed the typical elongated “on” and “off” subfields described in the studies of Hubel and Wiesel. To our knowledge this is the first time such a calculation was performed in a cortical cell.

Measurements of the spatio-temporal impulse response by cross correlation with spatio-temporal white noise were also attempted in the retina by Hida and Naka (1982). They stimulated retinal ganglion cells in the catfish retina with white noise and derived spatial weighting functions at the time delay of peak response. They triggered on each ganglion cell’s spike and accumulated images at past times in a running sum, to estimate the cross correlation function. They observed asymmetries in the retinal ganglion cell receptive fields that had been unobservable before.

Other kinds of stimuli have been used to determine the spatio-temporal impulse response function. Jones and Palmer (1987) used small, rectangular bar stimuli, flashed at random locations in the receptive fields of V1 cells in cat visual cortex. By using reverse correlation between the evoked spikes and the positions and times of occurrence of the stimuli, they estimated the spatial impulse response at a fixed offset (50 ms). Fig. 4 illustrates the kind of (x, y) maps Jones and Palmer (1987) obtained for the spatio-temporal impulse response of cat V1 neurons. Jones and Palmer’s stimuli were approximately similar to white noise in power spectrum and autocorrelation, but differed greatly in other statistical measures. As has been discussed elsewhere (Reid, Victor, & Shapley, 1997), the stimuli of Jones and Palmer (1987) were “sparse” in time—most of the time the stimulus at each location (x, y) was zero. Similar results with similar techniques were obtained in later work by DeAngelis, Ohzawa, and Freeman (1993) who studied spatio-temporal receptive field properties in the adult cat as well as in kittens. The approach of Jones and Palmer and of DeAngelis et al. is very unlike

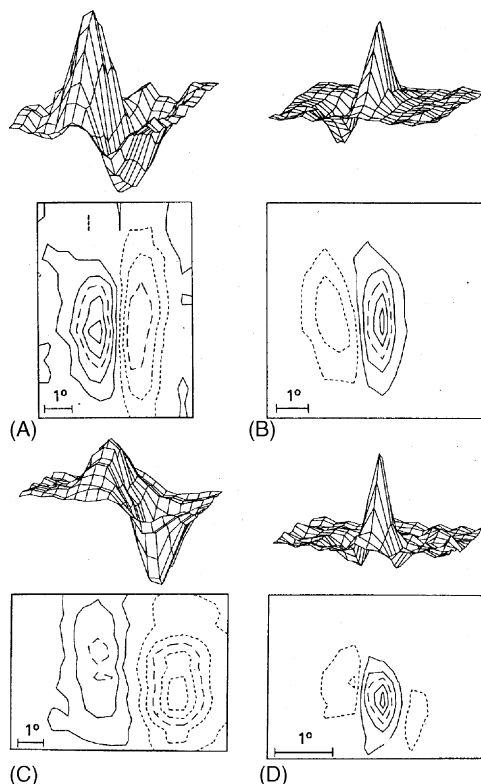


Fig. 4. Four examples (A–D) of spatio-temporal kernels of simple cells in cat area 17 measured by Jones and Palmer. The reverse correlation calculations in these examples were performed for a time lag of 50 ms.

the stimulus of Sutter (1975) or Hida and Naka (1982) in which each pixel was modulated like a Gaussian white noise. White noise stimuli like those used by Hida and Naka, have been termed “dense noise” as opposed to the “sparse noise” stimulus of Jones and Palmer (Reid et al., 1997). For a linear system, for example, an LMISOS, reverse correlation with either a dense or a sparse white noise would give the same result, and for both the cross correlation would equal the spatio-temporal impulse response. However, non-linear systems may have significantly different cross-correlations with the two different kinds of noise, and then one cannot easily equate the cross correlation with a spatio-temporal impulse function. For such non-linear systems, more analysis is required (see Reid et al., 1997; Victor, 1992).

It is worth comparing and contrasting the visual receptive field measurements made with the reverse correlation techniques and those obtained with “classical” receptive field mapping techniques. One cannot escape some kind of systems analysis technique if one is attempting to characterize a receptive field—it is like trying to speak without speaking in prose. In general, the transfer properties of a neuron can be characterized by cross-correlating stimulus with response whether the stimulus is a flashing spot, or a drifting bar, or a drifting or contrast-modulated sine grating. (Technically, one has to take into account the autocorrelation of the stimulus.) This is the reason for writing that all measurements of receptive fields are a kind of systems analysis, whether intentional or not.

The original idea of visual receptive fields came from H.K. Hartline in his studies of the visual properties of retinal ganglion cells in the frog retina (Hartline, 1940). This was an idea Hartline derived from E.D. Adrian, who in turn had used a term invented by C.S. Sherrington (cf. the scholarly review of the history of retinal ganglion cell physiology by Enroth-Cugell, 1993). The frog ganglion cells from which Hartline recorded had no maintained discharge of nerve impulses in the absence of stimulation, so the only nerve impulses they fired were either responses to spots of light that Hartline flashed on the frog's retina, or responses to the turning-off of the light. The area on the retina from which he could evoke nerve impulses from a ganglion cell Hartline termed the receptive field of that ganglion cell. Later work by Stephen Kuffler on cat retinal ganglion cells (Kuffler, 1953) required a modification of Hartline's definition of receptive field because cat ganglion cells have a maintained impulse rate in the absence of stimulation. The modified definition was that the receptive field was the region on the retina where stimulation could cause a *modulation of the impulse rate* around the average level of firing in the absence of stimulation.

It is also important that because of the maintained activity, Kuffler could observe stimulus-dependent modulations in spike rate above and below the maintained activity. For this reason, Kuffler was the first to recognize center-surround antagonism in the cat retinal ganglion cell: stimuli in the periphery of the receptive field caused a response modulation of the spike rate opposite in sign to that evoked by central stimuli (Kuffler, 1953).

From the beginning, it was clear that the receptive field was not an invariant characterization of the spatial, visual properties of a neuron. This is because, as defined, it varies in size with stimulus size and intensity and color and any other stimulus variable that determines the effectiveness of the stimulus in exciting the neuron. Hartline stated explicitly that the receptive field varied in size with size or intensity of stimulus spot (Hartline, 1940). For example, if one uses a stimulus spot with light intensity 1 unit, and then a second stimulus spot with light intensity 100, the receptive field mapped out with the second spot may be ten times bigger (or perhaps more!) than the field mapped out with the first spot. However, from his work on spatial summation, it is reasonable to suppose that already in the 1930s Hartline had the idea of spatial invariants of the visual properties of the ganglion cells he studied. These are the spatial distributions of sensitivity. A collection of equal sensitivity contours, one for each intensity of stimulus spot, will trace out the two-dimensional spatial sensitivity distribution, a surface, for the neuron studied. Hartline went further to predict the sensitivity for a compound stimulus which was the sum of two simpler stimuli. He found reasonably good agreement with linear summation of sensitivity (Hartline, 1940), the first result consistent with the concept of linear spatial summation weighted by the spatial sensitivity distribution.

But although Hartline and many others after him demonstrated some degree of linearity in the retinal network, the issue of linearity influences our judgments about what is the best way of doing systems analysis in the cerebral cortex. Most cells in the macaque monkey's primary visual cortex are not firing spikes until they are stimulated (Ringach, Bredfeldt, Shapley, & Hawken, 2002). But we know from the anatomy of the cortex that there are dense interconnections between cortical neurons. The multitude of cortico-cortical connections suggests that in the cortex we should expect cortico-cortical functional interactions to be important. But if we study the cortex with stimuli that excite only a small fraction of the cortical cells, we run the risk of missing the interactions and mistakenly overestimating the relative effectiveness of

feedforward coupling that is driven by the stimulus. This possible misestimation is traceable to the spike-firing threshold nonlinearity that is so important in the cortex. Therefore, using “classical” techniques like receptive field mapping with flashed spots and bars that work quite well in the retina may result in a serious misestimation of cortical spatial properties. This is where using “noisy” stimuli can come to the rescue. Such stimuli excite many cells simultaneously and thus allow us to observe the full range of cortico-cortical interactions that might be obscured with “classical” techniques applied to cerebral cortex, a highly-connected network of threshold elements.

6. Spatio-temporal m-sequences and reverse correlation mapping

Although systems analysis through cross correlation was an outgrowth of a theory based on random “noise” input signals, similar techniques have been developed for non-random input signals that have the right properties to be used for the same purpose. One particular type of signal has been used in many studies: the maximal length shift register sequence, or m-sequence (Sutter, 1987, 1992). An m-sequence is a string of -1 's and 1 's that looks like a random telegraph signal, but it is not random. Rather each m-sequence is calculated according to a deterministic algorithm that describes a series of operations on a shift register (for a detailed discussion, see Reid et al., 1997). As a consequence of the rules of generation of an m-sequence, each m-sequence has four properties that are important for systems analysis. These are: (1) the mean value of the m-sequence is close to 0; (2) the autocorrelation function of the m-sequence is also almost zero at all points, meaning equivalently that an m-sequence approximates white noise; (3) for an m-sequence of order n (meaning its total length is $2^n - 1$), every string of 1 's and -1 's of length n occurs exactly once in the m-sequence; (4) the product of an m-sequence with a time-shifted copy of itself is the same m-sequence but with a different starting point. Because of point 2, that an m-sequence has the same power spectrum as a white noise, m-sequences are very well suited to spatio-temporal measurements for visual neurons. Each pixel in the visual field can be modulated by the same m-sequence, with different starting points in time for each pixel. Then the cross correlation of the spatio-temporal stimulus, composed of all these time shifts of the one m-sequence, with the neuron's spike train, will yield cross correlation functions for each pixel that will be good estimates of the spatio-temporal impulse response at each pixel location.

Reid and Shapley (1992, 2002) used m-sequence analysis to study the receptive fields of visual neurons in the lateral geniculate nucleus (LGN) of macaque monkeys. The m-sequence is intrinsically a binary signal, and these investigators made use of this fact by varying the meaning of the two binary states of the m-sequence. They mapped the receptive fields with spatio-temporal m-sequences in which the two binary states were: black-white, red-green (isolating the L (565 nm) cone photoreceptor), red-green (isolating the M (535 nm) cone photoreceptor), and violet-yellow (isolating the S (440 nm) cone photoreceptor). Thus they were able to measure cone-specific spatio-temporal impulse responses for each neuron for all the cone inputs to that neuron, as well as its spatio-temporal impulse response for black-white stimuli. In this way they were able to establish that macaque LGN parvocellular cells were all single-opponent neurons that received one sign of input from one cone type, and another sign of input from an opponent cone type, for example, L+, M-. Furthermore, they could follow

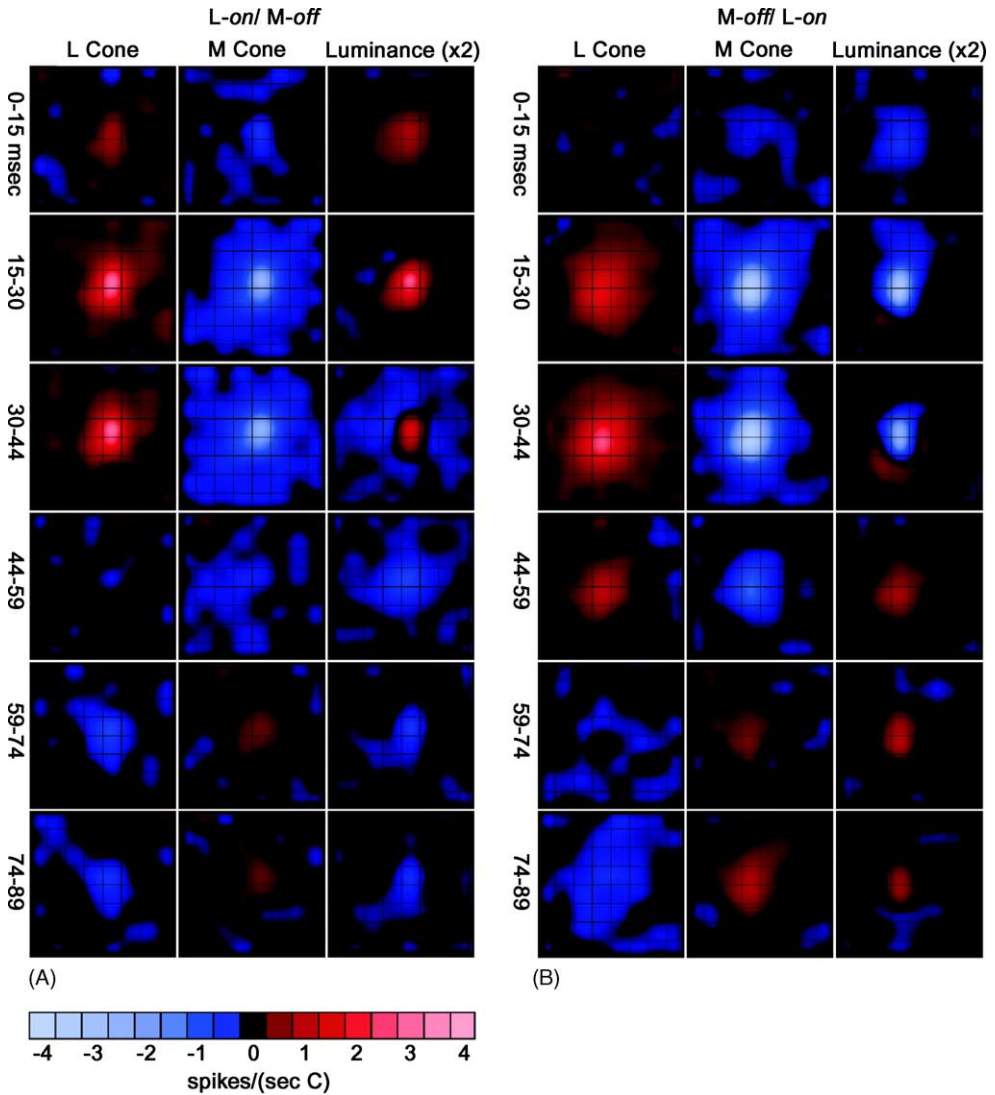


Fig. 5. (A and B) Spatial kernels for parvocellular LGN cells obtained by performing reverse correlation in cone-isolating directions (Reid & Shapley, 2002).

the input to the neuron from each cone type in time, at each location in the visual field. Such results are shown in Fig. 5 as a set of response “movie strips,” a column of slices through the spatio-temporal impulse response at different times (labeled in the figure).

7. Subspace reverse correlation and cortical receptive fields

As one moves from the retina to higher visual areas the receptive field of neurons become more complex. For example, some studies have found that neurons in V2 can be tuned to

two-dimensional stimuli such as angles, arcs, circles, and intersecting lines (Hegde & VanEssen, 2000; Pasupathy & Connor, 2002). Such cells would give very poor responses to white noise and therefore the applicability of reverse correlation methods appears limited. However, if we have some *a priori* knowledge of the type of information processed by a particular class of neurons, reverse correlation can be modified to take this into account. It has long been known, for example, that many neurons in V1 are selective to the orientation of the stimulus. It makes sense in some cases, then, to restrict the class of stimuli used to those that have one orientation. The reason such stimuli would be more effective in driving V1 neurons is that one has to wait a long period of time before an oriented pattern appears (by chance) in a sequence of a white noise images. We termed the idea of restricting the stimulus set to some smaller subset than the set of all possible random images “sub-space” reverse correlation. In the first application of the method the set of stimuli were gratings of various spatial frequencies, orientations and spatial phases (also called Hartley basis functions). It was shown that even though the set was restricted and quite different from white noise, the standard reverse correlation calculation would also lead to an estimate of the spatio-temporal receptive field of the neuron (under the condition that the maximum spatial frequency of the gratings used was beyond the window of visibility for the receptive field at hand).

Fig. 6a shows a sequence of Hartley basis functions. Such a sequence of images was presented to V1 neurons. The receptive field of a two simple V1 cells is shown in Fig 6b. Regions in red represent areas of the receptive field where increases in light cause the firing rate of the neuron to increase, and regions in blue indicate that increases in light at those locations cause the firing rate of the neuron to decrease. Restricting the space of stimuli used allowed us to increase the signal-to-noise of our measurements thereby achieving measurements of the first-order kernels of cortical cells within 10–15 min of experimental time (which would not be possible using white noise).

A slightly different analysis of the same data can provide a view of the receptive field in the Fourier domain. In these analyses, instead of calculating the mean image before the spike of a neuron, we calculate the mean amplitude spectrum of the stimulus. The result of such calculation in one neuron is shown in Fig. 6c. Here the mean two-dimensional spectrum of the image was calculated 60 ms before the firing of a spike during the presentation of a Hartley sequence. Regions in red indicate that gratings with the spatial frequency and orientation that are represented at that location induced the cell to increase its firing rate. Regions in blue, located at orientations orthogonal to the preferred, show that the cell was actually inhibited by the orthogonal orientation. Thus, selectivity in visual cortex arises not only because of enhancement of the response to the preferred stimulus but also by suppression of the response to non-optimal stimuli.

Sub-space reverse correlation can be extended to study shape selectivity in higher areas. For example, the sub-space of Cartesian, hyperbolic and parabolic gratings could be defined to study the dynamics of shape selectivity in V2 and V4. In a recent study, for example, Borghuis et al. (2003) used the reverse correlation method on the image velocity domain to study motion processing, and Cottaris and DeValois (1998) adopted our method to study the dynamics of color tuning in V1. In the following section we summarize our recent studies of orientation selectivity in V1 cortex applying the concept of subspace reverse correlation in the orientation domain.

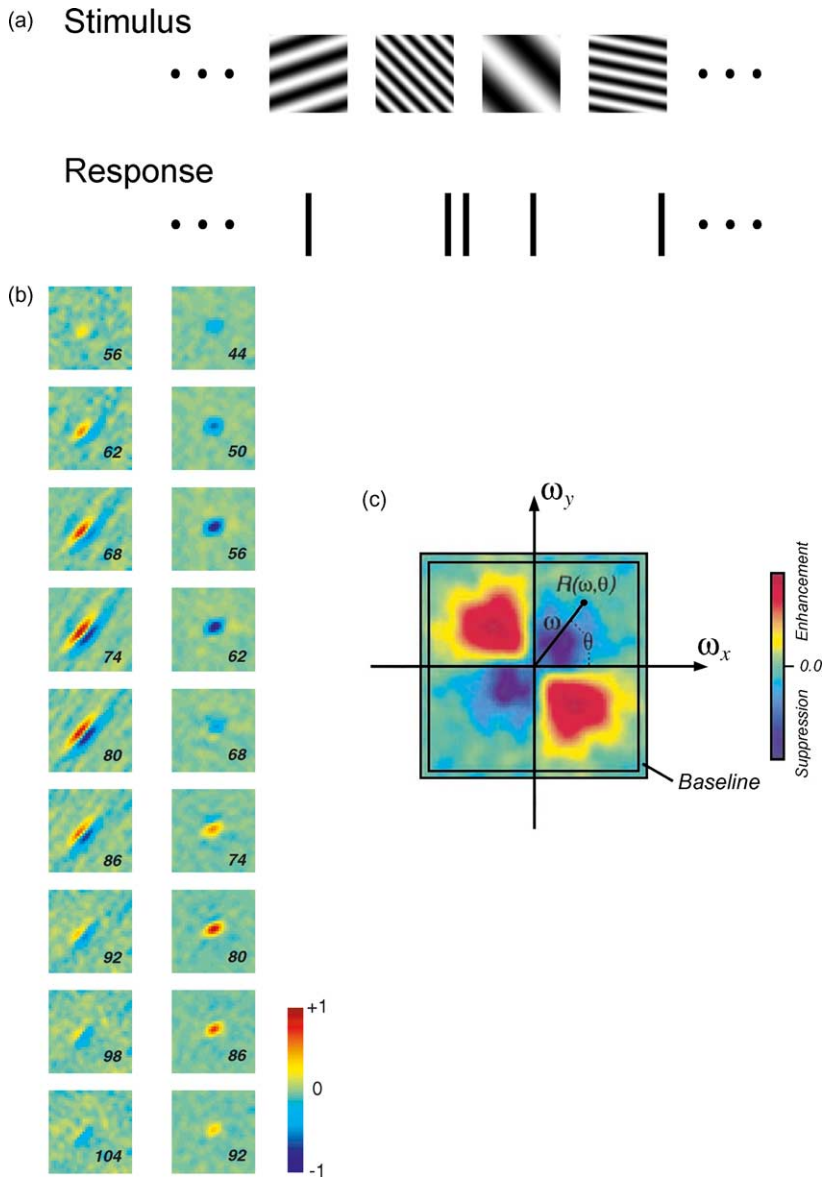


Fig. 6. (a) A sample sequence of Hartley basis functions. (b) Examples of the full impulse response, $h(x, y, t)$, of simple cells in macaque V1 obtained by the sub-space reverse correlation method (Ringach et al., 2002). (c) The probability of observing a particular combination of spatial frequency and orientation 50 ms before a spike in a V1 cell. Both enhancement and suppression are evident (from Ringach et al., 2002).

8. Orientation dynamics—categorical reverse correlation

In an attempt to provide data to test models of orientation selectivity, we used a reverse correlation method developed based on subspace reverse correlation in the orientation domain.

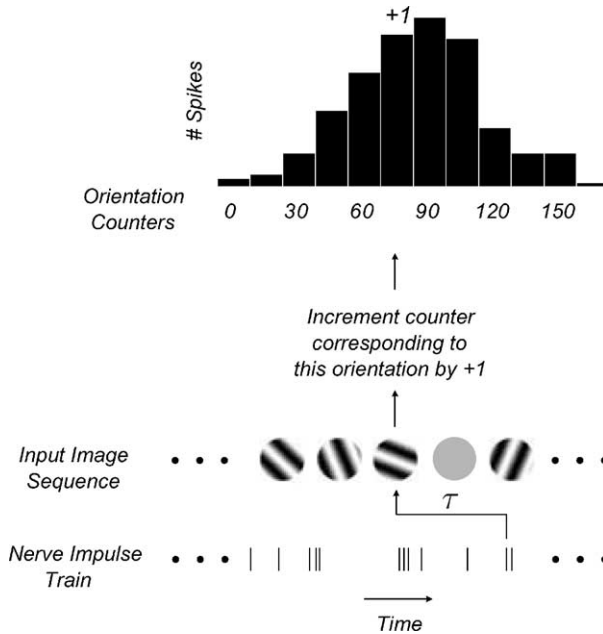


Fig. 7. Reverse correlation in the orientation domain.

The idea was to measure the time evolution of orientation selectivity extracellularly in single V1 neurons, with a technique that drove most cortical neurons above threshold. The technique is illustrated in Fig. 7. The input image sequence is a stimulus “movie” that runs for 15–30 min. Sinusoidal grating patterns (optimal size, spatial frequency, and high contrast) of orientations drawn randomly from a set of equally spaced orientations around the clock (usually in 10° steps) were presented for a fixed time (17 ms = 1 frame) at a 60 Hz refresh rate in the early experiments reported in Ringach, Hawken, et al. (1997), and 20 ms = 2 frames at 100 Hz refresh rate in the more recent experiments reported in Ringach, Hawken, and Shapley (2003). There is no interframe correlation in the stimulus between the grating patterns that are drawn at random. Furthermore, the cross correlation between any two grating patterns of different orientation is zero. Therefore such a stimulus approximates a white noise and the system’s properties can be obtained through cross correlation without correcting for autocorrelation within the stimulus (Victor, 1992). Each orientation was presented at four spatial phases and the response was phase averaged. For each time offset, the probability distribution for orientation $p(\theta, \tau)$ was calculated by reverse correlation: by incrementing the orientation bin corresponding to the orientation that preceded each of the N spikes at time offset τ , and then dividing the bin counts by N . N was usually of the order of 5,000 spikes. This calculation was done for each time offset between spike and stimulus to create a sequence of probability distributions for orientation, one for each time offset—an “orientation selectivity movie.”

The probability distribution $p(\theta, \tau)$ derived from reverse correlation in the case of the orientation dynamics experiment does not have the status of a spatio-temporal impulse response. Rather, it reflects the relative preference for a given stimulus (in this case an oriented grating

pattern) among a set of stimuli. The different patterns were evenly spaced along the dimension of orientation, θ . But in principle the set of patterns could have been any set of independent patterns, and a probability distribution could have been calculated by reverse correlation in a similar manner. So this is a more general application of the reverse correlation approach than in all the examples cited before in this review article, that aimed to calculate a spatio-temporal impulse response. This technique can be applied to a wider range of problems than can the linear systems approach that aims to measure the spatio-temporal impulse response of a linear transducer. The orientation dynamics measurement is applicable to non-linear systems as well as linear. The utility of the method is in testing models of neuronal networks that could generate probability functions $p(\theta, \tau)$ that can match the observed function. For instance, McLaughlin et al. (2000) tested feedforward and feedback models against the orientation dynamics data of Ringach, Hawken, et al. (1997) and found that a feedforward model could not account for major features of the probability distribution $p(\theta, \tau)$.

In our earlier work with the reverse correlation technique applied to the study of orientation dynamics (Ringach, Hawken, et al., 1997) we reported that most cells in the input layers 4C α and β have simple, “unimodal” dynamics and are relatively broadly tuned for orientation. By unimodal dynamics we meant that, after a time delay, the probability distribution for orientation simply had a single maximum in time and, after that peak, simply relaxed back to baseline. However, some cells in the output layers 2, 3, 4B, 5, and 6 showed “multimodal dynamics”: rebound responses, sharpening of the orientation tuning with time, and/or transient peaks of probability at off-optimal orientation. Also, in a few neurons in the output layers we observed a shift of the peak of the orientation probability distribution with time. These resemble the “shifter” cells described by Shevelev, Sharaev, Lazareva, Novikova, and Tikhomirov (1993). But shifter cells are the exception not the rule in macaque V1.

In more recent experiments on orientation dynamics (Ringach et al., 2003), we used a refined technique that revealed more about the basic mechanisms of orientation selectivity. As shown in Fig. 7, an additional pattern was added to the sequence—a blank stimulus at the mean luminance of the grating patterns. This allowed us for the first time to measure global enhancement and suppression because, with this new technique, one could estimate whether the effect of one of the oriented patterns was greater or less than that of the blank pattern. If the probability of producing a spike by a pattern of orientation θ is greater than that of a blank, we view that as evidence that a pattern of orientation θ produces net excitation, while if the probability of producing a spike by a pattern of orientation θ is less than that of a blank, we take this as an indication of suppression. Specifically, we take $R(\theta, \tau) = \log[p(\theta, \tau)/p(\text{Blank}, \tau)]$. If the probability that angle θ evokes a spike is greater than that of a blank screen, then the sign of R is +. If the probability that angle θ evokes a spike is less than that of a blank screen, then the sign of R is –. If all angles evoke a response above what a blank does, then $R(\theta)$ will have a positive value for all θ . A visual neuron equally well excited by stimuli of all orientation angles would produce a constant, positive $R(\theta)$.

We can estimate several useful features of the tuning curve $R(\theta, \tau)$, as shown in Fig. 8. These include: (a) the orientation angle of the peak response, θ_{pref} , and its magnitude $R(\theta_{\text{pref}}, \tau)$, (b) the orientation angle and magnitude of the minimum response, θ_{min} and $R(\theta_{\text{min}}, \tau)$, (c) the angle orthogonal to θ_{pref} , denoted θ_{ortho} , and its magnitude $R(\theta_{\text{ortho}}, \tau)$, (d) the “modulation depth” of the tuning curve as a function of time τ , $A(\tau) = R(\theta_{\text{pref}}, \tau) - R(\theta_{\text{min}}, \tau)$, and (e)

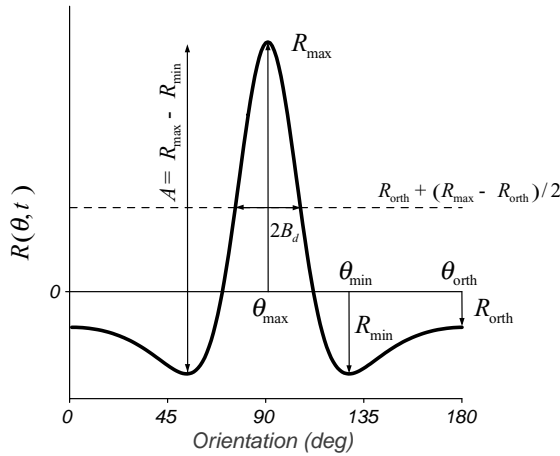


Fig. 8. Analysis of the log probability function $R(\theta, \tau)$.

the dynamic half-bandwidth defined by half the width of the tuning curve at the “half-height” which is equal to $1/2[R(\theta_{\text{pref}}, \tau) - R(\theta_{\text{ortho}}, \tau)]$. The orientation modulation depth $A(\tau)$ is a global measure of orientation selectivity because it is comparing the values of the tuning curve at two widely separated values of the angle θ . The bandwidth is a local measure of selectivity around the peak of the tuning curve.

The shape of the orientation tuning curve $R(\theta, \tau)$ changed with time, τ . This dynamic behavior has a number of important properties that are revealed in Fig. 9 for two representative V1 neurons. The black curves in Fig. 9 are graphs of $R(\theta, \tau)$ at the time offset τ_{peak} when the orientation modulation depth $A(\tau)$ reaches its maximum value. The red and blue curves are graphs of $R(\theta, \tau)$ at the two times bracketing τ_{peak} at which $A = 0.5 \times A_{\text{peak}}$. The red curve was measured at the development time τ_{dev} , the earlier of the two times when the modulation depth first rises from 0 to $0.5 \times A_{\text{peak}}$; the blue curve was taken at the declining time τ_{dec} when the response had declined back from A_{peak} to $0.5 \times A_{\text{peak}}$. One striking feature of these curves is that the dynamic tuning curve at the earlier time, $R(\theta, \tau_{\text{dev}})$, had a large positive pedestal of response, a sign of global excitation early in the response. This is just what one might predict from the analysis of feedforward models of V1 orientation selectivity (e.g., Hubel & Wiesel, 1962), if indeed the earliest responses measurable were predominantly feedforward excitation. But then, as the response evolved in time, the maximum value of $R(\theta, \tau)$ at the preferred orientation grew only a little, while the responses at non-preferred orientations declined substantially. Thus Fig. 9 demonstrates that the maximum orientation modulation depth occurred at a time when inhibition had suppressed non-preferred responses. Because such inhibition suppressed all responses far from the preferred orientation, we infer that it was global (untuned) inhibition. It is also reasonable to infer that tuned excitation near the preferred orientation counteracted the global inhibition to maintain the peak value of $R(\theta, \tau)$. While this kind of evidence for global inhibition was apparent in most V1 neurons we studied, a significant fraction of the cells exhibited a different kind of inhibition, as shown in responses of the cell illustrated in the lower panel of Fig. 9. For this cell, at τ_{dec} the tuning curve $R(\theta, \tau_{\text{dec}})$ had the shape of

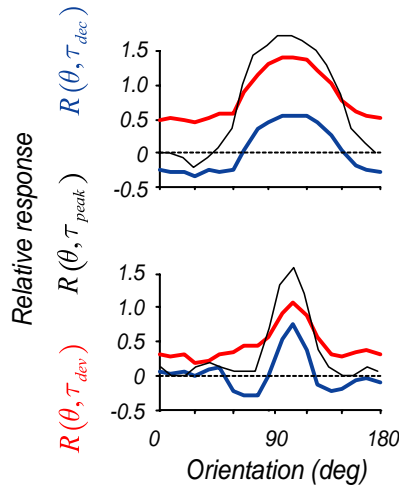


Fig. 9. Time development of $R(\theta, \tau)$ in two example cells in macaque V1.

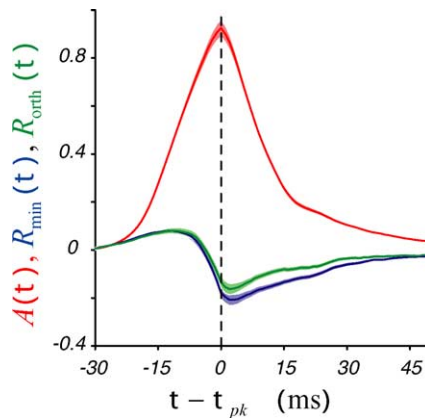


Fig. 10. Average dynamics of the modulation amplitude, minimum and orthogonal response in macaque V1.

a “Mexican hat” meaning that $R(\theta_{\min}, \tau_{\text{dec}}) < R(\theta_{\text{ortho}}, \tau_{\text{dec}})$. We interpret this to mean that there was also tuned inhibition in such neurons, that is, inhibition that does not extend out to orientations far from the preferred. Both the hypothesized inhibitions, global and tuned, would have to be relatively rapid in time course to have the effects on the tuning curve seen in Fig. 9 at the time of peak selectivity.

While orientation bandwidth often has been the focus of interest in previous research, it is rather the global shape of the orientation tuning curve at all orientations that differentiates between different theoretical mechanisms. Therefore, we studied $R(\theta_{\min}, \tau)$, $R(\theta_{\text{ortho}}, \tau)$ and the modulation depth $A(\tau)$ in V1 neurons. The average behaviors of $R(\theta_{\min}, \tau)$, $R(\theta_{\text{ortho}}, \tau)$ and $A(\tau)$ averaged over a population of 242 V1 neurons are depicted in Fig. 10. The modulation depth, $A(\tau)$, normally increased to reach a peak and then declined back to baseline over a time course of 50 ms. An important feature is the positive sign of $R(\theta_{\min}, \tau)$ and $R(\theta_{\text{ortho}}, \tau)$ early in the response,

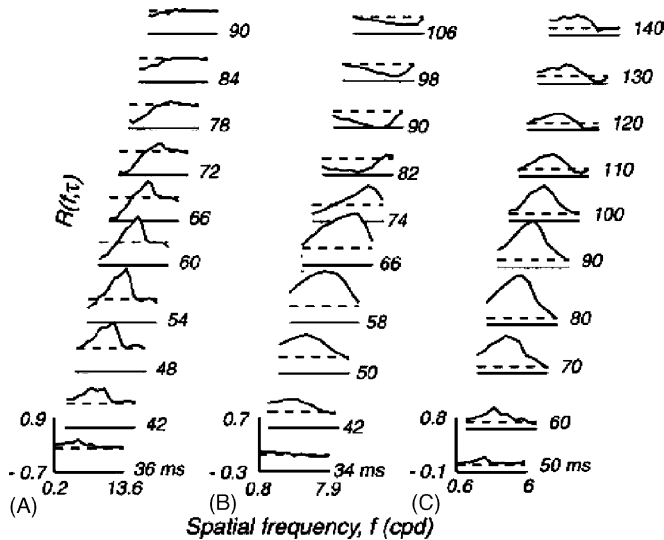


Fig. 11. (A–C) Examples of the dynamics of spatial frequency tuning in macaque V1 (Bredfeldt & Ringach, 2002).

indicating that, on average, V1 cells tended to respond to *all* orientations early in the response. Another important feature of the data was the sharp downward change in time course of $R(\theta_{\min}, \tau)$ and $R(\theta_{\text{ortho}}, \tau)$ before $A(\tau)$ reached its peak value. This is evidence for rapid inhibition and also for the likely influence of inhibition on the magnitude of orientation modulation depth. Eventually both $R(\theta_{\min}, \tau)$ and $R(\theta_{\text{ortho}}, \tau)$ declined to negative values meaning that later in the response orientations far from the preferred orientation were suppressive not excitatory.

9. Reverse correlation mapping of spatial frequency selectivity and its dynamics

A final example of how reverse correlation was applied to the study of V1 neurons is the measurements of spatial frequency dynamics by Bredfeldt and Ringach (2002). Here, a sequence of gratings having random spatial frequency and spatial phase (but fixed orientation) is presented to a cortical neuron. One can then calculate the relative probability that each of the spatial frequencies appeared in the input sequence a few milliseconds before the neuron spiked. This study found that a large proportion of V1 cells show inseparable responses in spatial frequency and time (see Fig. 11). In several cases, tuning becomes more selective over the course of the response, and the preferred spatial frequency shifts from low to higher frequencies. Many responses also showed suppression at low spatial frequencies (Fig. 11), that correlates with the increases in response selectivity and the shifts of preferred spatial frequency. The shift in preferred spatial frequency from low to high spatial frequencies was later replicated by Mazer, Vinje, McDermott, Schiller, and Gallant (2002) using similar methods in the awake monkey. These findings showed that it is incorrect to interpret the behavior of V1 as a bank of spatial frequency filters with labeled lines. Instead, spatial frequency information is multiplexed over time over the same line, in a fashion suggestive of coarse-to-fine processing of the visual image.

10. Conclusions

Cross-correlation provides a powerful method for studying neuronal signal processing. When neurons and neural networks act as if they were linear systems, cross correlation with noisy inputs allows one to measure the impulse response, or the spatio-temporal impulse response, and in that way to characterize the neuron or neural network completely. Spiking neurons allow experimenters to calculate the cross correlation function quickly and efficiently using the algorithm of reverse correlation, or spike triggered averaging. This technique has opened new windows into the dynamic function of neural networks in the brain, and it has great possibilities for future research. The method is also applicable to situations where a system is linear except for its last state that may include a static nonlinearity (Chichilnisky, 2001; Nykamp & Ringach, 2002). The resulting reverse-correlation kernels can be evaluated by calculating the proportion of explained variance in the data (DeAngelis et al., 1993). If only a small portion of the variance is explained, it is likely that the data cannot be well approximated by a linear-nonlinear system and further methods of analysis may be required (see Reid et al., 1997; Victor, 1992).

References

- Borghuis, B. G., Perge, J. A., Vajda, I., van Wezel, R. J., van de Grind, W. A., & Lankheet, M. J. (2003). The motion reverse correlation (MRC) method: A linear systems approach in the motion domain. *Journal of Neuroscience Methods*, 123(2), 153–166.
- Bracewell, R. (1998). *The Fourier transform and its applications*. New York: McGraw-Hill.
- Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *Journal of Neuroscience*, 22, 1976–1984.
- Brodie, S. E., Knight, B. W., & Ratliff, F. (1978). The spatiotemporal transfer function of the *Limulus* lateral eye. *Journal of General Physiology*, 72, 167–202.
- Bussgang, J. J. (1952). *Correlation functions of amplitude distorted Gaussian signals*. Technical Report 216, MIT Research Laboratory of Electronics.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, 12(2), 199–213.
- Cottaris, N. P., & DeValois, R. L. (1998). Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature*, 395(6705), 896–900.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69, 1118–1135.
- de Boer, R., & Kuypers, P. (1968). Triggered correlation. *IEEE Transactions of Biomedical Engineering*, 15, 169–179.
- Enroth-Cugell, C. (1993). The world of retinal ganglion cells. In D. K. Lam & R. M. Shapley (Eds.), *Contrast sensitivity* (pp. 149–179). Cambridge: MIT Press.
- Hartline, H. K. (1940). The receptive fields of optic nerve fibers. *American Journal of Physiology*, 130, 690–699.
- Hegde, J., & VanEssen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *Journal of Neuroscience*, 20(5), RC61.
- Hida, E., & Naka, K. (1982). Spatio-temporal visual receptive fields as revealed by spatio-temporal random noise. *Zeitschrift für Naturforschung*, 37, 1048–1049.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160, 106–154.
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat's visual cortex. *Journal of Neurophysiology*, 58, 1187–1211.

- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.
- Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., & Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 1645–1650.
- McLaughlin, D., Shapley, R., Shelley, M., & Wielaard, J. (2000). A neuronal network model of sharpening and dynamics of orientation tuning in an input layer of macaque primary visual cortex. *Proceedings of the National Academy of Science United States of America*, 97, 8087–8092.
- McLean, J., Raab, S., & Palmer, L. A. (1994). Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neurosciences*, 11, 271–294.
- Naka, K. I., Chan, R. Y., & Yasui, S. (1979). Adaptation in catfish retina. *Journal of Neurophysiology*, 42, 441–454.
- Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, 2(1), 1–11.
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, 5(12), 1332–1338.
- Reid, R. C., & Shapley, R. M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature*, 356, 716–718.
- Reid, R. C., & Shapley, R. M. (2002). Space and time maps of cone photoreceptor signals in macaque lateral geniculate nucleus. *Journal of Neuroscience*, 22, 6158–6175.
- Reid, R. C., Victor, J. D., & Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. *Visual Neurosciences*, 14, 1015–1027.
- Ringach, D. L., Bredfeldt, C. E., Shapley, R. M., & Hawken, M. J. (2002). Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1. *Journal of Neurophysiology*, 87(2), 1018–1027.
- Ringach, D. L., Hawken, M. J., & Shapley, R. M. (1997). The dynamics of orientation tuning in the macaque monkey striate cortex. *Nature*, 387, 281–284.
- Ringach, D. L., Hawken, M. J., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: The role of global and tuned suppression. *Journal of Neurophysiology*, 90(1), 342–352.
- Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse correlation method for the study of visual neurons. *Vision Research*, 37, 2455–2464.
- Ringach, D. L., Shapley, R. M., & Hawken, M. J. (2002). Orientation selectivity in macaque V1: Diversity and laminar dependence. *Journal of Neuroscience*, 22, 5639–5651.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5, 583–601.
- Shevelev, I. A., Sharaev, G. A., Lazareva, N. A., Novikova, R. V., & Tikhomirov, A. S. (1993). Dynamics of orientation tuning in the cat striate cortex neurons. *Neuroscience*, 56, 865–876.
- Sutter, E. (1975). A revised conception of visual receptive fields based on pseudorandom spatio-temporal pattern stimuli. In *Proceedings of the conference on testing and identification of nonlinear systems, March 18–20*. Caltech/NSF.
- Sutter, E. (1987). A practical non-stochastic approach to nonlinear time-domain analysis. In *Advances in methods of physiological systems modelling* (Vol. 1). University of Southern California.
- Sutter, E. E. (1992). A deterministic approach to nonlinear systems analysis. In R. Pinter & B. Nabet (Eds.), *Nonlinear vision: Determination of neural receptive fields, function and networks* (pp. 171–220). Cleveland, OH: CRC Press.
- Victor, J. D. (1992). Nonlinear systems analysis in vision: Overview of kernel methods. In R. Pinter & B. Nabet (Eds.), *Nonlinear vision: Determination of neural receptive fields, function and networks* (pp. 1–37). Cleveland, OH: CRC Press.
- Yasui, S., Davis, W., & Naka, K. I. (1979). Spatio-temporal receptive field measurement of retinal neurons by random pattern stimulation and cross correlation. *IEEE Transactions of Biomedical Engineering*, 26, 263–272.